Towards sustainable land and water management: Eco-hydrological and global scale modelling

A thesis submitted to The University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Science and Engineering

2021

Amirhossein Hassani

Department of Chemical Engineering and Analytical Science

Contents

List of Figures

List of Tables

Abstract

More than 40% of the Earth' surface is covered by drylands which are home to nearly three billion people. Climate change and over-pressure on land and water resources as a result of population growth have endangered the ecosystem functioning in drylands. Ecosystem degradation could adversely affect people's livelihoods and well-being. This study helps with decision-making to address two land and water resources managerial challenges in drylands: desiccating saline lakes and soil salinization. Saline lakes as the main natural aquatic feature of dryland landscapes are desiccating at alarming rates in the recent decades, predominantly due to human interventions and climate change. The result is increased water salinity levels which negatively impacts the ecosystems integrity in the nearby regions. To sustainably restore and preserve the saline lakes in the short- (e.g. 10 years) and long-term periods (e.g. 30 years), a four-step eco-hydrological framework, primarily based on land use strategies is proposed here: (1) projecting the future governing climatic patterns in the basin under study; (2) evaluating the needed water for restoration of the saline lake; (3) prioritising and allocation of the water to all environmental, municipal, industrial, and agricultural water users; and (4) optimisation of the agricultural land use considering the total available land and freshwater resources. The applicability of the framework was examined by the case of Lake Urmia in Iran — known formerly as the second (by volume) hyper-saline lake in the world. The results show that for restoration of Lake Urmia, annually 3,648 Mm^3 (~70% increase) and 3,692 Mm^3 (~73% increase) surface water inflow to the lake is required to restore the lake in 30 years under the two greenhouse gases emission scenarios of RCP 4.5 and RCP 8.5, respectively. Provision of these inflow volumes needs respective reductions of 95,600 ha and 133,687 ha in the Urmia basin irrigated area. The proposed framework can inform decision-making for adapting and enhancing the resilience of the saline lakes to negative consequences of the over-exploitation of water resources, particularly in the context of projected climate change.

The second part of this research estimates the extent, severity, and long-term trends in soil salinization as one of the land degrading threats in drylands. Soil salinization is a dynamic and common environmental issue; however, there remains considerable uncertainties regarding its large-scale spatio-temporal variability and relation to the future climate change. Due to unavailability of the detailed soil and plant data for application of physical- or numerical-based methods, Machine Learning (ML) techniques were used for prediction of the soil salinization trends on a global scale. Using the ML techniques, measured legacy soil profile data were related to a set of environmental predictors to estimate the soil salinity (and sodicity where Na^+ is higher compared to other soluble salts) at other locations, depths, and times. With a similar approach and based on outputs of the Global Circulation Models, the trends in drylands' soil salinization were predicted to the end of the 21st century. Between 1980 and 2018, the results indicate that soil salinity and sodicity have been temporally and geographically highly variable. Additionally, compared to the 1961 - 1990 period, the results show that primary (naturally-occurring) soil salinity will be a major issue in the drylands of South America, southern and Western Australia, Mexico, southwest United States, and South Africa by the end of the 21st century. The results provided by this study could contribute to sustainable land and water management in dryland regions of the world.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

Acknowledgments

Foremost, I would like to thank my supervisory team, Professors Adisa Azapagic and Nima Shokri, who always supported me, monitored my progression, and gave me credit whenever it was necessary during these three and almost a half years. Thank you Nima for trusting me and giving me the opportunity to come to Manchester. My PhD was funded by the Presidential Doctoral Scholarship Award provided by The University of Manchester which is highly acknowledged. Also, thank you my dearest family and friends for being always there for me.

Chapter 1 Introduction

Lands, including water bodies located within them, form the basis for our livelihoods, well-being, and biodiversity by providing multiple benefits to net primary production and serving diverse ecosystem services (Shukla et al. 2019). The Millennium Ecosystem Assessment (2005) defines these ecosystem services in four major groups including (1) provisional services which provide food, water, or fibre, (2) regulating services which have impacts on air and water quality, (3) cultural services such as aesthetic values, recreation, and ecotourism, and (4) supporting services that help photosynthesis, soil formation, and nutrient cycling. Climate change and direct and indirect human activities are among the crucial threats to the integrity of potential or existent ecosystem services provided by the lands and water bodies located within them (FAO 2015; Sivakumar et al. 2007). Loss of the land utility and production is a serious issue for the environment around the world (Ravi et al. 2010; Nkonya et al. 2016; Cerretelli et al. 2018). Land degradation, or negative trend in the long-term land quality/functionality caused by direct or indirect anthropogenic activities will remain high on the agenda in the 21st century (Eswaran et al. 2001).

To a large extent, climatic stressors, growing population, unsustainable agricultural activities, and rapid urbanisation/industrialisation are the main drivers of the land and water resources degradation (Ferreira et al. 2018; Blaikie et al. 2015). Globally, there is a large uncertainty in the estimated extent and spatial distribution of the degraded lands to the point that estimates vary between < 10 Mkm^2 and > 60 Mkm^2 (Gibbs et al. 2015). Degradation of the lands and their water bodies are particularly of concern in the world's drylands where all the provided supporting, regulating, provisioning, and cultural ecosystem services to humans are highly vulnerable to variability in precipitation, air temperature, and soil fertility (Enfors et al. 2008; Mortimore 2005).

Projected population growth and irrigation expansion has unprecedentedly increased the demand for water resources across the world and particularly the drylands (Sterling et al. 2013). Since the 1980s, many rivers have been dammed to meet the water demands (Pekel et al. 2016). The water scarcity in drylands has been exacerbated by recent observed drier trends in climatic conditions (Santos et al. 2014; Haddeland et al. 2014) to the point that water usage in many basins has exceeded the water availability (Molle et al. 2010). Accordingly, from the management point of view, optimal allocation of water resources to meet urban, industrial, and agricultural needs, whilst assuring that all other major ecosystems services are provided, remains a grand challenge in basins located in drylands (Shadkam 2017). One of the forms of water resources degradation in drylands is the desiccation of saline lakes, which has negatively impacted the ecosystem well-being in their surrounding areas (Williams 2002b; Wurtsbaugh et al. 2017).

On the other hand, the dryland populations' livelihoods are primarily dependent on agriculture, which is one of the sectors highly vulnerable to climate change (Rosenzweig et al. 2014). Land degrading processes such as soil salinization, soil erosion, and overgrazing have adversely impacted the drylands' provided ecosystem services (Morgan 2009; Majeed et al. 2019; Mirzabaev et al. 2016). Similar to land degradation on a global scale, currently, the estimates on the extent and severity of the land degradation in drylands are just crude approximates with substantial uncertainties. This is majorly due to data unavailability and methodological limitations (Cherlet et al. 2018a).

This thesis is particularly focused on two aspects of water and land resources degradation in drylands, i.e. saline lakes' desiccation and soil salinization. How saline lakes can sustainably be restored and estimation of the past, current, and future trends in soil salinization processes are the research questions addressed in this thesis. In what follows, first drylands are introduced and then a brief overview of the saline lakes' drying and soil salinization issues are provided. This is followed by the motivations and aims of the study and finally, by an overview of the research methodologies used to address in this research.

1.1 Drylands

Drylands are traditionally characterised based on the Aridity Index (AI) – the ratio of average annual precipitation amount to potential evapotranspiration amount (Middleton et al. 1997). Drylands are the regions with an AI ≤ 0.65 mm mm^{-1} (Middleton et al. 1997). Aridity is different from drought. Aridity is a long-term climatic feature of drylands with low precipitation and/or available water while drought is a short-term climatic event not restricted to drylands (Le Houérou 1996). Based on the AI, drylands can be classified into four climate areas including hyper-arid (AI between 0 and 0.05 mm mm^{-1}), arid (0.05 - 0.2 mm mm^{-1}), semi-arid (0.2 - 0.5 mm mm^{-1}) and dry sub-humid (0.5 - 0.65 mm mm^{-1}) (UNCCD 1994). Africa and Asia together have 70% of the global drylands (Figure 1-1). Excluding hyper-arid regions, grasslands, forests, and croplands are the largest land cover types in drylands. Together with hyper-arid regions, drylands cover about 46.2% (±0.8%) of the global land area (Prăvălie 2016).

Currently, close to 38% of the global population (near to 3 billion people) are residents of the drylands (Koutroulis 2019), the majority of them located in the drylands of South Asia, sub-Saharan Africa, and Latin America (van der Esch et al. 2017). Developing countries are home to close to 90% of this population (Shukla et al. 2019). It is estimated that population in the drylands will reach 4 billion people by 2050 (van der Esch et al. 2017). This will be twice the population increase projected for non-drylands. Estimates of Le et al. (2016) show that between 1980s and 2000s, approximately 500 (±120) million people living in the drylands faced the negative impacts of land degradation on bio-mass productivity.

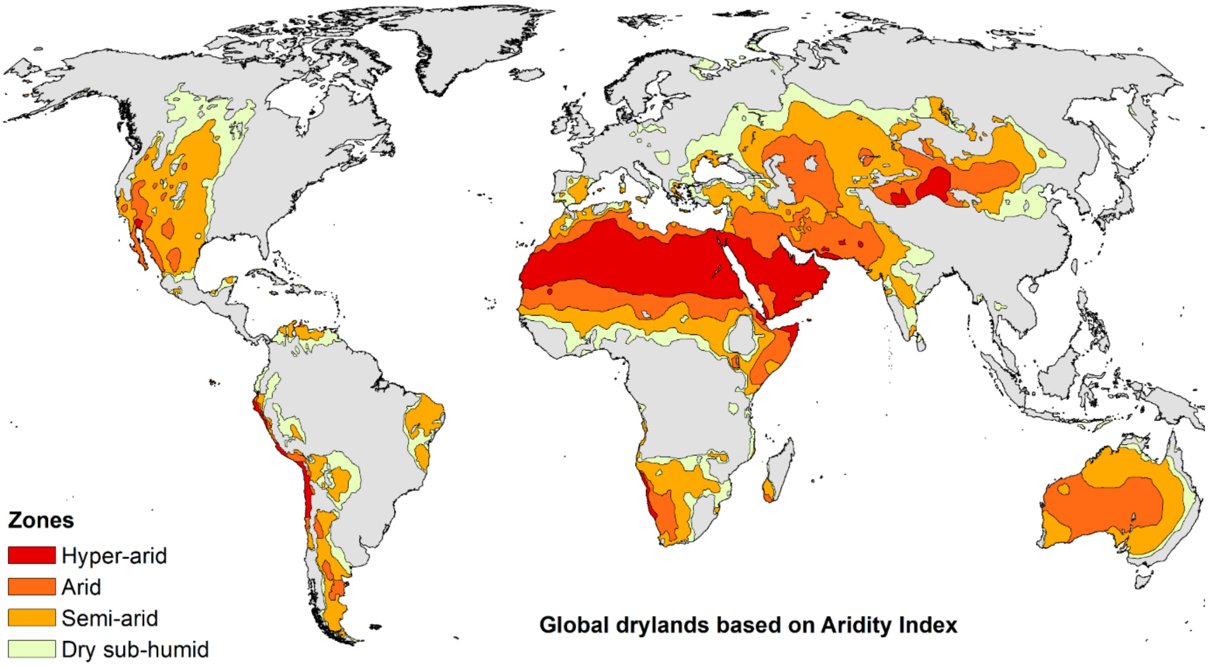


Figure 1-1: Global distribution of the drylands.

Traditionally, the response of the dryland populations to environmental degradation factors has been described in the context of two general frameworks (Shukla et al. 2019). The first is the “Desertification Syndrome” framework, which puts the stress on the human role and recognises the unsustainable anthropogenic activities as the main factor for poverty of the drylands’ livelihoods and subsequent inability for investment in resource conservation (Millennium Ecosystem Assessment 2005). The second is “Dryland Development”, which additionally considers the role of natural processes and recognises interactions between human activities and climate variabilities as the driver of resource degradation (Reynolds et al. 2007). In general, the consensus is that complex interactions between the human- and climate-induced drivers result in degradation of ecosystem services provided by drylands; which one is more significant, human or climate however, differs from place to place (Ravi et al. 2010; D’Odorico et al. 2013). Accordingly, analysing the environmental resource degradation in the drylands needs an integrated approach to both natural- and human-related parameters as neglecting either of these may result in under- or over-estimation of impacts (Harrison et al. 2016).

1.2 Saline lakes

Saline lakes or salt lakes are the water bodies surrounded by land with salt concentrations of sodium chloride and other dissolved salts of more than 3 g L^{-1}. In “Saline Lake Ecosystems of the World”, Hammer (1986) classifies the lakes into four groups in terms of salinity: (1) sub-saline: 0.5 - 3 g L^{-1}, (2) hypo-saline: 3 - 20 g L^{-1}, (3) meso-saline: 20 -50 g L^{-1}, and (4) hyper-saline: greater than 50 g L^{-1}. Saline lakes are the main aquatic landscape feature of the (semi-) arid and subtropical regions (Williams 1996) and represent nearly 44% of the total volume and 23% area of the all inland water bodies (Messager et al. 2016b). Saline lakes are majorly found in closed (endorheic) basins where the evaporation is excessive over the precipitation (Hammer 1986; Williams 2002a). The water input of the saline lakes is usually supplied by one or more rivers. In the majority of cases, the supplying rivers to the saline lakes are the main source of dissolved salts in their water. The entering salts cannot leave as the lake is terminal and excess of evaporation over water input to the lake increases the salt concentration. By far, the Caspian Sea is the largest saline lake in the word which is filled by approximately 41% of the total inland saline water in the world (Wurtsbaugh et al. 2017). Lake Urmia in Iran used to be the second largest hyper-saline lake by volume, but is now near to 90% desiccated, as discussed further below.

As saline lakes are predominantly located in water scarce areas, they provide a diverse range of socio-economic and ecological services (Wurtsbaugh et al. 2017). Saline lakes provide economic services, such as shipping and mineral extraction. Compared to freshwater lakes, saline lakes can better accumulate and recycle nutrients (Blomqvist et al. 2004; Wurtsbaugh et al. 2017) and therefore, they can be suitable habitats for diverse fish species. In the case of higher salt concentrations where the water salinity levels are not tolerable by the fishes (such as hyper-saline lakes), dormant eggs (cysts) of the invertebrate food organisms, such as brine shrimp (*Artemia spp.*) can be harvested as a source of food in the fish cultivation industry. Saline or hyper-saline lakes can also serve as the habitat for migratory birds and invertebrate food organisms, such as brine shrimp and brine flies (*Ephedra spp.*) can be used by migrating birds during long migration cycles (Roberts 2013). For example, Andean salars and the above-mentioned Lake Urmia are home to the migrating flamingos (Wurtsbaugh et al. 2017). Saline lakes are also of the aesthetic values and, similar to freshwater bodies, can be used for recreational activities, such as boating, fishing, swimming, and other aquatic sports (Case 2013; Micklin et al. 2008). In a broader sense, saline lakes can affect regional climate patterns, supporting the biological diversity and productivity, and maintaining environmental and human health (Williams 1996; Hammer 1986; Hassani et al. 2020).

As the formerly fourth largest lake and the second largest salt lake in the world, the Aral Sea started to shrink in 1960s and in 1997, retaining only 10% of its original volume (Micklin 2007). A great portion of this shrinkage is attributed to agricultural development. Due to the ever-growing demand for water, more or less similar drying patterns, or so-called Aral Sea syndrome (AghaKouchak et al. 2015) has been recorded for other major saline and hyper-lakes across the globe, except Antarctica (Messager et al. 2016b). The desiccation of saline lakes is not a recent-observed phenomenon (Williams 1993). For example, the diversion of water by the city of Los Angles for agricultural expansion and urban water use completely dried the Owens Lake in western US before the 1940s (Wurtsbaugh et al. 2017). The combination of climate change and climate variability with unsustainable human activities has accelerated the desiccation of saline lakes (Shadkam 2017). Since 1996, the Caspian Sea surface water level (above the mean sea level) has been dropping by almost 7 cm per year (Chen et al. 2017). Similarly, due to a combination of human interventions and climate change, Iran's Lake Urmia lost almost 90% of it volume within only 20 years since 1995 (AghaKouchak et al. 2015; Schulz et al. 2020). Furthermore, due to a series of unsustainable water management decisions and reduction in water input, the California's Salton Sea water level has declined more than 7 m since 2000 (Case 2013).

Severe desiccation of the saline lakes can negatively impact the health and economic resilience of the people living in their surroundings (Micklin 2007; Frie et al. 2019). The saline dried-up lands (playa) remaining after the severe desiccation of the saline lakes can be a source for saline dust and air pollutants (Abuduwaili et al. 2010; Griffin et al. 2004). Saline particulate matters originated from the dried bed of the saline lakes can be transferred to the regions far from the location of the saline lake and harm human health, agriculture, and soil bio-productivity. Exposure of 12,700 km^2 of the Aral Sea bed to wind resulted in substantial ecological loss, deterioration of the bio-diversity, and associated difficulties for the local communities (Crighton et al. 2011; Wiggs et al. 2003). An increase in the number of lung infections and respiratory diseases was also reported as a consequence of the wind-blown saline dusts originated from the Owens Lake in California (Cahill et al. 1996). By further shrinkage, the salt concentration of lakes' water increases and salinity levels become unbearable even by salinity-tolerant micro-organisms — resulting in the loss of fisheries or economic profits from harvesting of brine shrimp eggs. Salinities beyond the tolerance of the fishes in the Aral Sea cut around 60,000 jobs related to the fishing industry, leading to 40,000 t reduction in annual fish production (Micklin et al. 2008). In the case of Lake Urmia in Iran, brine shrimp was entirely eliminated from the lake at the water salinities above 250 g L^{-1} and thus, the flamingos on the top of the food chain vanished (Abbaspour et al. 2007). Although desiccation of the saline lakes can initially benefit the mineral extraction industries, complete desiccation or further recession of the shorelines can cut the linkage between the solar evaporation ponds and the lake's water body (Wurtsbaugh et al. 2017). Monetizing the loss of other ecological and biological services provided by (hyper-) saline lakes is not easy as the consequences are mostly irreversible, geographically widespread, and have resulted in degraded ecosystems (Hammer 1986).

1.3 Soil salinity and sodicity

Soil salinity refers to amount of soluble salts in the soil (Soil Survey Staff 2010). Cations including sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), and magnesium (Mg^{2+}), and anions including chloride (Cl^-), sulphate (SO_4^{2-}), bicarbonate (HCO_3^-), carbonate (CO_3^{2-}), and nitrate (NO_3^-) compose the major soluble minerals in the soil (Shahid et al. 2018). Soil sodicity, however, refers to high concentration of sodium salt (Na^+) compared to concentration of other exchangeable cations in the soil, such as calcium, magnesium, and potassium (Richards 1954; Soil Survey Staff 2010). In fact, soil sodicity is a special form of the soil salinity in which high levels of Na^+ accumulates at the expense of other soil exchangeable cations. The origin of these soluble salts in the soil is different. In primary soil salinization, physical and chemical naturally-occurring processes are the main sources of the soluble salts in the soil. These processes include surface evaporation, weathering of the saline parent material (rocks from which the soil is formed), and transferring of the saline minerals by wind or (sub-) surface waters to the location of soils. In secondary soil salinization, however, the origin of soluble salts is human activities. Irrigation with brackish/saline water, land use/cover change, dryland management activities, seawater intrusion as a result of coastal groundwater over-exploitation, over-use of fertilisers, dumping industrial waste brine, and poor drainage conditions in agricultural lands are all the examples of such human activities.

Conventionally, electrical conductivity of water-saturated soil paste extract (EC_e) has been used as indicator of soil salinity (Soil Survey Staff 2010). As laboratory measuring of the EC_e is time consuming and needs trained technicians, electrical conductivity of other soil to water (w/v) extracts including 1:1, 1:2, 1:2.5, 1:5, and 1:10 have been used to measure the extent of soil salinity (Shahid et al. 2018). These extractions can be prepared in field and then using the pre-defined correlations, the measured electrical conductivities can be converted to EC_e. However, still EC_e is preferred to EC of extracts prepared by other soil to water ratios as it is the best representative of the soil solution at the real field conditions. Additionally, the correlations for converting the EC of other extractions to EC_e are highly site-specific and there is no universal correlation for all soils.

On the other hand, the level of soil sodicity has been traditionally defined by exchangeable sodium percentage (ESP) (Soil Survey Staff 2010). Soils have a capacity to adsorb positively charged constituents of dissolved salts called cation exchange capacity (CEC). ESP is the extent to which sodium occupies the adsorption complex of a soil. ESP is usually represented as a percentage and calculated by: $\text{ESP} = (Na_{exc} / CEC) \times 100$ (Richards 1954), where ESP is the soil exchangeable sodium percentage (ratio), Na_{exc} is the soil exchangeable sodium in *mEq*/100 g soil, and *CEC* is the soil cation exchange capacity in *mEq*/100 g soil.

Soil salinization is a land degrading process that increases the concentration of soluble salts in the soil. Soil salinity is particularly important in the root zone as it can affect crops and plants productivity. A saline soil is a soil with excessive concentration of the soluble salts that negatively impacts the soil's fertility and stability. Higher salinity in soil solution increases the energy needed by soil micro-organisms and plants for withdrawal of water from the soil — leading to death of the plants and further soil erosion (Shrivastava et al. 2015). The presence of sodium, chlorine, and boron has toxic effects on plants' leaves. Excessive soil salinity in the root zone can increase the ion toxicity in plants' leaves, nutrient deficiency (N, Ca, K, P, Fe, Zn), and lower the bio-mass productivity (Munns et al. 2008). Soils with an EC_e less than 2 dS m^{-1} are known to be non-saline. A soil salinity between 2 and 4 dS m^{-1} can negatively affect the yields of sensitive crops (fruit trees and vegetables). Soil salinities between 4 and 8 dS m^{-1} restrict the growth of the majority of crops and plants and only salt tolerant crops can sustain at the salinities higher than 8 dS m^{-1} (Soil Survey Staff 2010). Similarly, high levels of soil sodicity can adversely affect the crop productivity and soil stability (Richards 1954). Sodicity causes decreased permeability and poor drainage — leading to a higher soil surface erosion during heavy rainfall or irrigation cycles (Daliakopoulos et al. 2016). Soils with an ESP more than 6% are considered to be sodic (Northcote et al. 1972; Isbell 2016); however, depending on the soil clay type, sodicity-related problems may emerge at other sodicity thresholds. Salt-affected soil is a generic term that includes both salinity and sodicity issues (Daliakopoulos et al. 2016).

1.4 Background and research motivations

The United Nations Sustainable Development Goal (SDG) 6 aims to ensure the freshwater availability, increase water-use efficiency, and restore/protect aquatic ecosystems (UN General Assembly 2015). Furthermore, the SDG 15 is to "protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt/reverse land degradation and biodiversity loss" (UN General Assembly 2015). However, as previously mentioned, due to population increase and growing food demand, sustainable management of land and water resources has become a grand challenge, particularly in drylands (Sakadevan et al. 2010).

The main reason for global drying of the saline lakes is that over the recent decades, provision of water for growing human water demands has been preferred to the environmental water needs (Williams 2002a). In response to the desiccation crisis of saline lakes, three strategies have been adopted by watershed/basin mangers to preserve or restore the lakes (Wurtsbaugh et al. 2017) including (1) reducing upstream water consumption, particularly in agricultural sector, (2) partial restoration (restoration of some parts of the lakes instead of the whole lakes), and (3) transfer of water from the adjacent basins to compensate the water deficit in the saline lakes' basins.

Approaches such as reducing the irrigation area, increasing agricultural water productivity, and cultivating less water intensive crops have been traditionally suggested for reducing the water uptake at supplying river basins (Wada et al. 2014); however, success and sustainability of these approaches depends on the adaptability and commitment of all stakeholders, particularly the farmers who use upstream water for irrigation (Shadkam 2017). Although partial restoration may not be a long-term sustainable solution for saving the full range of the ecosystem services provided by saline lakes (Wurtsbaugh et al. 2017; Micklin et al. 2008), it can be used as a temporary solution while more sustainable solutions are under development (Wada et al. 2014). However, in many cases, the logistic and financial costs do not allow construction of ducts/dams and smaller lakes. Transfer of water from other basins cannot be a sole sustainable solution as these projects need substantial funding and the consequent water deficits in the donor basins negativity impact the environmental sustainability. Therefore, a combination of the two or three of these strategies is required for sustainable restoration/preserving a saline lake. Identifying long-term sustainable solutions for restoration and preserving the saline lakes requires a complete balance between the economic and environmental services provided by these lakes. A sustainable solution must take into account all relevant economic, social, and environmental aspects. A comprehensive analysis that takes into account these aspects for sustainable restoration and preservation of the desiccating saline lakes in the context of future climate uncertainties, particularly based on agricultural land use management, has not been addressed adequately in the literature so far.

On the other hand, as one of the widespread and serious forms of land degradation, significant uncertainties and knowledge gaps remain concerning the soil salinization extent, severity, and its relation to climate change (Shukla et al. 2019). This leads to a need for developing predictive tools to identify the locations of saline and sodic soils and their spatio-temporal variability on the global scale (FAO Soils Portal 2020). However, previous studies have primarily focused on the past or current trends in spatio-temporal variability of the soil salinity at the local scales. Complex interactions between socio-economic, bio-physical, and environmental parameters peculiar to each location make it a challenge to map the land degradation and its various forms on a global scale (Cherlet et al. 2018b).

Currently, Harmonised World Soil Database (Fao/Iiasa/Isric/Isscas/Jrc 2012) and WISE-30 (Batjes 2015) are the main databases that provide global quantitative information on distribution of the salt-affected soils. However, these databases are purely spatial and mostly based on outdated 1970s data. Neither can a more recent attempt of Ivushkin et al. (2019) on generating a spatio-temporal map from the world’s salt-affected soils capture quantitatively the complex dynamism involved in soil salinization processes. Their model is a classifier which classifies the soils into saline and non-saline classes and variations in soil salinity within the classes cannot be quantified. Additionally, their work does not differentiate between the salinity and sodicity. Therefore, an up-to-date and consistent analysis of the variability of different aspects of the salt-affected soils at the global scale is lacking. Moreover, a quantitative investigation of the future trends in global soil salinity in the face of future climatic uncertainties is rare. Such analyses can bring new insights into our understating of the global environmental change and is crucial to agro-ecological modelling, land assessment, crop growth simulation, and sustainable water resources management.

1.5 Aims and objectives

Taking a quantitative approach, this thesis aims to fill the research gaps on the management of land and water resources in drylands mentioned in the previous section. Particularly, in the face of future climatic uncertainties, the main aim of this research is assisting policy and decision-making in devising sustainable action plans against the two forms of land and water resources degradation in drylands, i.e. desiccation of saline lakes and soil salinization. The specific goals of this thesis are:

- to identify sustainable solutions for restoration and preservation of desiccating saline lakes in various time horizons, considering socio-economic, hydrologic, climatic, and land-related stressors;
- to identify the parts of a saline lake that should be prioritised for restoration;
- to estimate the extent and severity of soil salinity and sodicity on a global scale;
- to estimate the current soil salinization rates and trends in salt-affected lands, from the country to global levels;
- to develop a quantitative tool for predicting variations in primary soil salinity in the mid-term (2031 - 2060) and long-term (2071 - 2100) futures; and
- to identify the soil salinization hotspots around the globe by the end of the 21st century.

In summary, the novel contributions to knowledge provided by this work include:

- development of a decision-support framework based on optimal agricultural land use and cropping pattern at a river-basin scale for sustainable restoration of desiccating saline lakes across different time horizons (e.g. 10 or 30 years);
- development of a remote sensing-based technique for identifying the parts of a saline lake that should be prioritised for restoration in combination with semi-analytical models for estimation of the wind-blown dust flux;
- development of four dimensional predictive models/tools, taking into account the longitude, latitude, time, and depth, for spatio-temporal prediction of soil salinity and sodicity;
- generating high-resolution schematic maps of different aspects of variation in surface soil salinity and sodicity (referring to the top 30 cm of the soil layer) on a global scale between 1980 and 2018; and

- development of quantitative predictive models/tools based on the outputs of the Global Circulation Models for predicting the annual variability of primary soil salinity by the end of the 21st century.

1.6 Methodology

In this section, an overview of the methods and frameworks proposed to address the above questions is provided (Figure 1-2). Details on the applied methodological approaches can be found in chapters 2 to 4.

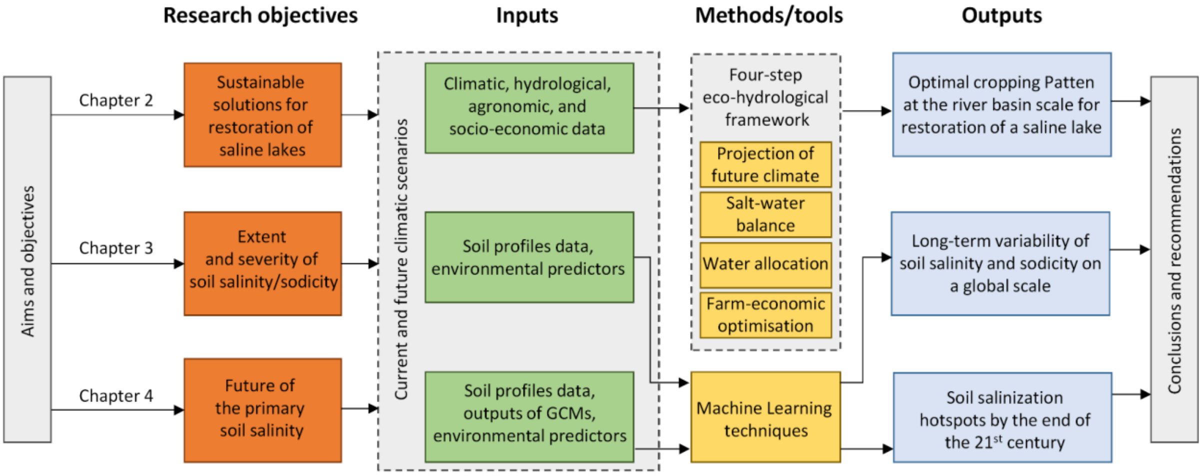


Figure 1-2: Schematic representation of the methodological framework used in the study (GCMs: Global Circulation Models).

1.6.1 Sustainable solutions for restoration of saline lakes

To provide sustainable solutions for restoration of the desiccating saline lakes, a four-step eco-hydrological framework is proposed here, which takes all climatic, hydrologic, agronomic, and socio-economic aspects of saline lakes' basins into account (Chapter 2). The proposed framework is principally based on reducing the upstream water uptake for agricultural activities and optimisation of cropping patterns in the river basins supplying a saline lake whilst trying to maximise the farmers' net income. The effect of diversion of water from other basins was also reflected in the development of the framework. In addition to this framework for restoration of desiccating saline lakes in long-terms, consideration is given to partial restoration as a short-term solution. The applicability of the proposed framework is illustrated by the case of Lake Urmia in north western Iran — as mentioned earlier, formerly the second largest hyper-saline lake of the world (by volume).

Depending on the time horizon considered for restoration of a saline lake, the first step in the framework is projection of governing climatic conditions in the lake's basin. In the analysis considered in this thesis, the target year for restoration of Lake Urmia is assumed to be 2050 (30-year period). Outputs of the Global Circulation Models (GCMs) are used for projection of the climatic parameters. Air temperature and precipitation over the basin area are the main required climatic parameters. Due to high uncertainties in the outputs of the GCMs, multi-GCM ensembles under different greenhouse gas emission trajectories are used for projecting the future climates.

The second step in the framework is development of a hydrologic balance over the lake area, coupled with salt balance models to estimate the water evaporation/subsurface flow from/into the lake. At large scales, these two parameters are not easily available for saline lakes and require using indirect techniques, such as application of hydrological or energy balance over the lake area, to be estimated. In this research, a calibrated salt-coupled hydrological balance over the lake's area is used to estimate the annual water needed to reach the restoration target water level. The target water level differs from lake to lake and depends on the ecosystem services needed to be returned by restoration of the saline lake. For the case of Lake Urmia, 1274.1 m (above the mean sea level) was selected as the target level as it could be the basis for returning the brine shrimp into the aquatic environment of the lake (Abbaspour et al. 2007). In the third step, based on the projected population growth, the municipal and industrial water demands within each supplying river basin is estimated and allocated.

Finally, in step four, alternative sustainable land use and cropping patterns in supplying river basins for reducing water uptake in the agricultural sector are analysed to provide annual required water inflow for the lake's restoration. From the total water available in each basin, the water remaining for agricultural sector is used as a constraint for an economic optimisation model at the river basin scale to obtain the optimal agricultural land use pattern, considering the total available land resources. The water diverted from the adjacent basins can be added to the water available for the agriculture sector. The final aim of the economic optimisation model is to maximise the net income for the farmers. Additionally, in this step, the potential greenhouse gas emissions as a result of the recommended land use change activities are considered in the optimisation model. As a result, the proposed land use/cropping pattern scheme can assure the preservation of the saline lake.

As reaching the main objectives of the agricultural land use optimisation may not be feasible in the short-term, partial restoration is considered as a parallel solution. For partial restoration of a saline lake, the priority is given to restoration of the locations at higher risk of wind-blown dust generation. In particular, a remote sensing-based technique coupled with semi-analytical models for estimation of the wind-blown dust flux is developed to recognise the parts of a saline lake that should be prioritised for restoration (Chapter 2).

1.6.2 Extent and severity of soil salinity/sodicity

As previously mentioned, measuring soil EC_e and ESP is the traditional method for assessing soil salinity and sodicity extent. Despite the accuracy and reliability, the applicability of this method remains at a farm-scale or for local assessments as it is a costly method and includes a large spatial discontinuity in results. Expert judgment and satellite remote sensing are the other two methodological approaches that have been usually used for estimation of the extent and severity of soil salinization at local to global scales. Expert judgment, however, cannot be consistent and quantitative for the variations of the soil salinity and sodicity over the time (Ivushkin et al. 2019). Remote sensing also has its own drawbacks for universal applications. In crop or vegetated lands, the remote sensing techniques cannot be reliable for assessing the soil salinity and sodicity as the soil spectral reflectance is highly blended by the vegetation reflectance (Shahid et al. 2010). Additionally, remote sensing is sensitive to top soil layer reflectance and salinity at deeper soil layers cannot be estimated. Application of model-based techniques for estimating the extent and severity of soil salinity and sodicity is another option (Feddes et al. 1988; Saito et al. 2006). However, numerical and analytical models for simulation of water and solute transport in the root zone are limited to short-term applications at the farm-scale as they need very detailed properties of soil and land data and are computationally demanding (Suweis et al. 2010).

To overcome some of the above difficulties, the Digital Soil Mapping (DSM) framework (Jenny 1994) is applied in this analysis to estimate the spatio-temporal variability in soil salinity and sodicity on a global scale. It is assumed that soil salinity and sodicity are governed by soil-forming factors including climate, organisms, relief, parent material, and time. If the relationship between soil EC_e and ESP and those soil-forming factors and the spatially explicit distribution of soil-forming factors are available, the distribution of soil salinity and sodicity can be estimated from the distribution of soil-forming factors (called predictors hereafter) (Omuto et al. 2013). Here, Machine Learning (ML) techniques are used for characterising the relation between the soil salinity/sodicity and predictors. ML models have been trained based on the experimentally measured EC_e and ESP values (soil profiles data) and a set of environmental predictors to estimate annual soil salinity and sodicity between 1980 and 2018 at high spatial resolutions (Chapter 3). The models' outputs are used to estimate the long-term variability of soil salinity and sodicity on a global scale.

1.6.3 Future of the primary soil salinity

Similarly, using ML techniques, the extent and severity of primary soil salinity is projected to the end of the 21st century using a set of environmental predictors and available projected values of some predictors, including precipitation frequency, precipitation intensity, evapotranspiration, and wet (rain) and dry (wind) deposition rates of sea salts, obtained from the outputs of Global Circulation Models (Chapter 4). Model predictions are used to identify the hotspots of soil salinization on a global scale

1.7 Thesis outline

This thesis is presented in the Journal Format comprising three papers, presented in chapters 2 to 4. The thesis author is the first author of all the three papers. The thesis structure and a short summary of what each chapter covers is provided below:

Chapter 2 (Paper 1): ***Desiccation crisis of saline lakes: a new decision-support framework for building resilience to climate change*** (published in Science of the Total Environment)

Chapter 2 presents the new eco-hydrological framework developed in this work for restoration of desiccating saline lakes due to human activities, which is of great importance for ecosystem functioning, agriculture, water management, and human well-being. Its applicability is illustrated by application to the Urmia basin. Furthermore, the developed method for identifying the locations of a saline lake with priority of restoration is also presented.

Chapter 3 (Paper 2): ***Predicting Long-term Dynamics of Soil Salinity and Sodicity on a Global Scale*** (Published in PNAS)

This chapter quantifies the spatio-temporal variability of soil salinity and sodicity on a global scale. A ML-based methodology is developed for spatio-temporal prediction of the soil salinity and sodicity and the results of the developed methodology are used to map the annual soil salinity and sodicity variation between 1980 and 2018.

Chapter 4 (Paper 3): ***Climate change and soil salinization: A global scale perspective for the 21st century***

The final paper included in this thesis is among the first attempts for projecting the spatio-temporal variations of the dryland soil salinity by the end of 21st century in response to climate change. ML-based models have been trained based on the current global trends in soil salinity to project the primary soil salinization hotspots to the end of the century.

References

Abbaspour, M., & Nazaridoust, A. (2007). Determination of environmental water requirements of Lake Urmia, Iran: an ecological approach. *International Journal of Environmental Studies, 64*(2), 161-169.

Abuduwaili, J., DongWei, L., & GuangYang, W. (2010). Saline dust storms and their ecological impacts in arid regions. *Journal of arid land, 2*(2), 144-150.

Agh, N. (2007). *Characterization of Artemia populations from Iran*: Ghent Univ.

AghaKouchak, A., Norouzi, H., Madani, K., Mirchi, A., Azarderakhsh, M., Nazemi, A., et al. (2015). Aral Sea syndrome desiccates Lake Urmia: call for action. *Journal of Great Lakes Research, 41*(1), 307-311.

Batjes, N. H. (2015). *World soil property estimates for broad-scale modelling (WISE30sec)*. (No. 2015/01). ISRIC-World Soil Information.

Blaikie, P., & Brookfield, H. (2015). *Land degradation and society*: Routledge.

Blomqvist, S., Gunnars, A., & Elmgren, R. (2004). Why the limiting nutrient differs between temperate coastal seas and freshwater lakes: a matter of salt. *Limnology and Oceanography, 49*(6), 2236-2241.

Cahill, T. A., Gill, T. E., Reid, J. S., Gearhart, E. A., & Gillette, D. A. (1996). Saltating particles, playa crusts and dust aerosols at Owens (dry) Lake, California. *Earth Surface Processes and Landforms, 21*(7), 621-639.

Case, H. (2013). *Salton Sea ecosystem monitoring and assessment plan*: US Department of the Interior, US Geological Survey.

Cerretelli, S., Poggio, L., Gimona, A., Yakob, G., Boke, S., Habte, M., et al. (2018). Spatial assessment of land degradation through key ecosystem services: the role of globally available data. *Science of The Total Environment, 628*, 539-555.

Chen, J., Pekker, T., Wilson, C. R., Tapley, B., Kostianoy, A., Cretaux, J. F., & Safarov, E. (2017). Long-term Caspian Sea level change. *Geophysical Research Letters, 44*(13), 6993-7001.

Cherlet, M., Hutchinson, C., Reynolds, J., Hill, J., Sommer, S., & von Maltitz, G. (2018a). World atlas of desertification (p. 295). Luxembourg: Publication Office of the European Union.

Cherlet, M., Hutchinson, C., Reynolds, J., Hill, J., Sommer, S., & von Maltitz, G. (2018b). World atlas of desertification. Luxembourg: Publication Office of the European Union.

Crighton, E. J., Barwin, L., Small, I., & Upshur, R. (2011). What have we learned? A review of the literature on children’s health and the environment in the Aral Sea area. *International Journal of Public Health, 56*(2), 125-138.

D’Odorico, P., Bhattachan, A., Davis, K. F., Ravi, S., & Runyan, C. W. (2013). Global desertification: drivers and feedbacks. *Advances in water resources, 51*, 326-344.

Daliakopoulos, I., Tsanis, I., Koutroulis, A., Kourgialas, N., Varouchakis, A., Karatzas, G., & Ritsema, C. (2016). The threat of soil salinity: A European scale review. *Science of the Total Environment, 573*, 727-739.

Enfors, E. I., & Gordon, L. J. (2008). Dealing with drought: The challenge of using water system technologies to break dryland poverty traps. *Global Environmental Change, 18*(4), 607-616.

Eswaran, H., Lal, R., & Reich, P. (2001). *Land degradation: An overview In: Bridges, EM, ID Hannam, LR Oldeman, FWT Pening de Vries, SJ Scherr, and S. Sompatpanit (eds.). Responses to Land Degradation.* Paper presented at the Proc. 2nd. International Conference on Land Degradation and Desertification, Khon Kaen, Thailand. Oxford Press, New Delhi, India.

FAO, I. (2015). Status of the world's soil resources (SWSR)–main report. *Food and agriculture organization of the United Nations and intergovernmental technical panel on soils, Rome, Italy, 650*.

FAO Soils Portal, S. R., Food and Agriculture Organization of the United Nations. (2020). Extent of salt-affected soils, http://www.fao.org/soils-portal/soil-management/management-of-some-problem-soils/salt-affected-soils/more-information-on-salt-affected-soils/en/.

Fao/Iiasa/Isric/Isscas/Jrc. (2012). Harmonized world soil database (version 1.2). *FAO, Rome, Italy and IIASA, Laxenburg, Austria*.

Feddes, R., Kabat, P., Van Bakel, P., Bronswijk, J., & Halbertsma, J. (1988). Modelling soil water dynamics in the unsaturated zone—state of the art. *Journal of hydrology, 100*(1-3), 69-111.

Ferreira, C. S., Walsh, R. P., & Ferreira, A. J. (2018). Degradation in urban areas. *Current Opinion in Environmental Science & Health, 5*, 19-25.

Frie, A. L., Garrison, A. C., Schaefer, M. V., Bates, S. M., Botthoff, J., Maltz, M., et al. (2019). Dust sources in the Salton Sea basin: a clear case of an anthropogenically impacted dust budget. *Environmental science & technology, 53*(16), 9378-9388.

Gibbs, H., & Salmon, J. M. (2015). Mapping the world's degraded lands. *Applied geography, 57*, 12-21.

Griffin, D. W., & Kellogg, C. A. (2004). Dust storms and their impact on ocean and human health: dust in Earth's atmosphere. *EcoHealth, 1*(3), 284-295.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., et al. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences, 111*(9), 3251-3256.

Hammer, U. T. (1986). *Saline lake ecosystems of the world* (Vol. 59): Springer Science & Business Media.

Harrison, P. A., Dunford, R. W., Holman, I. P., & Rounsevell, M. D. (2016). Climate change impact modelling needs to include cross-sectoral interactions. *Nature Climate Change, 6*(9), 885-890.

Hassani, A., Azapagic, A., D'Odorico, P., Keshmiri, A., & Shokri, N. (2020). Desiccation crisis of saline lakes: A new decision-support framework for building resilience to climate change. *Science of the Total Environment, 703*, 134718.

Isbell, R. (2016). *The Australian soil classification*: CSIRO publishing.

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., & De Sousa, L. (2019). Global mapping of soil salinity change. *Remote Sensing of Environment, 231*, 111260.

Jenny, H. (1994). *Factors of soil formation: a system of quantitative pedology*: Courier Corporation.

Koutroulis, A. G. (2019). Dryland changes under different levels of global warming. *Science of The Total Environment, 655*, 482-511.

Le Houérou, H. N. (1996). Climate change, drought and desertification. *Journal of Arid Environments, 34*(2), 133-185.

Le, Q. B., Nkonya, E., & Mirzabaev, A. (2016). Biomass productivity-based mapping of global land degradation hotspots. *Economics of land degradation and improvement–A global assessment for sustainable development*, 55.

Majeed, A., & Muhammad, Z. (2019). Salinity: a major agricultural problem—causes, impacts on crop productivity and management strategies. In *Plant abiotic stress tolerance* (pp. 83-99): Springer.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature communications, 7*(1), 1-11.

Micklin, P. (2007). The Aral sea disaster. *Annu. Rev. Earth Planet. Sci., 35*, 47-72.

Micklin, P., & Aladin, N. V. (2008). Reclaiming the Aral Sea. *Scientific American, 298*(4), 64-71.

Middleton, N., & Thomas, D. (1997). *World atlas of desertification.. ed. 2*: Arnold, Hodder Headline, PLC.

Millennium Ecosystem Assessment, M. (2005). *Ecosystems and human well-being: desertification synthesis*: World Resources Institute, Washington DC, USA.

Mirzabaev, A., Goedecke, J., Dubovyk, O., Djanibekov, U., Le, Q. B., & Aw-Hassan, A. (2016). Economics of land degradation in Central Asia. In *Economics of land degradation and improvement–A global assessment for sustainable development* (pp. 261-290): Springer, Cham.

Molle, F., Wester, P., & Hirsch, P. (2010). River basin closure: Processes, implications and responses. *Agricultural Water Management, 97*(4), 569-577.

Morgan, R. P. C. (2009). *Soil erosion and conservation*: John Wiley & Sons.

Mortimore, M. (2005). Dryland development: success stories from West Africa. *Environment: Science and Policy for Sustainable Development, 47*(1), 8-21.

Munns, R., & Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol., 59*, 651-681.

Nkonya, E., Mirzabaev, A., & Von Braun, J. (2016). *Economics of land degradation and improvement–a global assessment for sustainable development*: Springer Nature.

Northcote, K. H., & Srene, J. (1972). *Australian soils with saline and sodic properties*.

Omuto, C., Nachtergaele, F., & Rojas, R. V. (2013). *State of the Art Report on Global and regional Soil Information: Where are we? Where to go?* : Food and Agriculture Organization of the United Nations, Rome.

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature, 540*(7633), 418.

Prăvălie, R. (2016). Drylands extent and environmental issues. A global approach. *Earth-Science Reviews, 161*, 259-278.

Ravi, S., Breshears, D. D., Huxman, T. E., & D'Odorico, P. (2010). Land degradation in drylands: Interactions among hydrologic–aeolian erosion and vegetation dynamics. *Geomorphology, 116*(3-4), 236-245.

Reynolds, J. F., Smith, D. M. S., Lambin, E. F., Turner, B., Mortimore, M., Batterbury, S. P., et al. (2007). Global desertification: building a science for dryland development. *science, 316*(5826), 847-851.

Richards, L. A. (1954). *Diagnosis and improvement of saline and alkali soils. Handbook No. 60*: US Department of Agriculture, Washington, DC.

Roberts, A. J. (2013). Avian diets in a saline ecosystem: Great Salt Lake, Utah, USA. *Human–Wildlife Interactions, 7*(1), 15.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., et al. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences, 111*(9), 3268-3273.

Saito, H., Šimůnek, J., & Mohanty, B. P. (2006). Numerical analysis of coupled water, vapor, and heat transport in the vadose zone. *Vadose Zone Journal, 5*(2), 784-800.

Sakadevan, K., & Nguyen, M.-L. (2010). Extent, impact, and response to soil and water salinity in arid and semiarid regions. In *Advances in Agronomy* (Vol. 109, pp. 55-74): Elsevier.

Santos, R., Fernandes, L. S., Moura, J., Pereira, M., & Pacheco, F. (2014). The impact of climate change, human interference, scale and modeling uncertainties on the estimation of aquifer properties and river flow components. *Journal of hydrology, 519*, 1297-1314.

Schulz, S., Darehshouri, S., Hassanzadeh, E., Tajrishy, M., & Schüth, C. (2020). Climate change or irrigated agriculture–what drives the water level decline of Lake Urmia. *Scientific reports, 10*(1), 1-10.

Shadkam, S. (2017). Preserving Urmia Lake in a changing world: reconciling anthropogenic and climate drivers by hydrological modelling and policy assessment. In: Wageningen University, Netherlands.

Shahid, S. A., Abdelfattah, M. A., Omar, S. A., Harahsheh, H., Othman, Y., & Mahmoudi, H. (2010). *Mapping and Monitoring of Soil Salinization Remote Sensing, GIS, Modeling, Electromagnetic Induction and Conventional Methods–Case Studies.* Paper presented at the Proceedings of the International Conference on Soils and Groundwater Salinization in Arid Countries.

Shahid, S. A., Zaman, M., & Heng, L. (2018). Introduction to soil salinity, sodicity and diagnostics techniques. In *Guideline for salinity assessment, mitigation and adaptation using nuclear and related techniques* (pp. 1-42): Springer.

Shrivastava, P., & Kumar, R. (2015). Soil salinity: a serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi journal of biological sciences, 22*(2), 123-131.

Shukla, P., Skeg, J., Buendia, E. C., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D., et al. (2019). Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Sivakumar, M. V., & Stefanski, R. (2007). Climate and land degradation—an overview. In *Climate and land degradation* (pp. 105-135): Springer.

Soil Survey Staff. (2010). *Keys to soil taxonomy*: United States Department of Agriculture, Soil Conservation Service: Washington, DC.

Sterling, S. M., Ducharne, A., & Polcher, J. (2013). The impact of global land-cover change on the terrestrial water cycle. *Nature Climate Change, 3*(4), 385-390.

Suweis, S., Rinaldo, A., Van der Zee, S., Daly, E., Maritan, A., & Porporato, A. (2010). Stochastic modeling of soil salinity. *Geophysical Research Letters, 37*(7).

UN General Assembly. (2015). *Resolution adopted by the General Assembly on 19 September 2016*. A/RES/71/1, 3 October 2016 (The New York Declaration), 2015.

UNCCD. (1994). Elaboration of an international convention to combat desertification in countries experiencing serious drought and/or desertification, particularly in Africa. *General Assembly, United Nations*.

van der Esch, S., ten Brink, B., Stehfest, E., Bakkenes, M., Sewell, A., Bouwman, A., et al. (2017). Exploring future changes in land use and land condition and the impacts on food, water, climate change and biodiversity: scenarios for the UNCCD Global Land Outlook.

Wada, Y., Gleeson, T., & Esnault, L. (2014). Wedge approach to water stress. *Nature Geoscience, 7*(9), 615-617.

Wiggs, G. F., O'hara, S. L., Wegerdt, J., Van Der Meer, J., Small, I., & Hubbard, R. (2003). The dynamics and characteristics of aeolian dust in dryland Central Asia: possible impacts on human exposure and respiratory health in the Aral Sea basin. *Geographical Journal, 169*(2), 142-157.

Williams, W. D. (1993). Conservation of salt lakes. *Hydrobiologia, 267*(1-3), 291-306.

Williams, W. D. (1996). What future for saline lakes? *Environment: Science and Policy for Sustainable Development, 38*(9), 12-39.

Williams, W. D. (2002a). Environmental threats to salt lakes and the likely status of inland saline ecosystems in 2025. *Environmental conservation*, 154-167.

Williams, W. D. (2002b). Environmental threats to salt lakes and the likely status of inland saline ecosystems in 2025. *Environmental conservation, 29*(2), 154-167.

Wurtsbaugh, Wayne A, Craig Miller, Sarah E Null, R Justin DeRose, Peter Wilcock, Maura Hahnenberger, Frank Howe, and Johnnie Moore. 2017. 'Decline of the world's saline lakes', *Nature Geoscience*, 10: 816.

Chapter 2 Desiccation crisis of saline lakes: A new decision-support framework for building resilience to climate change

Amirhossein Hassani[a], Adisa Azapagic[a*], Paolo D'Odorico[b], Amir Keshmiri[c], and Nima Shokri[d*]

[a] Department of Chemical Engineering and Analytical Science, The University of Manchester, Sackville Street, Manchester M13 9PL, UK.
[b] Department of Environmental Science, Policy and Management, UC Berkeley, California, USA.
[c] Department of Mechanical, Aerospace and Civil Engineering, The University of Manchester, Pariser Building, Sackville Street, Manchester M13 9PL, UK.
[d] Hamburg University of Technology, Institute of Geo-Hydroinformatics, Am Schwarzenberg-Campus 3 (E), 21073 Hamburg, Germany

Science of the Total Environment, https://doi.org/10.1016/j.scitotenv.2019.134718
*Corresponding authors: adisa.azapagic@machester.ac.uk; nima.shokri@tuhh.de

Abstract

River flow reductions as a result of agricultural withdrawals and climate change are rapidly desiccating endorheic lakes, increasing their salinity and affecting the bio-diversity and human wellbeing in the surrounding areas. Here we present a new framework to guide eco-hydrological restoration of saline lakes and build their resilience to climate change by optimizing agricultural land use and related water withdrawals. The framework involves four steps: 1. selection of global circulation models for the basin under study; 2. establishment of a hydrological balance over the lake's area to estimate the amount of water required for its restoration; 3. water allocation modelling to determine the water available for restoration and allocation of the remaining water across different users in the lake's basin; and 4. basin-scale optimisation of land use and cropping patterns subject to water availability. We illustrated the general applicability of the framework through the case of the second largest (by volume) hyper-saline lake globally, Lake Urmia, which lost 96% of its volume in only 20 years, primarily as a result of upstream water withdrawals. Through the application of the framework, we estimated the amount of water needed to restore the lake, either fully or partially, and proposed a sustainable land use strategy, while protect farmers' income in the basin. Considering future climate change projections under two representative concentration pathways (RCP) 4.5 and 8.5, we found that an average annual surface inflow of 3,648 Mm^3 (~70% increase in RCP 4.5) and 3,692 Mm^3 (~73% increase in RCP 8.5) would be required to restore the lake by 2050, respectively. This would require the respective conversion of 95,600 ha and 133,687 ha of irrigated land to rain-fed cropland or grassland across the basin by 2050. The proposed framework can be used for building resilience to climate change and mitigating human-induced threats to other declining saline lakes.

Keywords: Ecosystem services; Lake restoration; Lake Urmia; Land use management; Optimal cropping patterns; Saline lakes.

Nomenclature

Symbol	Definition	Symbol	Definition
a, b	Constants in subsidy function	m_{pre}	Mass of precipitation over the lake area (kg)
A	Total area under cultivation (ha)	$m_{saltavailable}$	Available salt mass in the lake (kg)
a_c, b_c, c_c	Calibration parameters in the lake level-volume relation	m_{salt}	Change in the mass of lake water salt after Δt (kg)
AOT	Aerosol optical thickness	$m_{saltnew}$	Salt mass available in the lake after Δt (kg)
C_α, C_β	Constant parameters in Eq. 10	$m_{saltrivers}$	Salt mass added by rivers (kg)
C_a	Production cost (US\$ ha^{-1})	ω	Function used in Eq. 7
CC_x	Maximum canopy cover	p	Soil surface plastic pressure (Pa)
c_y	Dimensionless coefficient used in Eq. 11	P_c	Income received by farmers (US\$ t^{-1})
d_1	Lower saltating particle size (m)	$p(d_s)$	Soil particle size distribution
d_2	Upper saltating particle size (m)	$p_f(d_s)$	Fully disturbed particle size distribution
Δd_i	Particle-size bin interval length (m)	$p_m(d_s)$	Minimally disturbed particle size distribution
d_i	Particle-size bin mean diameter (m)	P_w	Price of irrigation water (US\$ m^{-3})
D_j	Mean diameter of *j*th mode (m)	Q	Total saltating particles flux (kg m^{-2} s^{-1})
d_s	Particle size (m)	$q(d_s)$	Horizontal sand flux of particle size d_s (kg m^{-2} s^{-1})
e_f	Freshwater pan evaporation (kg)	ρ_b	Bulk soil density (kg m^{-3})
e_s	Lake water pan evaporation (kg)	ρ_p	Particle density (kg m^{-3})
η_f	The total fraction of dust which can be released from unit soil mass	S	Crop subsidy (US\$ ha^{-1})
f	Dust fraction in soil	σ_p	Ratio of free dust to aggregated dust
F	Vertical dust flux (kg m^{-2} s^{-1})	σ_j	Standard deviation of the *j*th mode
$F(Y)$	Crop-field production function (t ha^{-1})	σ_m	The ratio between mass of impacting particle and mass ejected by bombardment
g	Acceleration due to gravity (m s^{-2})	TGM	Total gross margin (US\$)
i	Number of cultivated crops	u_*	Friction velocity (m s^{-1})
I	Number of particle-size bins	u_{*t}	Threshold friction velocity (m s^{-1})
J	Number of modes in calculation of fully/minimally particle size distribution	W	Available annual water in each river basin (m^3 ha^{-1}).
Δm	Change in the mass of lake brine after Δt (kg)	w_j	Weight of the *j*th mode
$m_{evaporation}$	Mass of evaporated water from the lake area (kg)	X	Area under cultivation by each crop (ha)
$m_{\inf}$	Mass of water entered to the lake by rivers (kg)	Y	Seasonal irrigation demand (mm)

2.1 Introduction

Saline inland water bodies are a recurrent landscape feature of mid-latitude and subtropical closed hydrological basins (Hammer 1986). With their current volume of 82,676 km^3, they account for 44% of the total lake water storage on earth (Messager et al. 2016a). These lakes play an important role in determining regional climate patterns, sustaining biotic productivity and diversity, maintaining environmental and human health, and providing recreational services, minerals, and other resources (Williams 1996; Hammer 1986). Therefore the dynamics of saline lakes are of great importance to a broad array of stakeholders.

Over the past decades, several anthropogenic and climatic drivers have disturbed the hydrological balance of many inland saline lakes, often resulting in complete desiccation (Wurtsbaugh et al. 2017). Excessive evaporation with respect to the natural inflow (i.e. negative water balance), resulting from unrestricted withdrawals of surface and sub-surface water in the upstream watersheds and global warming, have decreased the water levels of most large saline lakes to an alarming extent across the world. Saline lakes' basins are mostly located in regions categorised as arid and semi-arid climates (Williams 2002b). Some examples of the declining saline lakes around the world are provided in Figure 2-1. In most cases, unsustainable upstream water withdrawals have been a primary factor in the lake shrinkage (Williams 1996). Lake Urmia, the Aral Sea, and Owens Lake are some examples of the shrinkage as a result of anthropogenic interventions. The Aral Sea faced a reduced area of 74% and volume of 90% as a result of basin-wide irrigation expansion (Micklin 2007). Similarly, water use for human needs led to an average ~1.74 km^3 yr^{-1} reduction in water inflow to Lake Urmia from 1995 to 2010, resulting in ~86% decrease in the total lake volume (Chaudhari et al. 2018). Diversion of the supplying streams to meet agronomic and urban water demands desiccated California's Owens Lake completely by 1940 (Wurtsbaugh et al. 2017).

Comprehensive analyses of the trade-offs between ecological and economic benefits of the saline lakes are needed to identify sustainable options for managing these lakes, with the goal of restoring them while enabling economic activities in the surrounding regions (Wurtsbaugh et al. 2017). In addition to the loss of the above-mentioned ecosystem services, shoreline recession in theses saline lakes leaves salt-rich playas behind, which remain exposed to wind shear and are susceptible to aeolian transport of salt-rich dust. Snow melting in the surrounding mountains, soil salinization, vegetation degradation, poor air quality, and augmented risk of morbidity from respiratory diseases are some of the direct irreversible consequences of dust emissions from these desiccating lakes (Abuduwaili et al. 2010).

In response to concerns regarding the shrinking of saline lakes and the consequent impacts on the local environment, many ongoing national and sub-national efforts are aimed at increasing flows in upstream rivers. The endoreic nature of these basins makes them more vulnerable to changes in inflow regimes (Williams 2002b). Many approaches have been adopted by watershed managers to address the issue of saline lakes' shrinkage. One such approach is partial restoration of lakes by constructing physical barriers in the desiccated part, similar to the method used for the Aral Sea (Micklin 2007). Another example is California's Mono Lake where the water inflows were increased by limiting withdrawals in the lake's upstream tributaries (Ryan 2015). Multi-million to multi-billion-dollar projects for transferring water from the adjacent basins have been also considered as a method to preserve saline lakes, and examples include Lake Urmia, Dead Sea (Deatrick 2016), and Utah's Great Salt Lake (Miller 1987). In most cases, the required flow for restoration and preserving saline lakes can be provided by reducing water use, especially in the agricultural sector, although the projected climate change can help or exacerbate the management of the water demands (Wurtsbaugh et al. 2017). Despite utilizing various approaches for controlling the water withdrawals across saline-lake watersheds, a quantitative analysis of how agricultural land use would need to change to reduce water use, especially in face of future climatic uncertainties, is rare. Basin-wide technical and institutional improvements in the agricultural sector, reducing irrigated area, and cultivating less water-intensive crops can substantially conserve the water for restoring and preserving saline lakes. Hence, to assist policy-makers and evaluate the priority for investment in land use and cropping pattern change, a comprehensive analysis of the land, climate, and the economy nexus of saline lakes' basins is needed.

To address this need, we propose a framework to guide evaluation of possible solutions based on land use strategies that can facilitate restoration of saline lakes in the short- to long-terms (2030 - 2050). The framework considers hydrological models combined with a salt balance to determine the water inflows that would be necessary to restore and keep the lake at the recommended level to maintain its ecosystem services. It also enables consideration of alternative cropping patterns to facilitate rehabilitation of lakes, especially in the context of projected climate change, while striving to protect farmers' income. As a results, it is possible to determine possible future hydrological, agronomic, and socio-economic conditions in the lake's basin associated with an optimal land use change strategy, aiming to reduce water consumption in the basin and restore the lake. Because such land-based interventions have the potential to increase greenhouse gas (GHG) emissions (Houghton 2003; Smith 2008), the framework also quantifies the emissions related to the recommended land use change schemes.

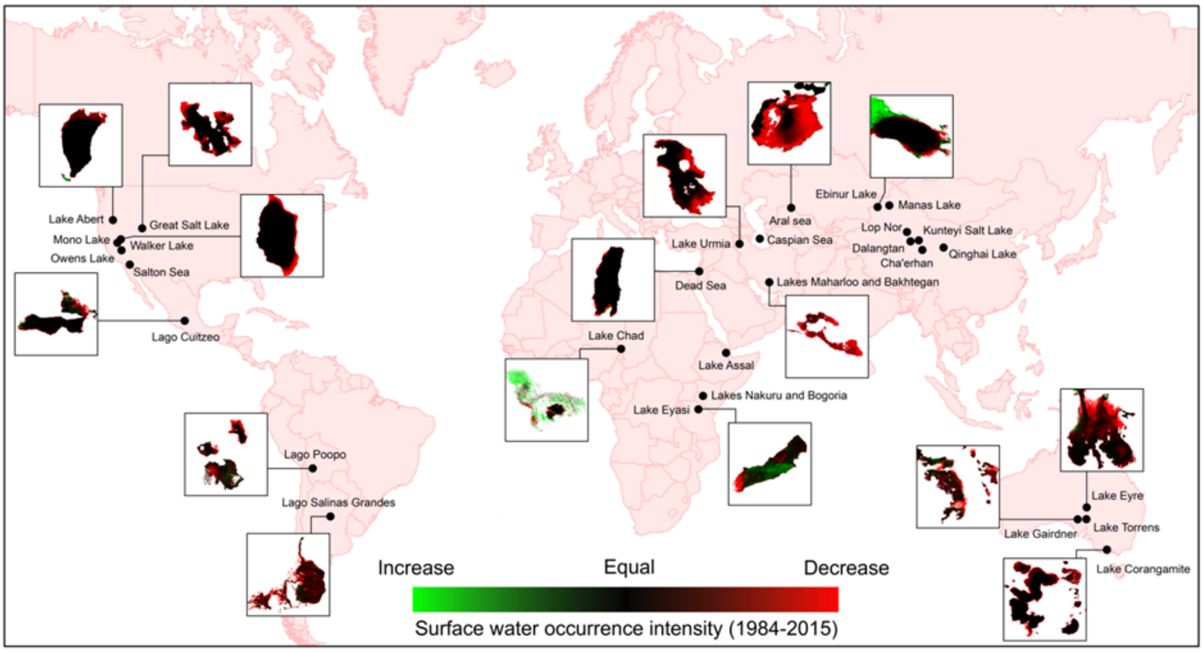


Figure 2-1: Examples of the observed decrease in surface water area in major saline lakes around the world over the period 1984 - 2015 (Pekel et al. 2016).

Although the lakes will differ and the specific analysis of each lake and its basin will require specific data, the proposed framework is generic and can be applied to similar saline-lake basins located in arid and semi-arid areas with unsustainable water withdrawals as a primary reason for decreased water flows into the lake. Some critical examples of these lakes are The Aral Sea, Great Salt Lake, Lake Urmia, Lake Abert, Walker Lake, Lake Poopó, and Owens Lake. If the basin-wide demographic, hydrological, climatic, and agronomic data for each river supplying the saline lake are available, the proposed framework can be applied for projecting the required adaptations in the agricultural land use. However, the proposed framework cannot be applied to lakes in arid regions where droughts and excess of evaporation relative to precipitation are the major causes of desiccation.

To mitigate the adverse consequences of lake recession in shorter time spans, the framework also enables consideration of partial restoration by identifying the part(s) of the lake that should be prioritised for restoration. This is based on the analysis of dust emissions from the lake bed since the dispersion of wind-blown saline dust from the playa, resulting from the desiccation of saline lake sediments, is the major concern for environmental and human health in nearby regions (Gillette et al. 1997). Here we prioritise the partial restoration option with the best potential for dust emission mitigation over other objectives, such as recreation or extraction of minerals. We quantify the potential for vertical emissions of saline dust from the lake bed in response to wind erosion to identify the zones with the higher priority for restoration. The framework is detailed in the next section, followed by an illustrative application in section 2.3.

2.2 Materials and methods

As illustrated in Figure 2-2, the proposed framework comprises four steps. Step 1 involves projection of future climate conditions, depending on the targeted time horizon for restoration of the lake. This is commonly carried out by using outputs of Global Circulation Models (GCMs). However, the variety of GCMs with different initial conditions and physics necessitates the application of multi-GCM ensemble for the basin under study to capture the uncertainties involved in the predictions of climate change. Therefore, this is taken into account in the framework as discussed in section 2.3.1.

Step 2 requires establishment of a historical salt and water balance over the lake's area to relate limnological and hydrological processes of a saline lake to climatic and other long-term meteorological drivers. These hydrological processes can include the surface and sub-surface inflow or sediment transport dynamics. Among the climatic parameters related to the water balance of a saline lake, a reliable estimation of the evaporation rate is a challenge (Zilberman et al. 2017), especially if measured data are not available. At the same conditions, the evaporation from saline water bodies is different from freshwaters due to the different surface activity of brine (Lensky et al. 2005). Dissolved salts reduce the free energy of the water molecules, which influences the saturated vapour pressure depending on the salt concentration (Shokri-Kuehni et al. 2017a; Shokri-Kuehni et al. 2017b). This in turn modifies the evaporative fluxes. Therefore, relationships commonly used for estimations of the evaporation flux from freshwater cannot be applied for the case of saline lakes. Instead, estimation of the evaporation rate can be resolved by direct measurements, as in the present study and in Mor et al. (2018), or implementation of long-term mass balance models — e.g. Winter et al. (2003) and Mohammed et al. (2012).

In step 3, the calibrated hydrological balance over the lake's area against system stressors provides the basis for estimation of the required inflow for restoration of the lake to the targeted water level for various time horizons. This will depend on how restoration is defined, what management objectives are preferred, and the constraints included. The restoration target level can be defined as the level that returns the degraded aquatic system to the conditions that are ecologically productive, protective, or aesthetically pleasing (Hobbs et al. 1996). It should be noted that the ecological target level for lake restoration is not a singular level and, in most cases, the ecological aspects are improved along a continuum of salinities and elevations. A vulnerability analysis of the ecological services provided by saline lakes to the water-level decline, descriptions of factors to which the saline aquatic system is sensitive, and quantification of trade-offs among various restoration objectives, can be the key elements in prioritizing and adopting a single elevation for the lake water. After the estimation of the required water, another model is needed to estimate the available water and allocate the remaining water (not used for restoration) among all stakeholders at the river basin scale. The water allocation approach needs to consider the future climatic, hydrological, agronomic, demographic, and other drivers in a saline lake basin.

Finally, step 4 aims at developing and evaluating alternative sustainable land use and cropping patterns to reduce water usage in the agricultural sector that is needed for increasing the lake's water level. The focus is on agricultural usage because in most cases unsustainable development of agronomic activities in saline lakes' basins has been the major reason for the decline of lakes, including Lake Urmia (AghaKouchak et al. 2015; Hassanzadeh et al. 2012; Chaudhari et al. 2018; Ghale et al. 2018). We define a land use strategy as sustainable if it allows the inflow of required water to the lake over relevant management time scales while meeting the economic, social, and environmental needs of the stakeholders in the basins. Water allocated to the agricultural sector (in step 3) becomes one of the inputs into a farm-economic optimisation model to obtain the optimal cropping pattern at the river basin scale, subject to available irrigation water and land resources needed to restore the lake by the targeted time horizon. Estimation of the cultivated crop yields as a function of irrigation water in the supplying river basins provides another major input for the farm-economic optimisation model (Figure 2-2).

The system is then optimised to maximise income at the farm level for the new land use and cropping patterns. Other optimisation objectives can also be defined, based on stakeholders' interests and data availability. As new cropping patterns involve land use change, it is important to estimate the resulting change in GHG emissions, which is also included in the framework. The optimal agricultural water management solutions, identified through the optimisation, enable decision- and policy-makers to evaluate how potential land use schemes may perform, explore the interconnections between different system components, and identify the key factors affecting the restoration strategies. They can also consider various challenges, such as ensuring that farmers comply with recommended solutions, maintaining equity among stakeholders, behavioural and institutional barriers, and financing needed for the proposed land management policy.

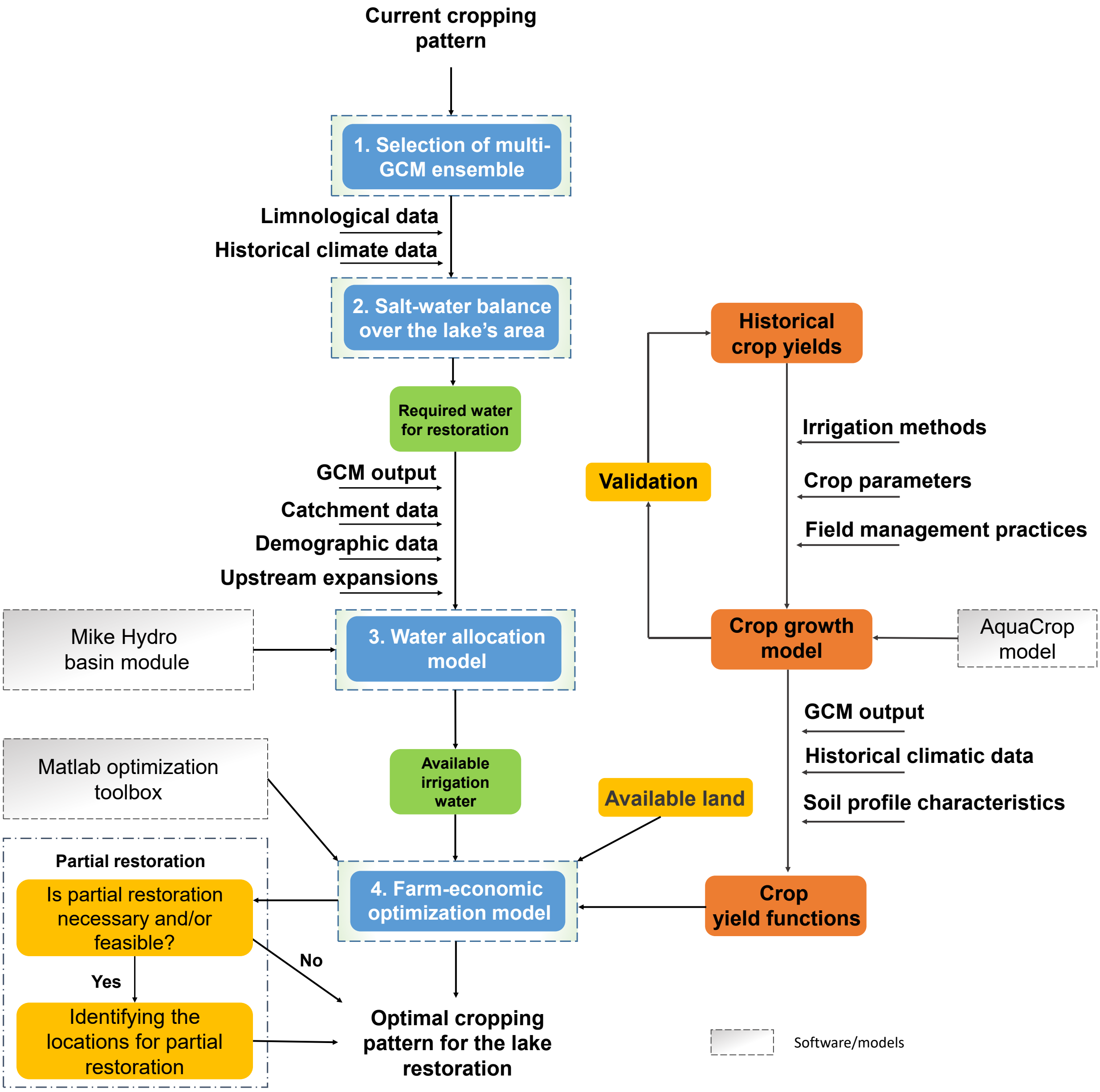


Figure 2-2: The proposed framework for restoration of saline lakes (GCM: Global Circulation Model).

In case that coping with these challenges is not possible in the short term, consideration can be given to feasible partial restoration scenarios in parallel with the land use optimisation process. Depending on available resources and services provided by the saline lake, locations with the higher priority for restoration can be identified. The decision-makers may also want to change the constraints or optimisation objectives. Therefore, the proposed framework can be used as an iterative decision-support tool which can be updated as new information or alternative watershed resource management frameworks come to light.

The applicability of the proposed framework is illustrated in the next section through the case of the second largest (by volume) hyper-saline lake in the world – Lake Urmia in Iran – which is at risk of complete desiccation. Strategies for both partial and complete restoration are considered in the short- (2030) and long-terms (2050), respectively.

2.3 Application of the framework: The case of Lake Urmia

The Urmia basin plays a pivotal role in Iranian national food supply. Nearly 10% of the country's agricultural area is located in the basin, producing about 6 Mt of crops annually. With an average annual precipitation close to 400 mm and average potential evapotranspiration of 530-680 mm, the basin's climate is classified as cold semi-arid (Kottek et al. 2006). The lake is located in a geological sink called graben where volcanic rocks caused by historical eruptions are widely distributed (JICA 2016). The lake is terminal and surrounded by mountainous areas with elevations ranging from 1,270 to 4,000 m (above the mean sea level). It is supplied by 12 major streams flowing into the lake, with an average annual run-off ratio of 0.24 (JICA 2016). The Talkhe and Zarineh rivers with the catchment areas of 12,717 and 11,838 km^2 compose 24.6% and 22.9% of the whole lake's basin area, respectively, while ~ 40% of the annual inflow to the lake is supplied by the Zarineh and Simineh rivers.

Construction of more than 44 reservoirs along the lake's tributaries (with the total storage capacity of ~1.413 km^3) and increasing the authorised groundwater withdrawals from 0.25 km^3 in 1980 to 1.6 km^3 in 2014, along with intermittent drought periods, doomed Lake Urmia to almost complete desiccation (Fathian et al. 2015; Shadkam et al. 2016; Jalili et al. 2016), which is unprecedented in the last 4,000 years (Kelts et al. 1986). As a result, the lake lost 80% of its area and 96% of its volume in just 20 years, declining by 0.4 m yr^{-1} in the water level (Figure 2-3). The growing in-field, modelling, and remote sensing evidence suggests that, although climate change and decreasing precipitation have exacerbated the decline of the water levels, intensified water withdrawal is one of the major reasons for the shrinkage of the lake (AghaKouchak et al. 2015; Hassanzadeh et al. 2012; Chaudhari et al. 2018; Ghale et al. 2018).

As mentioned earlier, airborne saline particles emitted from the desiccated parts of the lake, are a major concern for environmental and human health. Long-term fluctuations in hydrological and climatic parameters prevent an effective estimation of spatial distribution of the adjacent demographic and industrial centres exposed to the emissions. Hence, we evaluated the geospatial dispersion of PM_{10} (particulate matter $\leq$ 10um) emitted from the lake bed (section 2.3.4.2.1) in the context of a five-hour storm that struck the Urmia basin in March 2018 at the mean wind speed of 44 km h^{-1} (gusts up to 70 km h^{-1}). The simulations were based on the assumption that the lake bed moisture was < 5%. The HYSPLIT_4 (Hybrid Single-Particle Lagrangian Integrated Trajectory) model (Draxler 1999; Rolph et al. 2017) was used to simulate the dispersion and deposition trajectories (particle speed deposition = 0.001 m s^{-1}). The emission rate was set equal to the vertical dust flux (F) as a function of wind speed, calculated from the vertical dust parameterizations explained in section 2.3.4.2.2. The spatial scattering range of the emitted saline fine particles as a result of the storm (Figure 2-4, a to c) suggests a very significant influence of dust from the desiccated parts of Lake Urmia on air concentrations of PM_{10} at the ground level (0 to 100 m above the surface).

Annual depositions of 13.03 Mt of sediments, 1 Mt of salt, and 1,076.45 Mt of herbicides, pesticides and fertilisers (WRMC 2007), conveyed by inflowing rivers, dramatically exacerbate the environmental perils of wind-blown particulates. The composition of the deposited particles is of particular concern for human and animal health, as well as for vegetation. For example, chloride-containing airborne particles can lead to a range of metabolic and reproductive alterations to vegetation and respiratory-related threats for humans and livestock (McCune 1991). Our analysis of the elemental composition of 49 near-surface sediments at the lake bed (initial geochemical data after Alipour et al. (2018)) indicated significant contamination of deposits by arsenic and antimony and moderate contamination by rubidium and strontium (enrichment factors = 13.4, 11.9, 4.43, and 4.06, respectively).

Despite the emergence of the lake's recession between 2006 and 2014, the monetary incentives in agriculture and other sectors drawing on the water in the lake's basin hindered its conservation (Madani 2014). In 2014, the near-complete desiccation, with the water level declining to 1,270.1 m, received significant scientific and public attention (AghaKouchak et al. 2015; Garousi et al. 2013). As a consequence, the Iranian government and the UN issued a US$ 1.3 billion plan for rehabilitation of the lake and surrounding wetlands which saved the lake from complete desiccation. However, in the first four years of this ten-year restoration plan, the lake level increased by only 0.6 m to 1,270.7 m, corresponding to an increase in water volume of 525 Mm^3. This falls well below the targeted ecological level established by the Lake Urmia Restoration Committee of 1,274.1 m. This ecological target reflects the lake level and volume at which salinity is expected to fall below 240 g L^{-1} NaCl. This salinity represents the maximum tolerance level for brine shrimp, the only fauna at the top of the ecological pyramid of the lake (Abbaspour et al. 2007). Our calculations showed that an additional water volume of 11,850 Mm^3 is required to achieve the target water level of 1274.1 m (based on the lake level-area-volume relationships; see Appendix 1 for details).

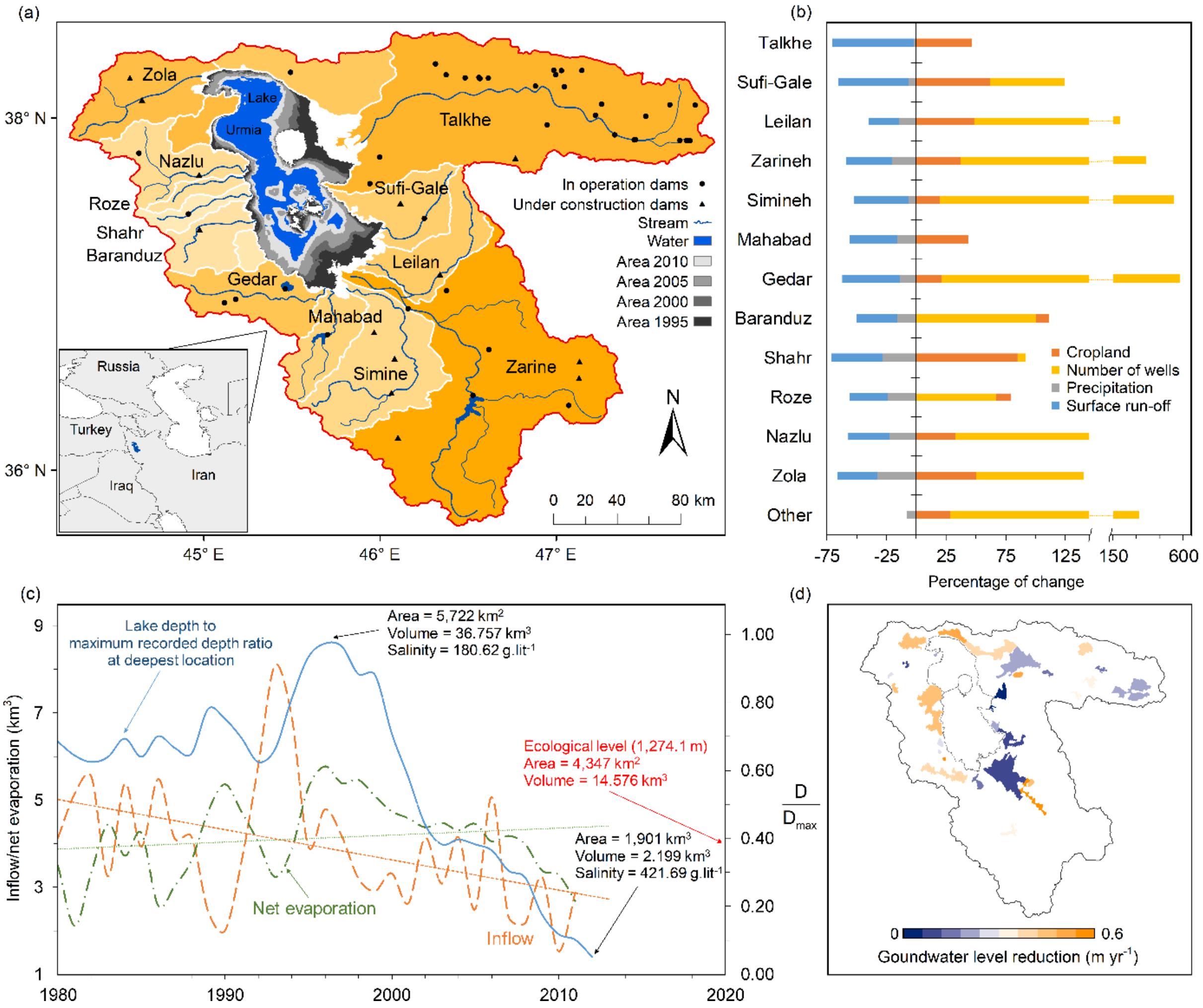


Figure 2-3: a: Lake Urmia's watershed and the river basins of its tributaries. **b**: Increase in the number of wells, decrease in precipitation, reduction in surface run-off, and increase in the area under cultivation in each river basin. The increase in the number of wells is obtained through comparison of two periods: before 1985 and from 1985 to 2001. For the surface run-off, average river discharges in the 1975 - 1995 and 1995 - 2011 periods are compared. The decrease in the annual precipitation in each basin is acquired through comparison of the 1980s and 2000s. Land use change is for the period 1987 - 2007. **c**: Annual water stocks and flows determining Lake Urmia's hydrological balance. For the period between 1980 and 1995, the difference between the inflow to the lake and net evaporation (evaporation minus precipitation on the lake's surface) was negligible; since 1995, the gap between the inflow and outflow from the lake has grown steadily. **d**: Annual decrease in the groundwater table level between 2004 and 2014.

To meet the water needs for short-term recovery of the lake (ten years), trans-basin water diversions were implemented by government as a part of the rehabilitation plan. One of these is a US$ 100 million tunnel (35.7 km) which will be used to transfer 0.64 km^3 of water per year from the adjacent basin (the lower Zab, one of the Tigris's tributaries). However, desertification and degradation of the Tigris's downstream marshes as a result of this water diversion, dam construction on other transboundary tributaries, and plans by Turkey and Iraq for impoundment of 34 km^3 of Tigris's water (Altinbilek 2004) cast serious doubts on the long-term efficacy of the solutions related to inter-basin diversion. Moreover, occurrence of 33 dust events per year in south western Iran, with hourly PM_{10} concentrations above 500 µg m^{-3} might be attributed to a 65% decrease in Tigris's historical discharge and desiccation of its deltaic wetlands compared to the historical values in the period 1931 - 1952 (Rahi et al. 2018). The following sections detail the framework steps as applied to the case study.

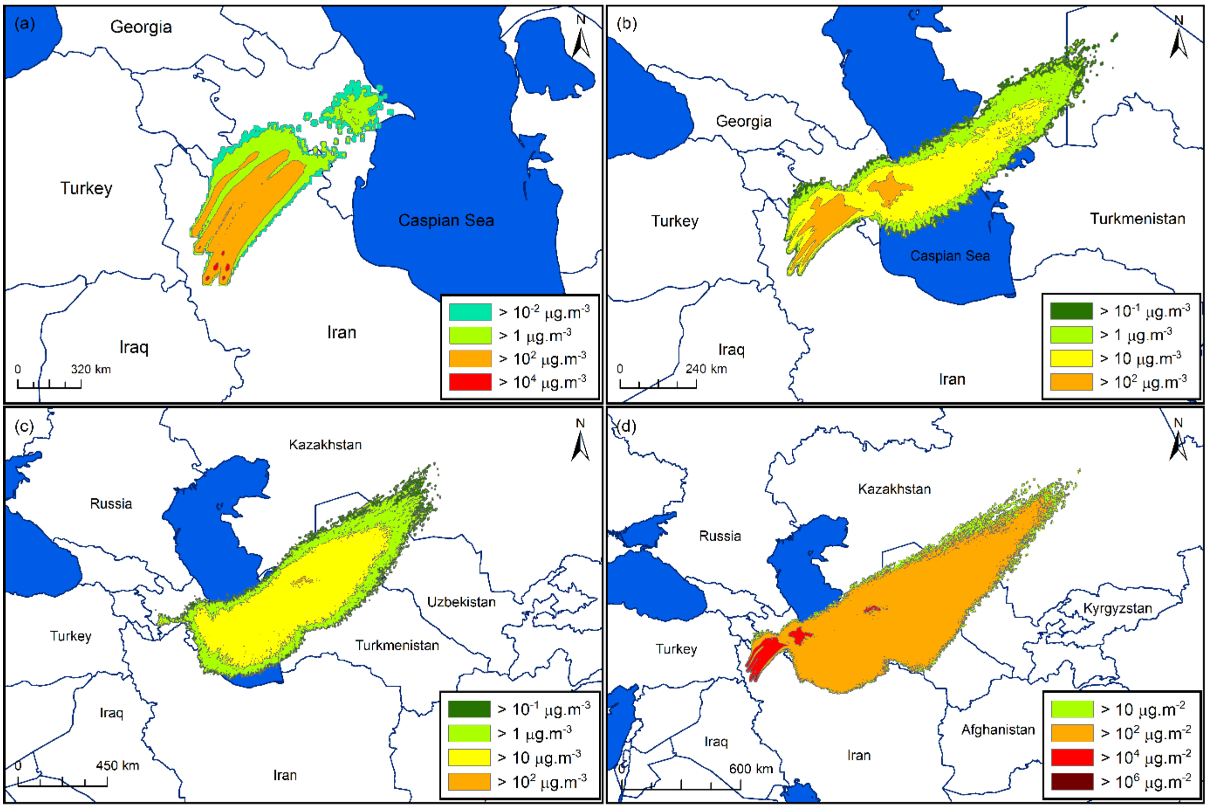


Figure 2-4: **a**, **b**, and **c**: Average concentrations of PM_{10} (particulate matter ≤ 10 µm) between the surface and 100 meters above ground level after 6, 12, and 18 hours from the beginning of the storm, respectively. **d**: Aggregate PM_{10} disposition rate 24 hours after the start of the simulation. The model simulates the PM_{10} dispersion originated from the Lake Urmia's saline playa (in a hypothesised case of the lake's desiccation) during a dust storm on 24/3/2018 (14:00 UTC). A mean of 27 ensemble members (offset is one meteorological grid point in the horizontal and 0.01 sigma units in the vertical direction) was calculated for the 18-hour dispersion simulation. For further details, see Appendix 1).

2.3.1 Step 1: Selection of multi-GCM ensemble

Required climatic parameters for projecting the future hydrological and agronomic conditions in the Urmia basin for the period 2020 - 2050 were obtained using an ensemble of GCMs from the IPCC's Fifth Assessment Report (Emori et al. 2016). The outputs of the 16 GCMs under two Representative Concentration Pathways (RCP) forcing scenarios (4.5 and 8.5) were considered to capture the uncertainties in the projected climates. Including different realizations of the GCMs, a total of 34 and 25 ensemble members were used for RCP 4.5 and 8.5, respectively. Details of the GCMs and their used realizations are provided in Table 2-1. They were statistically bias-corrected over the reference period of 1960 - 1995, i.e. before the lake's desiccation. The bias-corrections were based on Eq. 2 in Hawkins et al. (2013) using high-resolution gridded time-series dataset for observed climatic parameters at 0.5° resolution, based on Harris et al. (2014), version CRU TS 4.02. This method was selected as it is computationally more efficient than more advanced techniques because of the large number of ensemble members.

Table 2-1: Global Circulation Models (GCMs) used in this study for projecting future (2020 - 2050) hydrological and climatic conditions in the Urmia basin.

GCM	Used realizations under RCP 4.5[a]	Used realizations under RCP 8.5[a]	Spatial resolution (latitude × longitude; degrees)	Centre (Country)
ACCESS1-0	1	1	1.25 × 1.87	CSIRO-BOM (Australia)
CNRM-CM5	1	1	1.40 × 1.40	CNRM-CERFACS (France)
GFDL-CM3	1	1	2 × 2.5	NOAA (USA)
GFDL-ESM2G	1	1	2.02 × 2	NOAA (USA)
GFDL-ESM2M	1	1	2.02 × 2.5	NOAA (USA)
GISS-E2-H	1,2,3,4,5	1,2	2 × 2.5	NASA (USA)
GISS-E2-R	1,2,3,4,5	1,2	2 × 2.5	NASA (USA)
GISS-E2-R-CC	1	1	2 × 2.5	NASA (USA)
HadGEM2-CC	1	1	1.25 × 1.87	MOHC (UK)
HadGEM2-ES	1,2,3,4	2,3,4	1.25 × 1.87	INPE (Brazil)
IPSL-CM5A-LR	1,2,3,4	1,2,3,4	1.89 × 3.75	IPSL (France)
IPSL-CM5A-MR	1	1	1.26 × 2.5	IPSL (France)
IPSL-CM5B-LR	1	1	1.89 × 3.75	IPS (France)
MPI-ESM-LR	1,2,3	1,2,3	1.86 × 1.87	MPI (Germany)
MPI-ESM-MR	1,2,3	1	1.86 × 1.87	MPI (Germany)
NorESM1-M	1	1	1.89 × 2.5	NCC, NMI (Norway)

[a] Realization number indicates the initial conditions of the ensemble member.

2.3.2 Step 2: Salt-water balance over the lake's area

To plan for a sustainable restoration of a lake, the primary question is: how much water is required annually by each supplying tributary to restore the lake in the face of future hydrological, agronomic, and climatic conditions? To estimate the required inflow for the Lake Urmia, we established a hydrological model over the lake, combined with a salt balance. For simplicity, here we adhered to the recommended target level adopted by the Restoration Committee (1274.1 m) because it is expected that above this water level the lake's water salinity would fall below the maximum salinity that could be tolerated by brine shrimp (240 g L^{-1} NaCl).

Unlike previous analyses of the water balance over the lake and its tributaries' watershed (JICA 2016), here we accounted for the effect of projected climate change. Moreover, most of the existing models did not consider the effect of dissolved salt mass on the water balance and the variations in the depth of the lake, which were accounted for in the present study. The lake's hydrological balance was used to estimate the total annual and sub-surface inflow to the lake as well as the salt precipitation rate. The water needed for restoration of the lake under different scenarios (partial or complete restoration) was calculated based on this reconstruction of the hydrological balance. The hydrological model was first calibrated against the climatic parameters and the surface water input in the reference period, here taken to cover the period from 1996 to 2010 due to limited data availability. Then, using the calibrated model, the amount of water required for rehabilitation to the target ecological water level for different time and climate scenarios was estimated. This is discussed in more detail next.

2.3.2.1 Water and salt mass balance in the reference period (1996 - 2010)

In this step, the lake's mass balance, including both the mass of water and dissolved salt, was established first. Subsequently, the meteorological and bathymetry data over the period were implemented to calibrate the model parameters. The change in the mass of the brine (m) after a period Δt was determined as follows (Lensky et al. 2005):

$$\frac{\Delta m}{\Delta t} = -m_{evaporation} - m_{salt} + m_{\inf} + m_{pre} \tag{1}$$

In eq. (1), $m_{evaporation}$ is the mass of evaporated water according to the lake's area, m_{salt} is the change in the mass of salt available in the water, m_{inf} is the amount of water that enters the lake during a particular time period, and m_{pre} is the mass of precipitation over the lake's area. The change in the mass of salt, m_{salt}, can be positive or negative, depending on the deposition or dissolution of salt in water. It was calculated as:

$$m_{salt} = m_{saltavailable} + m_{saltrivers} - m_{saltnew} \tag{2}$$

where $m_{saltavailabe}$ and $m_{saltnew}$ represent respectively the total amount of salt present in the lake at the beginning and the end of a particular time span and can be obtained directly by measuring the concentrations of salt in the lake; $m_{saltrivers}$ is the amount of salt added to the lake by its rivers (Table 2-2). If m_{salt} is positive in a particular year, it means that the salt is precipitated on the lake bed and vice versa. The bathymetry data from 2010 were used to calculate the area, level, and volume of the lake, using ArcGIS surface volume tool. The volume of salt added to or removed from the lake at each time (assuming the salt deposition is related to the halite whose density is 2,160 kg m^{-3}) was taken into account to update the estimated area, level, and volume of the lake. However, the lake bed surface area (i.e. 3D surface) was assumed to be constant. The observed precipitation over the lake's region and its corresponding area during 1996 - 2010 was adopted from JICA (2016).

No data were available on the evaporation rates from the lake's surface. Therefore, the annual rate of evaporation was estimated based on pan evaporation data. While it is not possible to use pan evaporation values directly to calculate the evaporation from open water bodies like a lake, a correction factor is typically used, here assumed at 0.77 (Darvishi 2014). To account for the presence of dissolved salts, we applied two different methods for correction of the freshwater pan evaporation data. In both methods, a time series of pan evaporation between 1996 and 2010 was applied, including the average of annual freshwater pan evaporation measured experimentally at ten different stations around the lake and the only station within the lake.

Table 2-2: Salinity and average mass of salt added to the lake by different rivers during 1996 - 2010.

River	Average annual inflow to the lake (Mm^3)	Total dissolved solids ($mg\ L^{-1}$)	Average amount of added salt ($t\ yr^{-1}$)
Barandooz chay	174.5	288.282	50,322.89
Roze chay	25.8	437.526	11,294.58
Gedar chay	202.8	324.738	65,882.62
Nazlu chay	137.4	264.865	36,393.63
Shahr chay	51.1	403.800	20,673.75
Mahabad chay	118.1	436.872	51,623.41
Simineh rood	296.7	260.304	77,237.22
Zarineh rood	879.9	302.022	265,770.50
Zola chay	27.8	662.304	18,440.75
Sinikh chay	17.3	357.956	6,218.88
Ghale chay	27.5	339.060	9,348.78
Azar shahr	22.6	145.000	3,279.03
Mardoogh chay	60.4	250.000	15,111.33
Lilan chay	42.7	287.850	12,300.79
Javan chay	7.9	250.000	1,988.50
Soofi chay	30.4	286.851	8,755.05
Aji chay	166.4	8,863.020	1,475,385.57
Surrounding plains	316.4	150.000	47,467.05

In the first method, the lake water pan evaporation (e_S) was used as the annual evaporation from the lake's surface area. In the second method, saline water evaporation rate from the lake was assumed to be constant at 1,200 mm yr^{-1} (JICA, 2016). This value was calculated by establishing a water balance over the lake's area as:

Water balance = (Surface water inflow) + (Groundwater inflow: Groundwater recharge + Water from groundwater storage) – (Base flows of the rivers) + (Precipitation over the lake) – (Evaporation from the lake) – (Water uptake from surface water and groundwater).

Ten different cases of annual evaporation rates from the lake's surface were used for the annual water balance calculation. They ranged from 1,000 to 1,900 mm yr^{-1} with a 100 mm yr^{-1} interval, referring to the average annual freshwater pan evaporation value at stations adjacent to the lake, equal to 1,611 mm — obtained via the Thiessen method (Thiessen, 1911). The case with the evaporation rate of 1,200 mm yr^{-1} showed the highest correlation with the annual average lake water level and accumulated annual water balance. Using the total precipitation over the lake and the above two methods for estimation of the evaporation rate, net annual inflows to the lake were calculated as the sum of surface and sub-surface inflows.

Two other methods were also applied to calculate the annual evaporation from the lake (see Table 2-3): 1. correction of the evaporation values obtained in the first method described above by multiplying with the lake-effect correction factor (0.77); and 2. modification of the saline water pan evaporation values using the lake-effect correction factor. However, the calculated total annual inflow into the lake using these methods was not acceptable since the estimated values for the total inflow to the lake were lower than the total surface entry, measured by the discharge gauges at the periphery of the lake (~2,600 Mm^3 yr^{-1}).

Table 2-3: Freshwater and brine evaporation before and after application of salt and lake constants correction.

Year	Average of ten stations (freshwater) (mm)	Golmankhan station (freshwater) (mm)	Golmankhan station (saline) (mm)	Saline to freshwater ratio	Evaporated water from the lake calculated by JICA (Mm^3)	Average of ten stations modified for salt presence × saline-to-freshwater ratio (mm)	Golmankhan station modification (×0.77) (mm)	Average of ten stations (freshwater) modified for salt presence × 0.77 (mm)
1996	1,424.7	1,385.1	1,191.5	0.86	6,864.5	1,225.5	917.4	943.7
1997	1,370.9	1,181.7	918.9	0.77	6,814.8	1,066.0	707.5	820.8
1998	1,473.3	1,386.0	1,098.3	0.79	6,686.9	1,167.4	845.6	898.9
1999	1,417.2	1,168.2	1,116.0	0.95	6,311.6	1,353.8	859.3	1,042.4
2000	1,503.7	1,385.9	997.7	0.71	5,900.7	1,082.5	768.2	833.5
2001	1,615.4	1,489.3	1,391.1	0.93	5,476.6	1,508.9	1,071.1	1,161.8
2002	1,505.7	1,381.9	1,115.6	0.80	5,266.2	1,215.5	859.0	935.9
2003	1,406.0	1,557.3	1,153.6	0.74	5,241.6	1,041.5	888.2	801.9
2004	1,433.9	1,433.5	1,248.6	0.87	5,319.9	1,248.9	961.4	961.7
2005	1,515.4	1,338.8	1,205.7	0.90	5,138.9	1,364.7	928.3	1,050.8
2006	1,484.5	1,267.6	1,181.4	0.93	5,108.5	1,383.5	909.6	1,065.3
2007	1,383.0	1,164.3	1,115.9	0.95	4,929.1	1,325.5	859.2	1,020.6
2008	1,636.7	1,687.0	1,452.7	0.86	4,565.0	1,409.4	1,118.5	1,085.2
2009	1,378.9	1,379.2	801.8	0.58	4,409.4	801.6	617.3	617.2
2010	1,430.0	1,265.1	1,070.9	0.84	4,069.6	1,210.5	824.5	932.1
Average	1,465.3	1,364.7	1,137.3	0.82		1,205.5	863.8	928.2

Although the aforementioned mass balance enabled us to estimate the total inflow and the under-surface water gain, there was still a need for updated relationships between the lake's level, area, and volume, which also consider the salt precipitated on the lake bed to determine the required water for restoration. We initially used the bathymetry data from the year 2010 to predict the lake's area and volume as a function of the lake's level (Appendix 1) and also the lake's level as a function of its volume, i.e. $Level = a_c \times (Volume)^{b_c} + c_c$. The constant parameters (a_c, b_c, and c_c) in this relationship were chosen as the calibrating parameters. The total annual inflow results obtained through the combined water and salt balances enabled us to calibrate these relationships. These sets of fitted functions and the calibrated lake level-volume relationships were then used to predict the required water for the restoration of the lake between 2020 and 2050.

We used Monte Carlo simulations to address the issue of the equi-finality and parameter uncertainty in the calibration process. We investigated the calibration of the three parameters (a_c, b_c, and c_c) ranging from 2.189×10^{-5} to 5.189×10^{-5}, 0.50 to 0.52, and 1267 to 1269, respectively. A 1,000 combinations of the calibrating parameters were derived employing uniform sampling strategy from each parameter within the above range. We then estimated the annual total inflow between 1999 and 2010 using the first evaporation method (preferred because the second method assumes a constant annual evaporation rate), precipitation data, and the water balance over the lake's area. We finally compared these results with the total inflow calculated for the same period by considering both the water and salt balances. A coefficient of determination (R^2) between the two computed ranges of the total annual inflow equal was set at 0.6 as an acceptable threshold for choosing a set of calibration parameters.

2.3.2.2 Mass balance in the period 2020 - 2050

The calibrated level-volume relationship was then used to estimate the required annual surface water inflows for complete restoration of the lake to the target water level (1274.1 m) by 2050. For the whole lake, the calculations of the amount of salt added to the lake by its rivers demonstrated that the volume of added salt was negligible compared to the volume of lake: the average total volume of added salt was close to 1 $Mm^3\ yr^{-1}$, compared to the surface water inflow of 2,606.5 $Mm^3\ yr^{-1}$. Hence, in the established hydrological cycle between 2020 and 2050, the salt balance was omitted. This was also because the amount of salt which dissolves from the lake bed into the water cannot be estimated precisely. Owing to a large difference between historical observations and the model output, bias-corrected precipitation and evaporation rates projected by the GCMs were utilised to estimate the evaporation and precipitation over the lake's corresponding area in each year in the period. The rainfall and evaporation time series from each GCM were bias-corrected as mentioned earlier (based on

Hawkins et al. (2013) method), using the observed precipitation data at synoptic stations adopted from Harris et al. (2014) and the evaporation rates calculated in section 2.3.2.1, respectively. The reference period for the bias-correction of the rainfall time series was arbitrarily chosen to be between 1960 and 1995. This period was long enough to cover inter-annual climatic variabilities before and after the lake's shrinkage. The annual inflow volumes required to attain the target level of 1,274.1 m by 2050 were estimated using the calibrated level-volume relationship discussed in section 2.3.2.1. Considering the annual water gain by the lake from precipitation and projected evaporation, we calculated how much water is required annually to refill the lake to the target level by 2050. The calculations were repeated for each multi-GCM ensemble member over the acceptable sets of calibration parameters to determine the uncertainty ranges for the required volume. Finally, the annual inflow required for the restoration was distributed across the tributaries proportionally to the discharge history of each river. The estimated water requirement for each river was then used as an input for the water allocation model in step 3 of the framework.

2.3.3 Step 3: Water allocation model

To capture the variabilities and uncertainties of GCM-based predictions of the future climatic and hydrological conditions in the Urmia basin, the water allocation modelling and optimisation (step 4) were based on a mean of the multi-GCMs with 34 ensemble members for the RCP 4.5 scenario and 26 members for RCP 8.5. All the required climatic parameters were computed as an ensemble mean with equal weighting of each ensemble member. The final annual water requirement for restoration of the lake by 2050 was set as the mean of both multi-parameter and multi-GCM ensemble predictions under RCP 4.5 and RCP 8.5 (i.e. 3,648 and 3,692 $Mm^3 yr^{-1}$, respectively).

We calculated the available runoff and the increase in water demand to meet the lake's inflow target at the river-basin scale and then prioritised the latter to achieve the aims of the land-based restoration plan. In each watershed, the remaining water (the difference between available runoff and lake inflow requirement to meet the restoration target) was distributed among the industrial, municipal, and agricultural sectors, leading to an estimate of the maximum water available for irrigation after the distribution.

The water-resources management model in the Mike HYDRO Basin module (Mike 2017) was used for the allocation of water within each river basin. The total projected allocation of the annual run-off to water users, hydropower plants, and reservoirs was prioritised by the global ranking algorithm implemented in the model. The highest global priority was assigned to the municipal and industrial water users, while irrigation water users were assumed to have lower priority because of the lower economic productivity of water in agriculture. The discharge at the river inlet into the lake was set equal to the annual required flow for lake restoration. The yearly surface water run-off in each tributary catchment and sub-catchment was estimated as a product of precipitation and the runoff ratio, using the average of the multi-GCM bias-corrected precipitations between 2020 and 2050, and calibrated historical run-off ratios extracted from Table 2.6.3 in JICA (2016). Annual run-off ratios were calculated utilizing observed precipitation data (provided by Iran's Water Resources Management Company) and recorded surface discharge at the neighbouring end points of the supplying river basins. The accuracy of the runoff ratios for each basin was checked by calculation of the runoff depth (annual discharge divided by the catchment area) for each river. The average runoff depth for the supplying rivers was 318.6 mm, with Mahabad, Shahr, and Talkhe having the highest runoff depths of 699.9, 472.7, and 419.9 mm, respectively (Table 2.6.4 in JICA (2016)).

There are no complete data on the proportion of the municipal water that is returned back as grey water or how much urban water is consumed. Equally, data on future projections do not exist. Therefore, the time series of the municipal and industrial water usage per capita and the increase in the number of residents in each river basin (see WRMC (2007) for details) were applied to predict the non-irrigation water demand until 2050. It was assumed that the current urban per-capita water is consumed completely and there is no return (as grey water) to the network. The expected annual outflow (evaporation and agricultural water supply) from each reservoir under construction was also accounted for as additional water. The outputs from this part of the framework were fed into the optimisation model as the maximum available water for irrigation, as discussed in the next section.

2.3.4 Step 4: Farm-economic optimisation

This step aims to identify optimal solutions for a full or partial restoration of the lake. It involves basin-scale optimisation of land use and cropping patterns subject to water availability, while protecting farmers' income. The latter is used as the objective function, as follows (García-Vila et al. 2012):

$$TGM = \sum_{i=1}^{N} [(P_{ci} \times X_i \times F_i(Y_i) + S_i \times X_i)] - (C_{ai} \times X_i + P_{wi} \times Y_i \times X_i \times 10)] \qquad (3)$$

where *TGM* is the total gross margin (US$), *i* represents each crop analysed (1,2,…,*N*), *X* is the area under cultivation (ha), *Y* is the seasonal irrigation demand (mm), *F(Y)* is the crop-yield production function (t ha^{-1}), P_c is the income received by farmers (US\$ t^{-1}), S is the crop subsidy (US\$ ha^{-1}), P_w is the price of irrigation water (US\$ m^{-3}), C_a is the production cost (US\$ ha^{-1}) and 10 is the units conversion factor.

The optimisation is subject to the constraints on the available cropping lands and irrigation water:

$$\sum_{i=1}^{N} X_i \leq A \tag{4}$$

$$\sum_{i=1}^{N} (X_i \times Y_i \times 10) \leq A \times W \tag{5}$$

where A represents the total area under cultivation in each river basin, and W is the available annual water in each river basin (m^3 ha^{-1}).

The model was optimised using MATLAB Optimization Toolbox® (Mathworks) and applying the ‘fmincon’ function. As this function minimises rather than maximises, the inverse of the objective function (*1/TGM)* was used and optimisation carried out for each year between 2020 and 2050, subject to the above constraints. The final optimal cropping pattern and land use for agriculture were estimated as an average of the optimal land use and cropping pattern calculated for each year. It should be noted that the optimisation model does not account for all the counter-measure projects employed by the Iranian government, such as facilitating the effluent transfer to the lake (for a complete list of these, see WRMC (2007)).

The various inputs into the optimisation model, shown in Figure 2-2, are discussed below.

2.3.4.1 Crop-yield functions

One of the inputs required for the optimisation is an estimate of the variations in crop yields with applied irrigation water in face of projected climatic conditions. The AquaCrop model (Steduto et al. 2009; Raes et al. 2009) was used for this purpose to determine the crop-yield responses (*F(Y)* in Eq. 3) to climatic and irrigation variability and to develop crop-water production functions. The following major cultivated crops in the Urmia basin were considered: winter wheat, barely, maize (grain), sugar beet, oilseeds, potato, tomato, cucumber, water melons, alfalfa, dry beans, apple, pear, and stone fruits. Their yield responses were analysed to determine the crop yield as a function of available irrigation water and other environmental factors. In addition, pistachio was selected as one of the alternative crops for cultivation as it has a lower irrigation demand and higher market value (also proposed by the Restoration Committee). Cultivar-specific crop parameters, including the time needed to reach the maximum canopy cover (CCx), canopy senescence, physiological maturity and flowering (or the start of yield formation) were adopted from Table 11 in Allen et al. (2005). This source was also used to obtain the average maximum effective rooting depth (Table 22 in Allen et al. (2005)). The average planting density, planting method (direct sowing or transplanting) and use of fertilisers were the field and management parameters tuned by the historical crop yields. Historical crop yields (Food et al. 1998) were compared with simulated crop yields to validate the AquaCrop model against field and irrigation management practices. Regression techniques were then applied to obtain the yield response function for each crop for the period 2020 - 2050. For other cultivated crops and where input data for AquaCrop model parameterization were missing, the average annual potential evapotranspiration was assumed at 700 mm. The latter was calculated using the Hamon method based on monthly data of three weather stations (JICA 2016).

Based on the soil map of the Urmia basin (Hengl et al. 2017), the soil type in this area can be categorised as Inceptisol (Service USDoASC, 1999). The soil moisture profile for each region was acquired from ERA-Interim reanalysis. Daily rainfall (mm) and climate time-series data, including daily relative humidity, minimum and maximum temperature (˚C), and wind speed (m s^{-1}) between 2020 and 2050 were obtained by taking an average from the bias-corrected output of the multi-GCM ensemble. As mentioned earlier, the climate time series were bias-corrected according to Hawkins et al. (2013) using the updated data from Harris et al. (2014).

2.3.4.2 Partial restoration and land use optimisation in the agricultural sector

As mentioned earlier, the optimisation model allows provision of the required water for restoration of the lake. However, in some cases, attaining the optimal land use and cropping patterns may not be viable in the short term. Accordingly, feasible partial restoration solutions in parallel with the agricultural land use optimisation would be beneficial to mitigating the adverse consequences of the lake's shrinkage. If partial restoration is needed and feasible, it is important to determine the parts of the lake that should be prioritised for restoration. This will depend on many factors, such as agriculture, extraction of minerals from the lake bed or recreational activities. Here we focus on the avoidance of dust emissions from the dried lake bed due to their adverse impacts on human health, animals, and vegetation. The following sections provide an overview of the dust emission sources and the methodology used to estimate the geospatial dispersion of PM_{10} through vertical dust parameterisation mentioned previously in section 2.3 and Figure 2-4.

2.3.4.2.1 Identification of dust emission sources

Locations vulnerable to wind erosion were characterised through the daily Aerosol Optical Thickness (AOT) obtained by Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra dataset (5 Min L2 Swath, 3 km resolution at nadir, both over land and ocean, MOD04_3K) over the lake's area between 2010 and 2016 (Levy et al. 2015). We chose this period because the lake had the lowest water level between 2010 and 2015. For each pixel, the average of daily AOT intensities was multiplied by the number of days which had the measured AOT values > 0.15 (intensity × frequency) to obtain an index for each pixel, here called "aerosol intensity-frequency". The pixels with an aerosol intensity-frequency index value higher than the 75-percentile of the calculated indices for all pixels between 2010 and 2016 were identified as dust sources within the lake's domain. These sites were mostly located on the lake's islands and on the margins of the southern half of the lake. The sediment at the lake's periphery can be classified as sandy loam and silty clay loam-textured, except the Jebel site on the west shore which has sand dunes (Table 2-4).

Table 2-4: Sediment composition of 12 sites on the margin of Lake Urmia. Soil textures in Figure 2-9 are based on these compositions.

Site	Sand (%)	Silt (%)	Clay (%)
1	4.7	77.5	17.7
2	8.4	75.2	16.3
3	22.9	60.2	16.9
4	42.0	45.0	13.0
5	4.3	78.7	17.0
6	30.5	57.0	12.5
7	31.0	56.0	13.0
8	46.0	42.0	12.0
9	43.0	49.0	8.0
10	7.0	76.0	17.0
11	4.0	66.0	30.0
12	73.0	18.0	9.0

2.3.4.2.2 Vertical dust parameterization

Two parameters should be determined first in estimations of the vertical dust flux: 1. the threshold friction velocity at which the soil particles start to move; and 2. the horizontal saltation flux (Kang et al. 2011). Threshold friction velocity (u_{*_t}) is defined as the minimum friction velocity required for the commencement of soil-particles movement and the saltation process. Here, the relationship proposed by Shao et al. (2000) was used to calculate the threshold friction velocity. For a particle with diameter d_s, six formulations were used to determine the horizontal sand flux, $q(d_s)$ (kg m^{-2} s^{-1}). Details of those equations can be found in White (1979), Kawamura (1951), Owen (1964), Sørensen (2004), Lettau et al. (1978), and

Durán et al. (2011). Aggregate flux intensity, Q of saltating particles of all sizes was then estimated as follows (Shao et al. 2002):

$$Q = \int_{d_1}^{d_2} q(d_s)p(d_s)\delta ds \tag{6}$$

where d_1 and d_2 are the lower and upper saltating particle size limits, respectively, and $p(d_s)$ is the soil particle size distribution. Shao (2001) assumed $p(d_s)$ as a combination of two idealised particle size distributions, known as minimally disturbed particle size distribution $p_m(d_s)$, and fully disturbed particle size distribution $p_f(d_s)$. During weak erosion, $p(d_s)$ is close to $p_m(d_s)$, while when the erosion is strong, $p(d_s)$ is close to $p_f(d_s)$. Shao (2001) represented $p(d_s)$ as:

$$p(d_s) = \omega p_m(d_s) + (1-\omega)p_f(d_s) \tag{7}$$

and

$$\omega = \exp(-(u_* - u_{*_t})^3) \tag{8}$$

where u_* is the friction velocity and u_{*t} is the threshold friction velocity.

The minimally and fully disturbed particle size distributions can be deemed as the sum of lognormal distributions, i.e.:

$$p_{m,f}(d_s) = \frac{1}{d_s}\sum_{j=1}^{J}\frac{w_j}{\sqrt{2\pi}\ln\sigma_j}\exp(-\frac{(\ln d_s - \ln D_j)^2}{2\ln^2\sigma_j}) \tag{9}$$

where J is the number of modes, and D_j, σ_j, and w_j are the mean diameter, standard deviation, and the weight of the jth mode particle size distribution, respectively (Shao 2001). Minimally and fully disturbed size distributions for different soil types were calculated using parameters for four soil textures, namely sand, loam, sandy clay loam, and clay, proposed by Shao (2004). Saltating particles can mobilize other particles with different size ranges by their impact on soil surface. Dust particles (defined as particulate matter < 70 μm) are not lifted by the direct effect of wind since the inter-particle cohesive forces are predominant compared to the aerodynamic forces (Kok et al. 2012). These particles are primarily ejected and lifted from the soil surface due to the impact of saltating particles on them (Gillette 1974). This ejection from the soil results in a vertical dust particle flux into the atmosphere and subsequent particle suspension (which can be either short-term or long-term, depending on the size of the particles).

In this study, the following two methods were applied to estimate the vertical dust flux, F, from the dried lake bed:

1. A simplified total vertical dust flux F was obtained by modelling the ploughing process of individual saltating particles (assuming the impact angle of a saltating particle is 13°) as (Lu et al. 1999):

$$F = \frac{C_\alpha g f \rho_b}{2p}(0.24 + C_\beta u_* \sqrt{\frac{\rho_p}{p}})Q \tag{10}$$

where p (500 for silty clay loam, 10,000 for loam, and 50,000 for sand dune) is the plastic pressure of the soil surface in N m^{-2} (surface hardness), f is the fraction of dust contained in the volume of the soil, ρ_b and ρ_p are the densities of the bulk soil and particles, respectively, and C_α and C_β are constants; Q is the aggregate flux intensity determined by Eq. (6). For loamy and silty-clay loam which form 90% of the lake's saline playa, C_α is equal to 0.0002 and 0.0006, respectively; for sand dunes, C_α is 5. For C_β, the value of 1.37, suggested by Lu et al. (1999), was applied for all soil textures. Here, f represents the fraction of mineral particles < 10 μm and is equal to 15.2%, the average percentage of clay available in 12 samples gathered from the sites around the lake (Table 2-4).

2. A simplification of a vertical dust emission parameterization was followed in the second method (Shao 2001). The proposed method for estimation of the vertical dust emission is based on saltation bombardment and aggregate disintegration mechanisms. Using this method, the dust emission rate for particle-size range with a mean value of d_i and an increment Δd_i induced by saltation of particles of size d_s, can be calculated as:

$$F(d_i, d_s) = c_y \eta_{fi}[(1-\omega) + \omega\sigma_p](1+\sigma_m)\frac{gq(d_s)}{u_*^2} \tag{11}$$

where c_y is a dimensionless coefficient, $q(d_s)$ is the stream-wise saltation flux of particle with diameter d_s and g is the acceleration due to gravity. η_{fi} can be obtained by:

$$\eta_{fi} = \int_{d_i - \Delta d_i/2}^{d_i + \Delta d_i/2} p_f(d)\delta d \tag{12}$$

and σ_m is:

$$\sigma_m = 12u_*^2 \frac{\rho_b}{p}(1 + 14u_* \sqrt{\frac{\rho_b}{p}}) \tag{13}$$

where p is soil plastic pressure and σ_p is the ratio of free dust to aggregated dust, i.e.:

$$\sigma_p = \frac{p_m(d_i)}{p_f(d_i)} \tag{14}$$

Assuming the particles are divided into I particle-size bins, each with a mean diameter d_i and an interval length Δd_i, the model proposed by Shao (2004) can be considered as a spectral dust emission model which can be utilised for estimation of the vertical dust flux of various particle size ranges. The vertical dust emission F from bin i is then:

$$F(d_i) = \int_{d_1}^{d_2} F(d_i, d) p(d) \delta d \qquad (15)$$

where d_1 and d_2 are the lower and upper size limits for saltating particles. Total vertical dust emission of fine particulate matter is the sum of all size bins. The following values were assumed for p (required for calculation of σ_m) and c_y for different soil types (Shao 2004): for sand dunes p = 1,500 Pa and $c_y = 5\times10^{-5}$; for silty clay loam p = 50,000 Pa and $c_y = 1\times10^{-5}$; for loam p = 10,000 Pa and $c_y = 5\times10^{-5}$. To compute the density of air, the average air temperature of Urmia city of 21˚C was used, measured at the climate station between May and September, the driest period for the lake and a higher dust emission possibility. The average dew temperature between May and September for the Urmia climate station is 7.56 ˚C. Based on this, the average air density at the location of the lake of 10.32 kg m^{-3} was used. For the calculation of dynamic viscosity, it was assumed that the pressure is equal to the atmospheric pressure at sea level (since the change in air pressure has a minimal impact on the viscosity of gases). The dynamic viscosity of air was estimated at 18.17×10^{-6} Pa s. Various parameterizations of vertical dust emission were compared with the wind tunnel (Roney et al. 2006) and in-field measurements (Nickling et al. 2001) of salty dust emission as a function of friction velocity to determine the best formulation for each soil texture (Figure 2-5). In order to include the effects of moisture and non-erodible elements, the saltation threshold friction velocity should be corrected, usually by multiplication of the threshold friction velocity by correction factors > 1. Soil moisture and the presence of non-erodible elements increase the saltation threshold friction velocity. Expanded parameterization of Fécan et al. (1998) was used for the correction of the soil moisture effect (Appendix 1, Eq. A1.3). Moreover, the threshold friction velocity correction factor (f_λ) proposed by Raupach et al. (1993) was applied to account for the presence of roughness elements (Appendix 1, Eq. A1.4).

2.3.4.3 Protecting farmers' income: irrigation costs and subsidies

Currently, the irrigation efficiency in the Urmia basin is 41% (WRMC 2007) and many endeavours are in progress to improve it to 70% or more for farms and 90% for horticultural gardens. However, in this study, a conservative assumption was made that the irrigation efficiency would increase only to 60%. Traditionally in Iran, the agricultural water price paid by farmers in modern, semi-modern, and traditional irrigation networks has been 1%, 2%, and 3% of the final value of the produced crop, respectively (WRMC 2007). With this tariff scheme, farmers only pay 25% of the real calculated cost of water in three different irrigation networks and there is no incentive for improving the water use efficiency. In our optimisation model, the irrigation water price of US$ 0.06 per m^3, recommended by the Iranian Ministry of Agriculture, was assumed to eliminate inefficiencies in the water consumption chain (WRMC 2007).

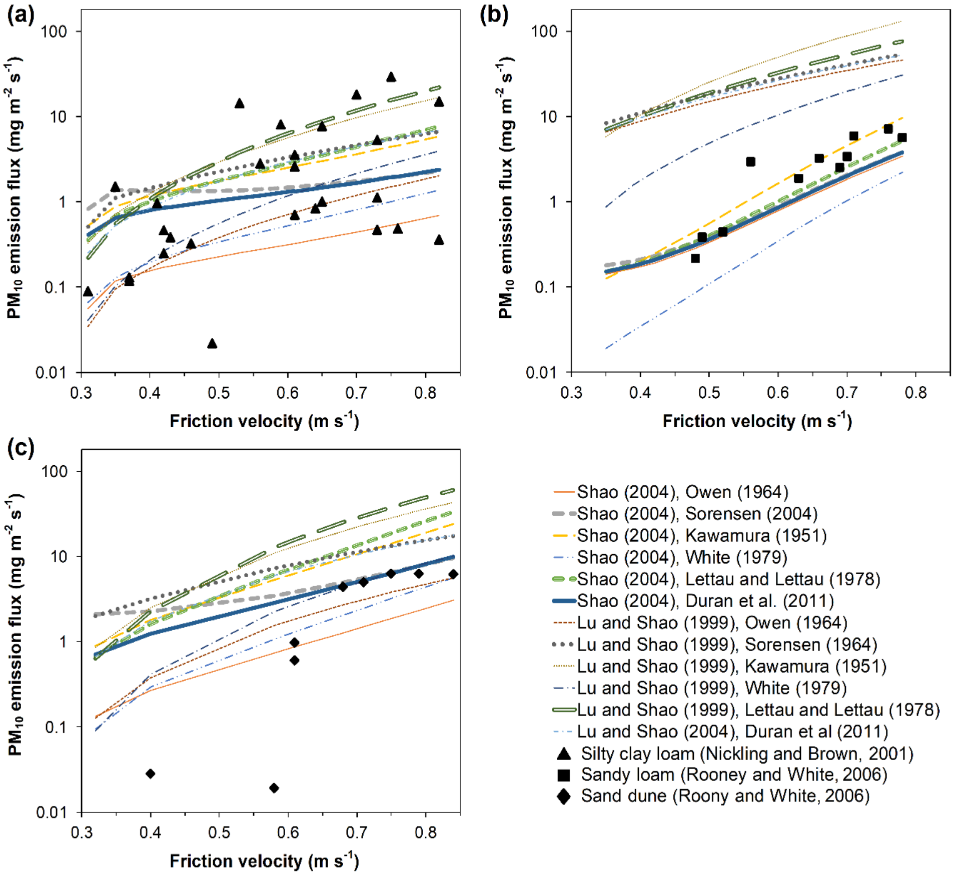


Figure 2-5: Observed and modelled vertical PM$_{10}$ emission from the Lake Urmia's saline playa: a: Silty clay soil. **b**: Sandy loam. **c**: Sand dune. Since there are no experimental data on saltation and vertical dust flux from the playa as a function of wind friction velocity, measured data for other saline lakes were used to assess and validate the performance of various vertical dust formulations. For each soil texture, measured values are shown as dots while model outputs are represented by the lines. For silty clay loam, the vertical PM$_{10}$ dust emission rates were observed in an in-field wind-tunnel at potentially high emission areas in Owens Lake (Nickling et al. 2001). Experimental data for sandy loam and sand dunes were measured in the saltation wind-tunnel at University of California Davis (Roney et al. 2006) where crustal sediments were conveyed from Owens Lake to the wind-tunnel site. For silty clay loam, the horizontal saltation flux from White (1979) and vertical dust formulation from Lu et al. (1999) were selected to model dust emissions from the Lake Urmia's bed. For sand dunes and sandy loam, White (1979) & Shao (2004) and Kawamura (1951) & Shao (2004) formulations were used, respectively.

Alternatively, to protect farmers against a new irrigation water pricing system and capture climate-induced risks in crop yields, a new subsidy scheme was considered here. Prior to 2010, irrespective of the yield of harvested crops, farmers were entitled to receive fixed annual subventions. To be eligible for receiving the funding, it was assumed in this analysis that the farmers' harvested annual crop yield must fall below the 30th percentile of the historical local annual crop yields between 1995 and 2015 (based on data availability for all crop yields). With this subsidy scheme, we tried to cover the agriculture-associated risks throughout the growing season, including adverse climatic conditions, natural disasters, and pest invasions. The subsidy was estimated in US\$ ha^{-1} as $Subsidy = a \times (yield)^{b}$, where yield is in t ha^{-1} and a and b are constant parameters for each cultivar. At lower yields, it was assumed that the payable subsidy increases with yield reduction compared to the reference yield. The final constant parameters used for each crop in the subsidy function based on historical yields are presented in Table 2-5. For some crops, including wheat, barely, and oil seeds, the final product price was calculated as a function of the yield to capture the high dependency of the final crop price on annual yields. In other words, the final crop price was assumed to be a function of the yearly crop yield because in years with low average crop yields, the final crop price increases.

Table 2-5: Historical annual yields, production costs, and average producer prices for different crops based on the data between 1995 and 2015.

Crop	Yield (30th percentile) (t ha^{-1})	Production costs [a] (US\$ ha^{-1})	Producer price[b] (US\$ t^{-1})	a	b
Apple/pear	14.17	2547.82	605.59	2.14×10^{21}	-17.79
Stone fruits	9.54	6085.25	482.77	1.41×10^{18}	-19.75
Pistachio	0.63	7473.92	10042.24	0.01873	-7.14
Winter wheat	2.90	892.44	-0.0768 (yield) + 473.9	9.72×10^{6}	-13.15
Barely	2.64	680.51	-97.44 (yield) + 482.2	2.36×10^{15}	-39.23
Potato	22.14	2725.56	284.67	4.28×10^{43}	-33.46
Sugar beet	30.33	3025.31	95.10	8.85×10^{16}	-10.07
Oil seeds	0.97	1201.65	-274.7 (yield) + 871.9	0.03969	-12.84
Tomato	27.21	2783.12	237.80	1.10×10^{46}	-33.17
Cucumber	13.52	1331.60	422.95	1.76×10^{45}	-40.67
Water melons	9.85	1331.60	158.36	4.64×10^{45}	-48.59
Maize	5.92	1878.25	371.60	3.30×10^{18}	-24.21
Bean (dry)	1.14	1096.85	1616.97	2.15	-11.45

[a] Source: The Iranian Ministry of Agriculture for year 2014.

[b] Producer prices are adopted from FAO (2016) and show the final price received by farmers. For winter wheat, barely, and oil seed the producer prices are presented as function of the yield to capture the dependency of the price to produced crop availability. a and b are the constant parameters in the subsidy function.

2.3.4.4 Greenhouse gas emissions

The potential GHG emissions associated with the suggested land use change were estimated at the optimal solution determined through the optimisation model. The focus was on two major plausible activities: 1. conversion of the current marginal irrigated croplands to rain-fed systems; and 2. conversion of the current marginal croplands to grasslands/rangelands. Quantification of the change in GHG emissions as a result of shifting from one cropping pattern to another in irrigated lands is not as easy; hence, the change in GHG emissions related to the cropping pattern alteration within the irrigated croplands was not considered. For the conversion of irrigated to rain-fed croplands, we assumed that the current irrigated croplands in the river basins are converted into rain-fed cultivation of cereal crops. Furthermore, the conversion factors for the warm temperate dry IPCC climate zone were used to evaluate the change in the GHG emission in the Urmia basin. The sequestration rate was calculated by subtracting the average net soil CO_2 eq. emissions from the whole soil profile (0 - 1.2 m depth), with the clay contents of 18%, 34%, and 40% in the Australian cereal belt (Dalal et al. 2001), from the net cropland soil CO_2 eq. emissions of the IT-BCi site (5200 ± 410 kg CO_2 eq. ha^{-1} yr^{-1}), reported by Schaufler et al. (2010). We used these empirical values since the estimated emission rates of their studied site's IPCC climate zone are similar to the Urmia basin's IPCC climate zone. For the restoration of cropland to grassland, we used the estimated emission factor range for the warm temperate dry IPCC climate zone (2,475 - 18,069 kg CO_2 eq. ha^{-1} yr^{-1}) in Diaz et al. (2012) to calculate the mean and standard deviation of the possible reduction in GHG emissions.

Conversion activities were studied in the context of two land use change rates: achieving the land use program targets in ten and in 30 years (Cameron et al. 2017). Monte Carlo simulation was used to propagate the uncertainty of the estimated emission rates for each land-based activity, with 50,000 iterations assuming normal distribution from the mean and standard deviation of the activity's net sequestration rate. The final probability distribution of each activity is the product of the Monte Carlo sampling results and the converted area rate. The 5th and 95th percentile of the cumulative reduction in GHG emissions of each activity during the conversion interval (ten or 30 years) were used to determine the confidence intervals.

2.4 Results and discussion

The following sections discuss the results obtained in the different steps of the framework.

2.4.1 Step 1: Projected climate in the Urmia basin

Figure 2-6 presents the overall variations in the projected changes in precipitation and near-surface (2 m) air temperatures (hereinafter referred to 'air temperature') across the Urmia basin, averaged by month. The results were computed by comparing air temperatures and precipitation outputs of each model for the reference (1960 - 1995) and future (2020 - 2050) periods. We chose 1960 - 1995 as the reference period because in that period the lake was in good condition and had high water levels. As can be inferred from Figure 2-6, almost all the

GCM ensemble members predict an increase in monthly air temperatures in the Urmia basin compared to the reference period. An increase in the basin's average monthly air temperatures was found for both 2050 scenarios: 1.95˚C for RCP 4.5 and 2.47˚C for RCP 8.5. A slightly lower increase is expected for the average annual temperature: 1.86˚C and 1.94˚C. Furthermore, a clear decreasing precipitation trend is projected to continue until 2050. The annual average precipitation in the basin is expected to decrease by 2.7% (RCP 4.5) and 11.6% (RCP 8.5) by 2050, relative to the reference period (1960 - 1995). The projected monthly precipitations show large seasonal variations. For RCP 4.5, the monthly precipitation ranges between -17.1% in August to 6.5% in December, while for RCP 8.5, all monthly precipitation are decreasing by up to -33.1% (September). In general, the decrease in precipitation rates is higher in spring and summer. However, it should be noted that the uncertainty analysis in Figure 2-6 indicates a spectrum of possibilities with respect to precipitation, including no change, increase, and a greater reduction than discussed above.

In addition to the temperature and precipitation trends, a 52% to 57% increase in the basin's population is anticipated by 2050 relative to the year 2007 (WRMC 2007), further exacerbating the desiccation conditions if left unaddressed.

2.4.2 Step 2: Salt water balance — the water required for restoration

As discussed in section 2.3.2, the results of the lake's salt balance in the period of shrinkage (1996 - 2010) suggest a considerable raise in the lake's floor due to the salt precipitation on the lake bed at a rate of 4 cm yr^{-1} (Figure 2-7). Therefore, it can be concluded that the net inflow into the lake during the shrinkage was even smaller than the inflow value estimated here based only on the lake level-area-volume relationships. The precipitation of the salt on the lake bed is a result of the negative water balance and excess evaporation over the net inflow to the lake. As the lake's volume decreases, supersaturation of the dissolved salts causes precipitation of excessive salts on the lake bed.

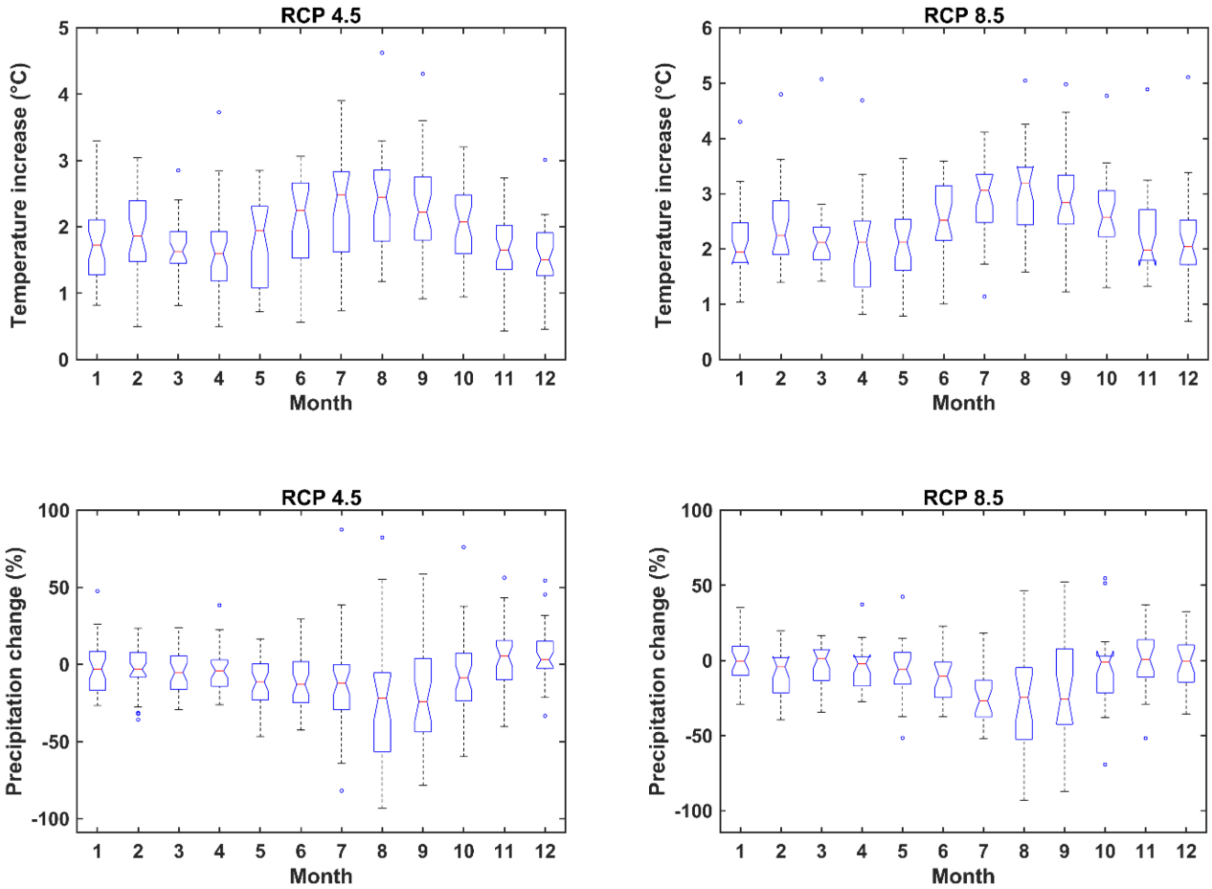


Figure 2-6: Overall monthly variations in the projected changes in precipitation and near-surface air temperatures across the Urmia basin, calculated by comparing the output of each GCM ensemble member for the reference (1960 - 1995) and future (2020 - 2050) period. The central mark in each box indicates the median, and the bottom and top edges of the box are the 25th and 75th percentiles, respectively. The outliers are shown as individual dots.

It is generally accepted that Lake Urmia receives a comparatively small portion of its annual inflow from groundwater discharge. However, there is disagreement on the estimates of groundwater inflows, ranging from 3% (Hasemi 2011) to 49% (JICA 2016) of the total water input into the lake. The sub-surface inflow to the lake is mainly through the wetlands around the lake because the thick halite bed does not allow for water flow through the lake bed (JICA 2016). According to the results of our model, the average annual sub-surface inflow calculated as the difference between the total inflow (from the water balance analysis) and the surface water inflow (from stream gauges) varies only between 12.8% and 15.3%, depending on the method used for the estimation of annual evaporation. This indicates that the lake's water budget is mostly dependent on the surface inflow so that any future restoration plan should focus on increasing the amount of surface water flowing into the lake.

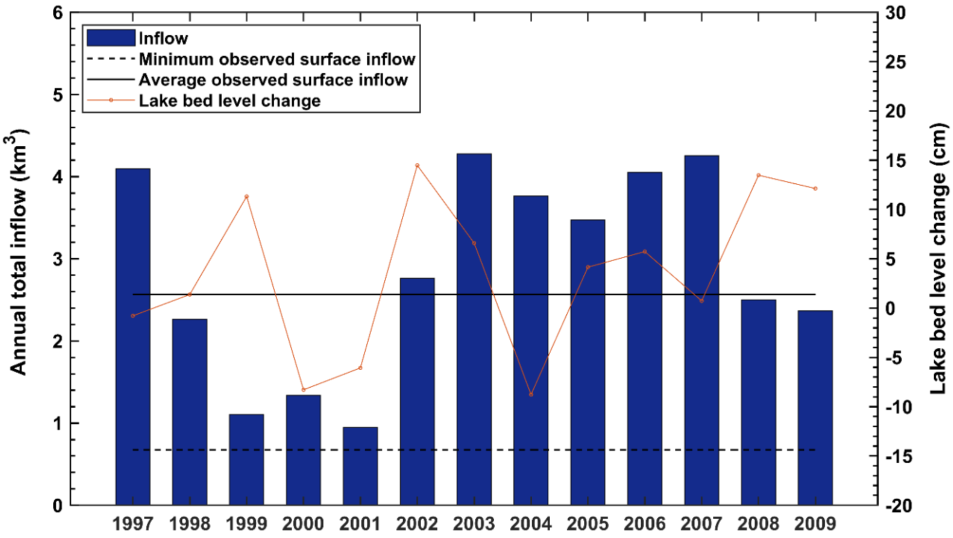


Figure 2-7: Precipitated/dissolved salt on the lake bed and average annual total inflow into Lake Urmia, including both surface and underground flows (1997 - 2009). The average salt precipitation rate on the lake bed during the desiccation period is 0.04 m yr^{-1}.

Figure 2-8, a shows the annual volume of surface water required for restoration of the lake by 2050 under the RCP 4.5 and RCP 8.5 scenarios for GCMs with the realization number of 1 (the other numbers are not shown for simplicity). These surface inflow requirements were estimated based on the bias-corrected evaporation and precipitation projected by the various GCMs. For each GCM, the uncertainty involved in the process of calibrating the hydrological balance over the lake's area was represented by the corresponding error bar (minimum and maximum required surface inflow values calculated by Monte Carlo simulations). According to the mass balance results, the average surface inflow of 3.648 km^3 yr^{-1} (standard deviation, SD = 0.271 km^3) is required under RCP 4.5 and 3.692 km^3 yr^{-1} (SD = 0.315 km^3) under RCP 8.5 for the complete 30-year lake restoration to the target level of 1274.1 m. Our predicted inflow is comparatively lower than that of JICA (2016) who estimated the annual river inflow volume of 4.95, 4.55, and 4.40 km^3 yr^{-1} for ten, 20, and 50 years' restoration scenarios, respectively. However, their model did not consider the projected climate variabilities to attain the target water level. Taking into account the annual sub-surface water gain of 12.8% in RCP 4.5 and 15.3% in RCP 8.5 (Figure 2-7), our model shows that the total annual inflow required to restore the lake in 30 years is 4.183 km^3 yr^{-1} and 4.358 km^3 yr^{-1}, respectively.

To put these results in perspective, the total current annual surface inflow to the lake is 2.13 km^3 yr^{-1}. Therefore, under the RCP 4.5 scenario, a 70.8% increase in the inflow would be needed to rehabilitate the lake fully in 30 years' time; the equivalent increase for RCP 8.5 is 73.2%. As the sub-surface water gain is limited, much of the required water would need to be supplied by the lake's tributaries, with the water allocations shown in Figure 2-8, b and c (full restoration). This sensitivity to the surface inflow, as well as the substantial technical, institutional, and behavioural barriers to reducing the upstream river withdrawals, make full restoration of the lake a grand challenge, particularly in the short term. Through a dynamic environmental inflow plan, Alborzi et al. (2018) predict that complete restoration of Lake Urmia under arid conditions may take up to 16 years even if a 40% decrease in the basin irrigation demand occurs. Therefore, partial restoration of the lake, similar to that implemented for the Aral Sea, could be a more promising approach to accelerate the rehabilitation process, as discussed next.

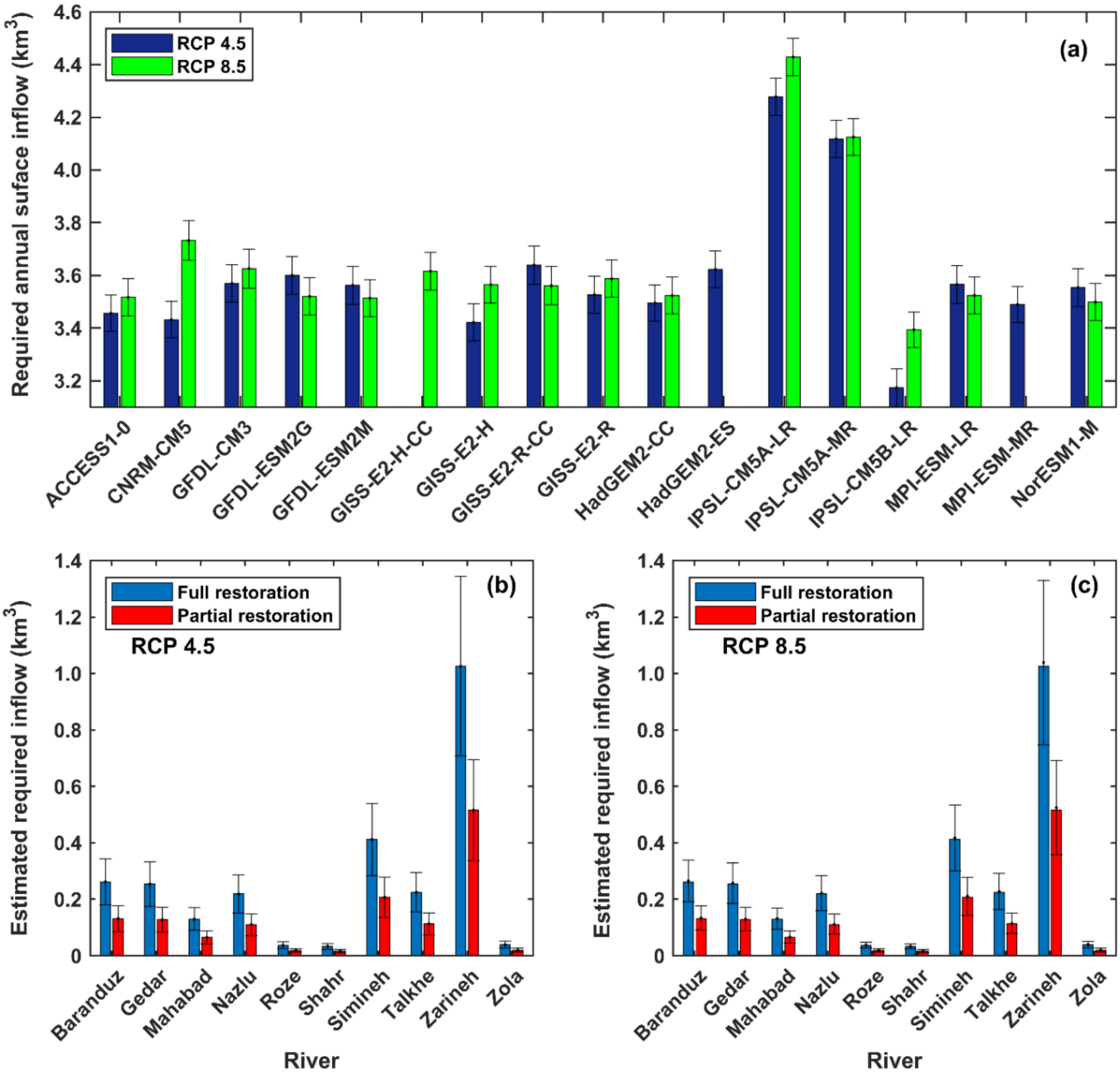


Figure 2-8: a: Required annual surface inflow for restoration of the lake by 2050 under the RCP 4.5 and RCP 8.5 scenarios, predicted by the GCMs for the realization number of one. **b** and **c**: Minimum, maximum, and mean of the estimated annual volume of water required from each river supplying the lake (shown on the x-axis) for a ten-year partial (southern half of the lake) and 30-year full restoration of the lake.

2.4.3 Steps 3 and 4: Water allocation and optimal land use for agriculture

2.4.3.1 Short-term: recommended partial restoration

Decreased evaporation as a result of reduced surface area of the lake and the required water volume suggest that the current river inflows would be sufficient for a partial restoration of the lake. However, it is not clear which part of the lake should be prioritised for partial restoration based on the current river inflows.

Some options have been proposed previously for partial restoration, including: 1. conserving the southern shores by building a dike connecting four major islands in the southern part of the lake (Razia et al.) (hatched area in Figure 2-9); and 2. dividing the lake into the northern and southern parts (Hamidi-Razi et al. 2018) (a two-lane highway embankment), and rehabilitating the southern part at the expense of the northern part. In addition to these, we also analysed direct transfer of water to the northern part as a third option.

Since the wind dispersion of mineral aerosols from the dried lake bed is of major concern for environmental and human health (Gillette et al. 1997), here we prioritised the partial restoration option with the best potential for dust emission mitigation over other objectives, such as water quality, migratory birds, island separation, recreation, social, economic, and political, that can be fully, partially, or not at all achieved in partial restoration. To that end, we explored the main sources of PM_{10} through an intensity-frequency index analysis of remotely-sensed AOT between 2010 and 2016 (the period with the highest shoreline recession). As a result, nine highly wind-erodible salt playa sites were detected, with an approximate aggregate area of 1,745.9 km^2 (Figure 2-9).

Parametrizations of the vertical dust models were employed (see section 2.3.4.2) to evaluate the potential for wind erosion from the lake bed as a function of soil moisture, vegetation cover, and texture. For the whole lake bed, meteorological records and vertical dust flux estimates show the potential for the release of 36,391 t PM_{10} yr^{-1}, contingent on the moisture of the bordering salty playa falling below the threshold moisture content for dust emissions (~5%). Based on our simulations, soil stabilization management, including ploughing, grooving, and planting salinity-tolerant seedlings to restore at least 5% of vegetative cover, were found to mitigate only 24.6% of the fugitive dust emissions. However, in the alternative scenario that keeps the soil moisture of the identified dust sources at 15%, the PM_{10} emissions would be reduced by 90.8%, to 3,342 t yr^{-1}. The likelihood of salt-entrained dust suspension can be reduced to almost zero by keeping the sediments moisture in the range of 35% to 40%.

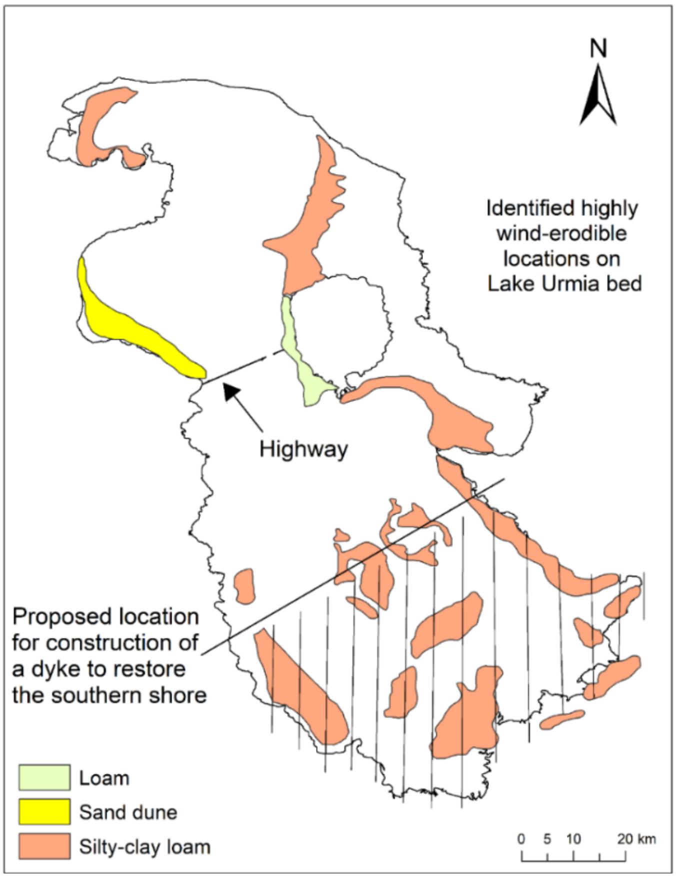


Figure 2-9: Detected dust sources (coloured areas), their soil texture, and the proposed area for partial restoration. White areas represent the regions where no dust emissions are expected. The hatched area denotes the proposed area for restoration of four major islands.

Restoration of the southern half of the lake could mitigate 77% of salt-rich dust emissions from the lake bed. Restoration of the northern part of the highway (see Figure 2-9) and areas below the main southern islands (hatched area in Figure 2-9) could prevent 22% and 39% of the total average dust emissions per year, respectively. Thus, based on these results, restoration of the southern part of the lake should be given a higher priority. Diking would change the bathymetry, connectivity, and physics of the diked area, as well as the water and salt balances, and water level where the salinity tolerance is reached. Still, there is a challenge of defining a singular salinity and water level within the diked area. However, in case the ecological target level for the whole lake (1,274.1 m) is attained in this southern part of the lake, the re-flooded areas would cover more than 70% of the detected dust-source locations of the whole lake. The Aji river in the northern part is responsible for 68% of the lake's annual dissolved salt entry (Table 2-2); therefore, the southern part receives a lower salt input. The decrease in salt concentration would allow for a recovery of the industry of harvesting brine-shrimp dormant eggs (or "cyst"). Under the RCP 4.5 and RCP 8.5 scenarios, our simulation shows that the respective annual water inputs of 1.83 and 1.86 km^3 can restore the southern part of the lake to the target level in ten years (Figure 2-8, b and c, partial restoration). Note that 1.58 km^3 of water is provided currently by the southern tributaries. The remaining volume of 0.32 km^3 can be supplied by the proposed land use management, as discussed below.

2.4.3.2 Long-term restoration: recommended land use change

In the short-term, partial restoration can help to redress the saline dust dispersion issue; however, the question about a sustainable solution for restoration of the lake still remains unanswered. Irrigation accounts for 93% of water consumption in the lake's basin, with an average irrigation efficiency of 37% for arable and 45% for horticultural lands (WRMC 2007). One pragmatic and viable intervention to reduce the withdrawals would be to improve the irrigation efficiency or cultivation of less water-intensive crops (Micklin 2007). It is unclear, however, how the current land use and crop distribution should be changed to use water more efficiently and achieve the required lake restoration target by 2050.

Using the optimisation approach described in section 2.3.4, we determined the optimal land use and cropping patterns that maximise the farmers' net income at the river-basin level, subject to land and water availability. We divided the irrigated area in the Urmia basin into six major zones, represented in Figure 2-10, with the suggested decrease in irrigated areas and change in cropping patterns in these sub-basins given in Figure 2-11. These results are based on the analysis of the effects of regional climate change (for RCP 4.5 and RCP 8.5), expansion of upstream irrigation reservoirs/networks, and population growth along the major tributaries of the lake, considering two plausible lake restoration options: intra-basin restoration (without any human-made water conveyance) and inter/trans-basin restoration (partly relying on water diversion from the Zab basin).

In the case of intra-basin restoration, our results show that optimal water use over the basin requires the conversion of 95,600 ha (RCP 4.5) and 133,687 ha (RCP 8.5) of irrigated land for rain-fed cropland or grassland. Under RCP 4.5, the optimal reduction in horticultural lands (31%) is slightly higher than in the arable lands (28%). Likewise, under RCP 8.5, the optimal reduction in horticultural lands is 34% and in arable 30%. The annual water requirement for restoration of the lake by 2050 are 3,648 and 3,692 Mm^3 yr^{-1} for RCP 4.5 and RCP 8.5, respectively.

For inter-basin restoration, a reduction of 78,700 ha and 114,826 ha in the total irrigated area is proposed under the RCP 4.5 and 8.5 scenarios. This is significantly reduced compared to the area currently used for cultivation in the lake's basin, which is close to 438,900 ha (Figure 2-12). The total inflow into the lake required for this restoration option is 3,048 and 3,092 Mm^3 yr^{-1} for the two respective scenarios.

For the optimal cropping pattern in the basin in both restoration options, ~624 mm (RCP 4.5) and ~681 mm (RCP 8.5) of water per year would satisfy the irrigation demand of the basin (at 60% irrigation efficiency); by comparison, 1,511 mm is used currently (WRMC 2007).

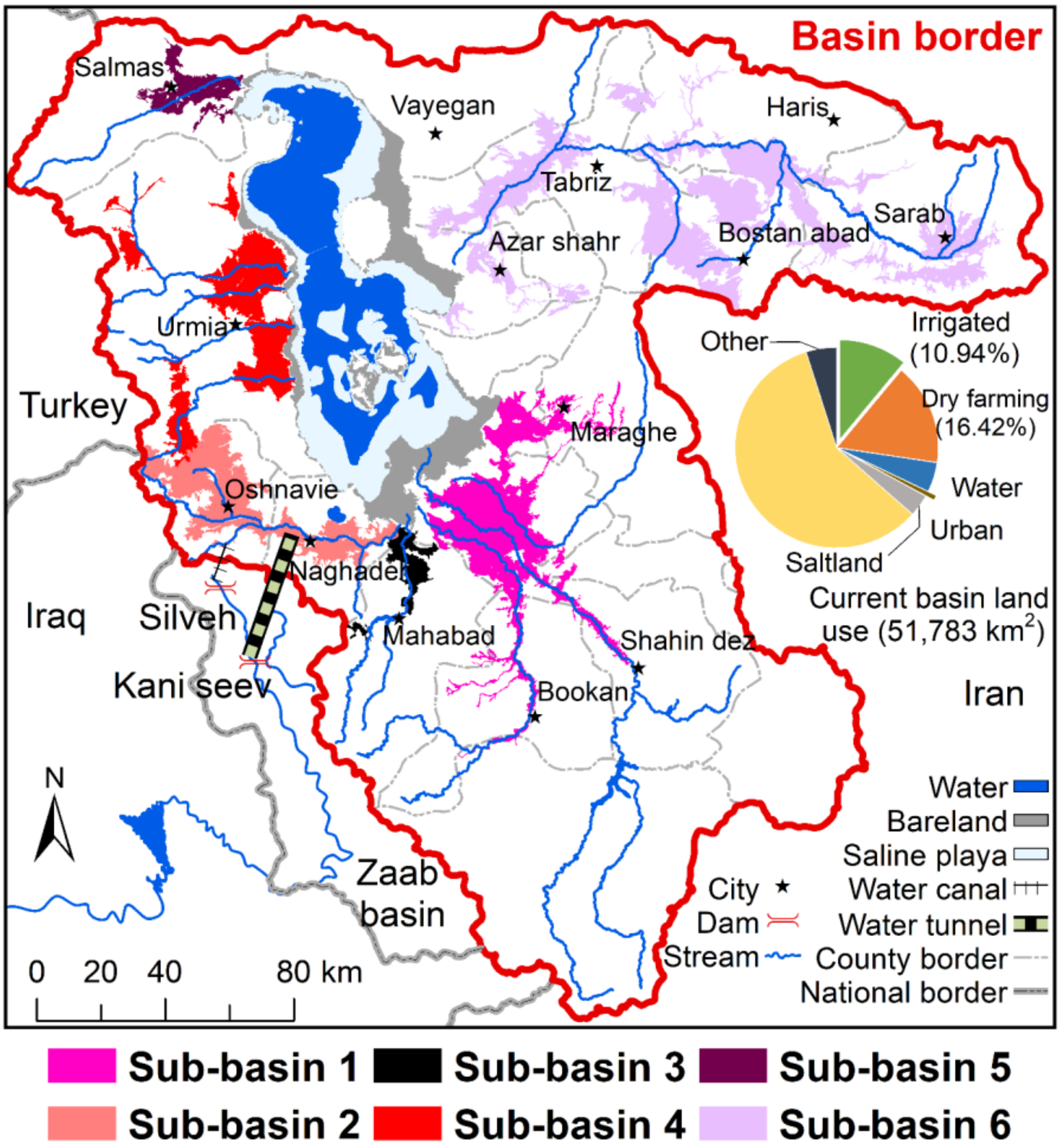


Figure 2-10: The Urmia basin and six major agricultural sub-basins considered in the study.

At the end of the 30-year restoration process, assuming the current final price for the crops considered, we estimated under RCP 4.5 that the unit water productivity would be 0.278 US\$ m^{-3} for the inter-basin water transfer and 0.304 US\$ m^{-3} for the intra-basin restoration. This represents a respective increase of 68% and 87% from the current water productivity of 0.16 US\$ m^{-3}. Similar increases were found for the two restoration options under RCP 8.5: 0.271 US\$ m^{-3} and 0.296 US\$ m^{-3}, respectively.

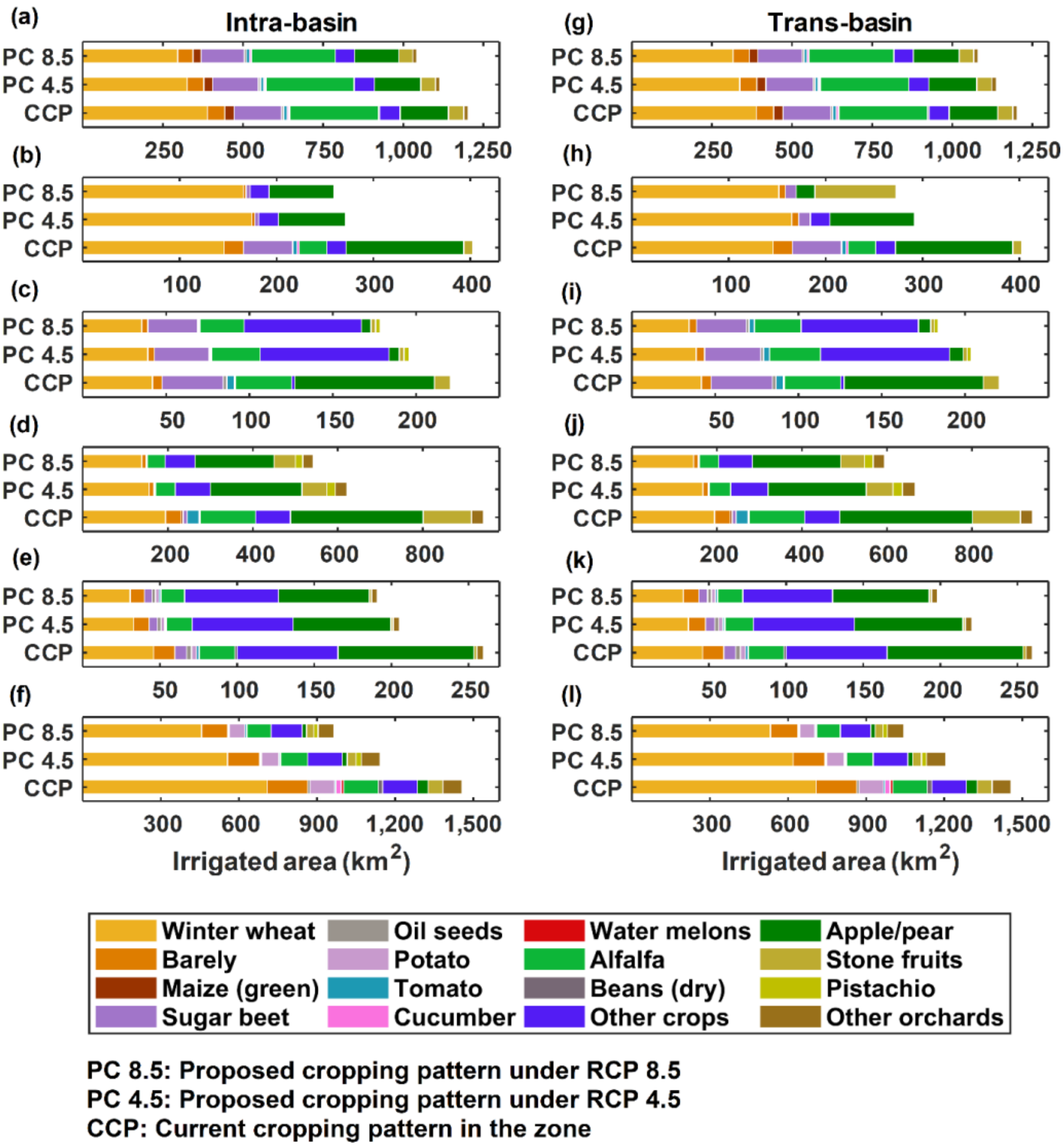


Figure 2-11: Proposed land use change in the Lake Urmia's basin for the restoration of the lake by 2050. Each bar chart shows the proposed change in the current area under cultivation in the agricultural sub-basins (shown in Figure 2-10) considering 14 major crop categories. New cropping patterns refer to two options: intra-basin restoration of the lake (**a** to **f**) and inter-basin restoration (**g** to **l**). Consideration was given to the role of the major stressors: rise in the basin's temperature, increase in population/dry years, and completion of upstream reservoirs and irrigation networks.

Regarding the change in cropping patterns, the share of each crop group would remain generally the same as currently (Figure 2-12). However, the share of individual crops would change, favouring crops, which require less water (Figure 2-11). One such crop is pistachio so its proposed cultivation would increase in intra-basin restoration from the current 718 ha to 4,415 ha (RCP 4.5) and 3,790 ha (RCP 8.5), mostly in areas downstream of the river basins. For inter-basin restoration, the pistachio cultivation would increase to 4,542 ha (RCP 4.5) and 4,038 ha (RCP 8.5). This would also lead to a large increase in farmers' revenue as pistachio has a high final (market) price.

According to the crop yields and the area under cultivation in each year until 2050, our calculations showed that the final annual subsidy cost to cover the new irrigation water cost would be US$ 1.179 million (RCP 4.5) and US$ 1.287 million (RCP 8.5) for intra-basin restoration. For the inter-basin option, these values decrease to US$ 0.995 million (RCP 4.5) and US$ 1.105 million (RCP 8.5), respectively. Overall, winter wheat, barley, and maize need the highest average annual subsidy (9.17, 1.25, and 4.64 US$ ha^{-1}).

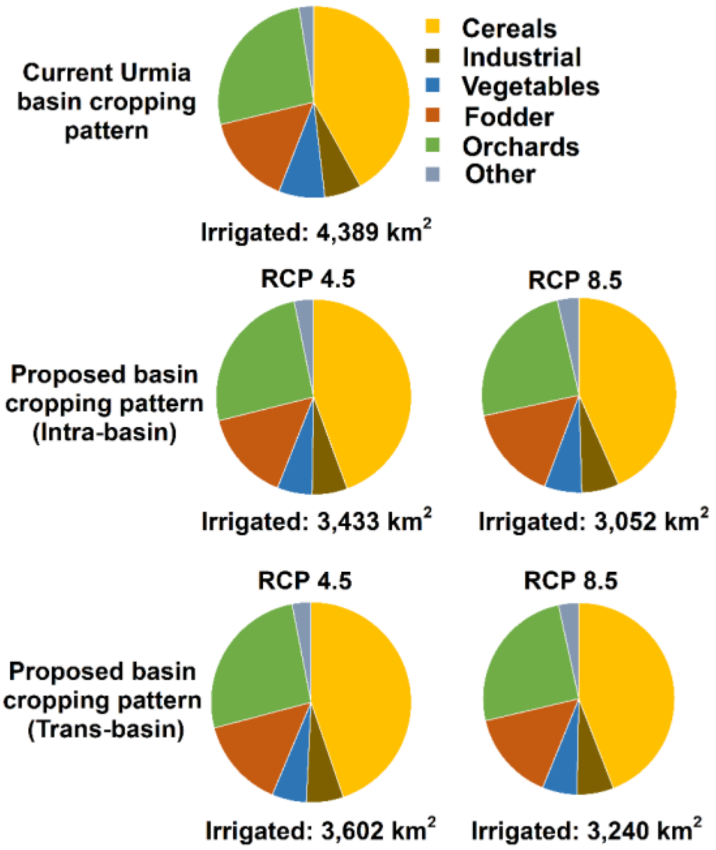


Figure 2-12: Proposed cropping pattern change in the Urmia basin (Cereals: wheat, barely, maize; industrial crops: sugar beet, oil seeds; vegetables: potato, tomato, cucumber, water melons, bean, sweet melon, onion, and other vegetables; fodder: clover, alfalfa, corn; orchards: apples, grapes, stone fruits, pistachio, and others).

The proposed change in land use and cropping patterns would lead to a reduction in GHG emissions (Table 2). This is due to the reduction in the area of land used for cultivation of 21% (RCP 4.5) to 30% (RCP 8.5) for the intra-basin restoration option. As indicated in Table 2, conversion of the irrigated croplands to rain-fed crops and to grasslands would reduce the cumulative GHG emission from 0.07 (0.05 - 0.09) Mt CO_2 eq. to 1.55 (0.65 - 2.44) Mt CO_2 eq., depending on the type, amount, and speed of conversion. On average, changing from irrigated croplands to grasslands saves six times more GHG emissions than converting to rain-fed cropland. As expected, the mitigated GHG emissions are higher over the ten-year restoration period than over 30 years.

It should be noted that the GHG savings refer only to land use change (section 2.3.4.4) and do not include GHG emissions as a result of farming activities. It is also noteworthy that there is a general lack of data on GHG emissions related to alteration of cropping patterns within the irrigated croplands for a region with climatic parameters similar to the Urmia basin. For example, there are no data for the change in GHG emissions due to a change from wheat to sugar beet cultivation. Therefore, with the available data it was not possible to include GHG emissions in the optimisation model and hence the GHG emissions have been estimated at the optimum solution. However, in other saline lakes' basins where detailed data on GHG emissions related to changing a crop pattern are available, the emissions can be included into the optimisation model.

The aforementioned reductions in the total area under cultivation, especially, without water transfer from adjacent basins, would entail some behavioural and institutional challenges that can jeopardise the accomplishment of the land use strategies proposed in this study. This is particularly the case in the Urmia, Gedar, and Aji sub-basins due to the required large-scale land use reduction in the agricultural sector (up to 42%, 35%, and 33%, respectively). Indeed, these strategies go against the current tendency to increase crop production at the farm level and therefore their implementation may meet major opposition. Monetizing the activities involved in conversion of the irrigated lands to grasslands/rangelands or rain-fed systems is difficult with available data; however, investing in entrepreneurial initiatives to reduce the local livelihood dependency on agriculture and industrializing the economy of the Urmia basin would presumably ensure long-term profits.

Table 2-6: Reductions in GHG emissions as a result of the proposed land use change in the agricultural sector over ten- and 30-year periods.

Restoration option/ Scenario		Period	Cropland to rain-fed (Mt CO_2 eq.)			Cropland to grassland (Mt CO_2 eq.)		
			5th percentile[a]	Median	95th percentile[a]	5th percentile[a]	Median	95th percentile[a]
Intra-basin	RCP 4.5	10 years	0.18	0.27	0.36	0.46	1.02	1.60
		30 years	0.06	0.09	0.12	0.47	1.11	1.74
	RCP 8.5	10 years	0.38	0.38	0.50	0.64	1.40	2.37
		30 years	0.08	0.12	0.16	0.65	1.55	2.44
Inter-basin	RCP 4.5	10 years	0.15	0.22	0.29	0.37	0.88	1.39
		30 years	0.05	0.07	0.09	0.39	0.90	1.43
	RCP8.5	10 years	0.22	0.33	0.43	0.55	1.29	2.02
		30 years	0.07	0.10	0.14	0.65	1.54	2.43

[a] Calculated through Monte Carlo simulation.

2.4.4 Method limitations and uncertainties

In addition to the uncertainties in future climate projections, there are some other limitation to the methodologies used in the proposed framework. For instance, there may be some bias in the historical climatic data used as observations to remove the significant bias in the output of GCMs. Use of the data from meteorological stations, combined with more advanced bias-correction methods (e.g. quantile-mapping (Gudmundsson 2016)) may mitigate this issue.

Another limitation is the reference period for calibration of the hydrological balance over the lake's area. Here, the period from 1995 to 2010 was considered with yearly time resolutions, while longer periods with finer time resolutions (e.g. monthly) would improve the accuracy of the results.

One of the sources of uncertainty in the estimations of the dust release from Lake Urmia bed is the data that we used for validation of the vertical dust parametrizations, which are not specific to this lake. Long-term experimental data on the response of the lake's playa to wind erosion are rare and require further laboratory and field studies. Similarly, most of the data that we used as crop tuning parameters were the general crop parameters adopted from Allen et al. (2005). More detailed data from the cultivars in the Urmia basin would improve the simulations of the crop-yield response.

Furthermore, the restored lake and its surrounding wetlands can act as carbon sinks. With the available data, quantifying their potential for carbon sequestration was not possible and was not considered in the study. This could be a subject of future work.

2.5 Conclusions

A new eco-hydrological framework was proposed to pave the way for preservation and restoration of desiccating saline lakes in the face of projected climate change, with the aim of mitigating the catastrophic consequences of their shrinkage. The framework comprises a suit of models used to identify optimal solutions for lake restoration, including global circulation, salt-water, water allocation, and land use and economic optimisation models. Considering the effects of global warming under different climate change scenarios, the framework enables estimations of the amount of water required for either partial or full restoration of saline lakes based on the water availability in the lake's tributaries. It also guides decision- and policy-makers in formulating an optimal land use strategy in the lake's basin with the aim of restoring the lake to recommended ecological levels up to 2050.

To illustrate its capabilities, the framework was applied in the context of a typical saline lake basin, in this case Lake Urmia. This lake used to be the second hyper-saline lake in the world and has faced a drastic water-level decline over the past two decades. The following conclusions can be drawn with respect to the restoration of Lake Urmia, based on the application of the proposed framework:

- The outputs of the GCMs suggest an increase of ~2˚C in the average temperature in the basin by 2050 compared to the reference period (1960 - 1995). A decreasing precipitation trend is expected by 2050, although the uncertainty analysis suggest a spectrum of possibilities, from no change, to an increase to a greater decline in precipitation.
- The mass and water balance model over the lake's area between 1997 and 2010 (shrinkage period) shows the salt precipitation of 4 cm yr^{-1} on the lake bed. The subsurface flow in this period was limited, ranging between 12.8% and 15.4% of the total inflow to the lake of 2,861.3 Mm^3 yr^{-1}.
- The average surface inflow of water required for a complete 30-year lake restoration to the target level of 1274.1 m was estimated at 3,648 Mm^3 $yr^{-1}$for the RCP 4.5 scenario and 3,692 Mm^3 yr^{-1} for RCP 8.5. From a policy perspective, increase in population, maintaining equity among stakeholders, and various financial, behavioural, and institutional barriers impede attaining such surface inflows in the short-term. Instead, partial restoration of the lake by 2030 might be a more feasible option.
- If the soil moisture of the lake bed falls below 5%, there is a potential for the release of 36,391 t yr^{-1} of saline PM_{10} from the lake bed. These emissions would reduce to 3,342 t yr-

^{1} with the soil moisture increasing to 15%. Restoration of the southern half of the lake could mitigate 77% of salt-rich dust emissions from the lake bed, while restoration of the northern and areas below the main southern islands could prevent only 22% and 39% of the total average dust emissions per year, respectively. Accordingly, the restoration of the southern part of the lake is recommended in the short-term (2030) to control the expected PM_{10} emissions from the lake bed.

- An annual water input of 1.83 km^3 (RCP 4.5) and 1.86 km^3 (RCP 8.5) can restore the southern part of the lake to the target level of 1274.1 m in ten years; currently, 1.58 km^3 of water is provided by the southern tributaries.
- At present, the area under cultivation in the lake's basin is close to 438,900 ha. In the case of restoration, including inter-basin water transfers, a reduction in the total irrigated area of 78,700 ha is proposed under RCP 4.5 and 114,826 under RCP 8.5. The conversion of 95,600 ha (RCP 4.5) and 133,687 ha (RCP 8.5) of irrigated lands to rain-fed cropland or grassland is suggested.
- At the end of the 30-year restoration plan, the unit water productivity would be 0.278 US$ m^{-3} for restoration with inter-basin water transfer and 0.304 US$ m^{-3} for the intra-basin restoration scenario (RCP 4.5). The equivalent values for RCP 8.5 are 0.271 US$ m^{-3} and 0.296 US$ m^{-3}, respectively.
- Conversion of the irrigated croplands to rain-fed systems and to grasslands proposed here is expected to reduce cumulative GHG emissions due to the land use change from 0.07 (0.05 - 0.09) Mt CO_2 eq. to 1.55 (0.65 - 2.44) Mt CO_2 eq., depending on the type, amount, and speed of land conversion.

As demonstrated in the paper, the proposed framework provides a comprehensive and powerful tool to aid decision- and policy-makers in identifying optimal solutions for restoring saline lakes at risk of desiccation. The framework is generic enough to be applicable to different regions, subject to data availability. It is recommended that future work explores its application to other saline lakes with the aim of demonstrating further its applicability and improving the methods.

Acknowledgements

Nima Shokri acknowledges the funding by Royal Academy of Engineering-The Leverhulme Trust Senior Research Fellowship (LTSRF1718\14\8). We acknowledge the use of imagery from the NASA Worldview application (https://worldview.earthdata.nasa.gov/) operated by the NASA/Goddard Space Flight Centre Earth Science Data and Information System (ESDIS) project. The authors would like to thank MIKE powered by DHI group for providing the Mike HYDRO basin management framework. The first author's PhD study is supported by Presidential Doctoral Scholarship Award at The University of Manchester which is gratefully acknowledged.

References

JICA, Japan International Cooperation Agency. (2016). *Data collection survey on hydrological cycle of lake urmia basin in the islamic republic of Iran*.

Abbaspour, M., & Nazaridoust, A. (2007). Determination of environmental water requirements of Lake Urmia, Iran: an ecological approach. *International Journal of Environmental Studies, 64*(2), 161-169.

Abuduwaili, J., DongWei, L., & GuangYang, W. (2010). Saline dust storms and their ecological impacts in arid regions. *Journal of arid land, 2*(2), 144-150.

AghaKouchak, A., Norouzi, H., Madani, K., Mirchi, A., Azarderakhsh, M., Nazemi, A., et al. (2015). Aral Sea syndrome desiccates Lake Urmia: call for action. *Journal of Great Lakes Research, 41*(1), 307-311.

Alborzi, A., Mirchi, A., Moftakhari, H., Mallakpour, I., Alian, S., Nazemi, A., et al. (2018). Climate-informed environmental inflows to revive a drying lake facing meteorological and anthropogenic droughts. *Environmental Research Letters, 13*(8), 084010.

Alipour, S., Mosavi-ovenlegi, K., Hosseini, E., Aslanpour, S., & Haseli, Z. (2018). Geochemistry of major, trace and rare earth elements in bed-sediments of Urmia lake. *Journal of Geoscience, 27*(107), 51-62.

Allen, R. G., Pereira, L. S., Smith, M., Raes, D., & Wright, J. L. (2005). FAO-56 dual crop coefficient method for estimating evaporation from soil and application extensions. *Journal of irrigation and drainage engineering, 131*(1), 2-13.

Altinbilek, D. (2004). Development and management of the Euphrates–Tigris basin. *International Journal of Water Resources Development, 20*(1), 15-33.

Cameron, D. R., Marvin, D. C., Remucal, J. M., & Passero, M. C. (2017). Ecosystem management and land conservation can substantially contribute to California's climate mitigation goals. *Proceedings of the National Academy of Sciences, 114*(48), 12833-12838.

Chaudhari, S., Felfelani, F., Shin, S., & Pokhrel, Y. (2018). Climate and anthropogenic contributions to the desiccation of the second largest saline lake in the twentieth century. *Journal of Hydrology, 560*, 342-353.

Dalal, R., & Chan, K. (2001). Soil organic matter in rainfed cropping systems of the Australian cereal belt. *Soil Research, 39*(3), 435-464.

Darvishi, J. (2014). *Lymphology and Paleolymology Report of Lake Urmia: The balance of water and salt and the sedimentation rate of salts, Final report (in persian)*.

Deatrick, E. (2016). Can a controversial canal stop thousands of sinkholes from forming around the Dead Sea. *Science, 22*.

Diaz, D., Rashford, B., De Gryze, S., Zakreski, S., & Dell, R. (2012). Evaluation of Avoided Grassland Conversion and Cropland Conversion to Grassland as Potential Carbon Offset Project Types. *Portland, Oregon: The Climate Trust*.

Draxler, R. R. (1999). *HYSPLIT radiological transport and dispersion model implementation on NCEP Cray*: Silver Spring, Maryland : U.S. Department of Commerce, National

Oceanic and Atmospheric Administration, National Weather Service, Office of Meteorology, Science Division, 1999.

Durán, O., Claudin, P., & Andreotti, B. (2011). On aeolian transport: Grain-scale interactions, dynamical mechanisms and scaling laws. *Aeolian research, 3*(3), 243-270.

Emori, S., Taylor, K., Hewitson, B., Zermoglio, F., Juckes, M., Lautenschlager, M., et al. (2016). CMIP5 data provided at the IPCC Data Distribution Centre. In: Geneva.

Fathian, F., Morid, S., & Kahya, E. (2015). Identification of trends in hydrological and climatic variables in Urmia Lake basin, Iran. *Theoretical and Applied Climatology, 119*(3-4), 443-464.

Fécan, F., Marticorena, B., & Bergametti, G. (1998). *Parametrization of the increase of the aeolian erosion threshold wind friction velocity due to soil moisture for arid and semi-arid areas.* Paper presented at the Annales Geophysicae.

Food, & Agriculture Organization of the United, N. (1998). FAOSTAT statistics database: Rome : FAO, 1998.

García-Vila, M., & Fereres, E. (2012). Combining the simulation crop model AquaCrop with an economic model for the optimisation of irrigation management at farm level. *European Journal of Agronomy, 36*(1), 21-31.

Garousi, V., Najafi, A., Samadi, A., Rasouli, K., & Khanaliloo, B. (2013). Environmental crisis in Lake Urmia, Iran: a systematic review of causes, negative consequences and possible solutions. *Proceedings of the 6th International Perspective on Water Resources & the Environment (IPWE) Izmir, Turkey.*

Ghale, Y. A. G., Altunkaynak, A., & Unal, A. (2018). Investigation anthropogenic impacts and climate factors on drying up of Urmia Lake using water budget and drought analysis. *Water resources management, 32*(1), 325-337.

Gillette, D. A. (1974). On the production of soil wind erosion aerosols having the potential for long range transport. *J. Rech. Atmos, 8*(3-4), 735-744.

Gillette, D. A., Fryrear, D., Xiao, J. B., Stockton, P., Ono, D., Helm, P. J., et al. (1997). Large-scale variability of wind erosion mass flux rates at Owens Lake: 1. Vertical profiles of horizontal mass fluxes of wind-eroded particles with diameter greater than 50 μm. *Journal of Geophysical Research: Atmospheres, 102*(D22), 25977-25987.

Gudmundsson, L. (2016). qmap: Statistical transformations for post-processing climate model output. *R package version 1.0-4.*

Hamidi-Razi, H., Mazaheri, M., Carvajalino-Fernández, M., & Vali-Samani, J. (2018). Investigating the restoration of Lake Urmia using a numerical modelling approach. *Journal of Great Lakes Research.*

Hammer, U. T. (1986). *Saline lake ecosystems of the world* (Vol. 59): Springer Science & Business Media.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International journal of climatology, 34*(3), 623-642.

Hasemi, M. (2011). *A socio-technical assessment framework for integrated water resources management (IWRM) in Lake Urmia Basin, Iran.* Newcastle University,

Hassanzadeh, E., Zarghami, M., & Hassanzadeh, Y. (2012). Determining the main factors in declining the Urmia Lake level by using system dynamics modeling. *Water Resources Management, 26*(1), 129-145.

Hawkins, E., Osborne, T. M., Ho, C. K., & Challinor, A. J. (2013). Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. *Agricultural and Forest Meteorology, 170*, 19-31.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one, 12*(2), e0169748.

Hobbs, R. J., & Norton, D. A. (1996). Towards a conceptual framework for restoration ecology. *Restoration ecology, 4*(2), 93-110.

Houghton, R. A. (2003). Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. *Tellus B, 55*(2), 378-390.

Jalili, S., Hamidi, S. A., & Namdar Ghanbari, R. (2016). Climate variability and anthropogenic effects on Lake Urmia water level fluctuations, northwestern Iran. *Hydrological sciences journal, 61*(10), 1759-1769.

Kang, J. Y., Yoon, S. C., Shao, Y., & Kim, S. W. (2011). Comparison of vertical dust flux by implementing three dust emission schemes in WRF/Chem. *Journal of Geophysical Research: Atmospheres, 116*(D9).

Kawamura, R. (1951). *Study of sand movement by wind Translated.* (1965) as University of California, Hydraulics Engineering Laboratory Report HEL 2-8, Berkely.

Kelts, K., & Shahrabi, M. (1986). Holocene sedimentology of hypersaline Lake Urmia, northwestern Iran. *Palaeogeography, Palaeoclimatology, Palaeoecology, 54*(1-4), 105-130.

Kok, J. F., Parteli, E. J., Michaels, T. I., & Karam, D. B. (2012). The physics of wind-blown sand and dust. *Reports on progress in Physics, 75*(10), 106901.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift, 15*(3), 259-263.

Lensky, N., Dvorkin, Y., Lyakhovsky, V., Gertman, I., & Gavrieli, I. (2005). Water, salt, and energy balances of the Dead Sea. *Water Resources Research, 41*(12).

Lettau, K., & Lettau, H. (1978). Experimental and micrometeorological field studies of dune migration. In H. H. Lettau & K. Lettau (Eds.), *Exploring the world's driest climate*: Madison.

Levy, R., & Hsu, C. (2015). Modis atmosphere l2 aerosol product, NASA modis adaptive processing system, goddard space flight center, USA.

Lu, H., & Shao, Y. (1999). A new model for dust emission by saltation bombardment. *Journal of Geophysical Research: Atmospheres, 104*(D14), 16827-16842.

Madani, K. (2014). Water management in Iran: what is causing the looming crisis? *Journal of Environmental Studies and Sciences, 4*(4), 315-328. journal article.

Mathworks, T. Matlab Optimization Toolbox User's Guide.

McCune, D. (1991). Effects of airborne saline particles on vegetation in relation to variables of exposure and other factors. *Environmental Pollution, 74*(3), 176-203.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature communications, 7*, 13603.

Micklin, P. (2007). The Aral sea disaster. *Annu. Rev. Earth Planet. Sci., 35*, 47-72.

Mike. (2017). Hydro Basin User Guide.

Miller, J. R. (1987). The Political Economy of Western Water Finance: Cost Allocation and the Bonneville Unit of the Central Utah Project. *American Journal of Agricultural Economics, 69*(2), 303-310.

Mohammed, I. N., & Tarboton, D. G. (2012). An examination of the sensitivity of the Great Salt Lake to changes in inputs. *Water Resources Research, 48*(11).

Mor, Z., Assouline, S., Tanny, J., Lensky, I., & Lensky, N. (2018). Effect of water surface salinity on evaporation: The case of a diluted buoyant plume over the Dead Sea. *Water Resources Research, 54*(3), 1460-1475.

Nickling, W. G., & Brown, L. J. (2001). *PM10 Dust Emissions At Owen Lake, CA 2001, Final Report*.

Owen, P. R. (1964). Saltation of uniform grains in air. *Journal of Fluid Mechanics, 20*(2), 225-242.

Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). AquaCrop—the FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agronomy Journal, 101*(3), 438-447.

Rahi, K. A., & Halihan, T. (2018). Salinity evolution of the Tigris River. *Regional Environmental Change, 18*(7), 2117-2127.

Raupach, M., Gillette, D., & Leys, J. (1993). The effect of roughness elements on wind erosion threshold. *Journal of Geophysical Research: Atmospheres, 98*(D2), 3023-3029.

Razia, H. H., Mazaherib, M., Samanic, J. M. V., & Fernandezd, M. C. (2016). Investigating Urmia Lake Partial Restoration and Ecological Water Level Using MOHID-2D Water Hydrodynamic Model. Geographic and Environmental Impacts of Urmia Lake Conditions Conference, Tabriz, Iran.

Rolph, G., Stein, A., & Stunder, B. (2017). Real-time environmental applications and display system: Ready. *Environmental Modelling & Software, 95*, 210-228.

Roney, J. A., & White, B. R. (2006). Estimating fugitive dust emission rates using an environmental boundary layer wind tunnel. *Atmospheric Environment, 40*(40), 7668-7685.

Ryan, E. (2015). The Public Trust Doctrine, Private Water Allocation, and Mono Lake: The Historic Saga of National Audubon Society v. Superior Court. *Envtl. L., 45*, 561.

Schaufler, G., Kitzler, B., Schindlbacher, A., Skiba, U., Sutton, M., & Zechmeister-Boltenstern, S. (2010). Greenhouse gas emissions from European soils under different land use: effects of soil moisture and temperature. *European Journal of Soil Science, 61*(5), 683-696.

Service, U. S. D. o. A. S. C. (1999). *Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys*: US Department of Agriculture.

Shadkam, S., Ludwig, F., van Oel, P., Kirmit, Ç., & Kabat, P. (2016). Impacts of climate change and water resources development on the declining inflow into Iran's Urmia Lake. *Journal of Great Lakes Research, 42*(5), 942-952.

Shao, Y. (2001). A model for mineral dust emission. *Journal of Geophysical Research: Atmospheres, 106*(D17), 20239-20254.

Shao, Y. (2004). Simplification of a dust emission scheme and comparison with data. *Journal of Geophysical Research: Atmospheres, 109*(D10).

Shao, Y., Jung, E., & Leslie, L. M. (2002). Numerical prediction of northeast Asian dust storms using an integrated wind erosion modeling system. *Journal of Geophysical Research: Atmospheres, 107*(D24).

Shao, Y., & Lu, H. (2000). A simple expression for wind erosion threshold friction velocity. *Journal of Geophysical Research: Atmospheres, 105*(D17), 22437-22443.

Shokri-Kuehni, S. M., Rad, M. N., Webb, C., & Shokri, N. (2017). Impact of type of salt and ambient conditions on saline water evaporation from porous media. *Advances in water resources, 105*, 154-161.

Shokri-Kuehni, S. M., Vetter, T., Webb, C., & Shokri, N. (2017). New insights into saline water evaporation from porous media: Complex interaction between evaporation rates, precipitation, and surface temperature. *Geophysical Research Letters, 44*(11), 5504-5510.

Smith, P. (2008). Land use change and soil organic carbon dynamics. *Nutrient Cycling in Agroecosystems, 81*(2), 169-178.

Sørensen, M. (2004). On the rate of aeolian sand transport. *Geomorphology, 59*(1-4), 53-62.

Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agronomy Journal, 101*(3), 426-437.

Thiessen, A. (1911). Precipitation for large areas. *Monthly Weather Rev*. 39, 1082-1084.

White, B. R. (1979). Soil transport by winds on Mars. *Journal of Geophysical Research: Solid Earth, 84*(B9), 4643-4651.

Williams, W. D. (1996). What future for saline lakes? *Environment: Science and Policy for Sustainable Development, 38*(9), 12-39.

Williams, W. D. (2002). Environmental threats to salt lakes and the likely status of inland saline ecosystems in 2025. *Environmental conservation, 29*(2), 154-167.

Winter, T. C., Buso, D. C., Rosenberry, D. O., Likens, G. E., Sturrock, A. J. M., & Mau, D. P. (2003). Evaporation determined by the energy-budget method for Mirror Lake, New Hampshire. *Limnology and Oceanography, 48*(3), 995-1009.

WRMC, Water Resources Mangemnet Company (2007). *Master Plan Report on Agricultural Water in the Lake Urmia Basin (In persian)*.

Wurtsbaugh, W. A., Miller, C., Null, S. E., DeRose, R. J., Wilcock, P., Hahnenberger, M., et al. (2017). Decline of the world's saline lakes. *Nature Geoscience, 10*(11), 816.

Zilberman, T., Gavrieli, I., Yechieli, Y., Gertman, I., & Katz, A. (2017). Constraints on evaporation and dilution of terminal, hypersaline lakes under negative water balance: The Dead Sea, Israel. *Geochimica et Cosmochimica Acta, 217*, 384-398.

Chapter 3 Predicting Long-term Dynamics of Soil Salinity and Sodicity on a Global Scale

Amirhossein Hassani[a], Adisa Azapagic[a*], and Nima Shokri[b*]

[a] Department of Chemical Engineering and Analytical Science, The University of Manchester, Sackville Street, Manchester M13 9PL, UK.
[b] Hamburg University of Technology, Institute of Geo-Hydroinformatics, Am Schwarzenberg-Campus 3 (E), 21073 Hamburg, Germany

Proceedings of the National Academy of Sciences, https://www.pnas.org/content/early/2020/12/09/2013771117
*Corresponding authors: adisa.azapagic@machester.ac.uk; nima.shokri@tuhh.de

Abstract

Knowledge of spatio-temporal distribution and likelihood of (re)occurrence of salt affected soils is crucial to our understanding of land degradation and for planning effective remediation strategies in face of future climatic uncertainties. However, conventional methods used for tracking the variability of soil salinity/sodicity are extensively localised, making predictions on the global scale difficult. Here, we employ machine learning techniques and a comprehensive set of climatic, topographic, soil, and remote sensing data to develop models capable of making predictions of soil salinity (expressed as electrical conductivity of saturated soil extract) and sodicity (measured as soil exchangeable sodium percentage) at different longitudes, latitudes, soil depths, and time periods. Using these predictive models, we provide the first global-scale quantitative and gridded dataset characterising different spatio-temporal facets of soil salinity and sodicity variability over the past four decades at a ~1 km resolution. Analysis of this dataset reveals that a soil area of 11.73 Mkm^2 located in non-frigid zones has been salt-affected with a frequency of reoccurrence in at least three-fourths of the years between 1980 and 2018, with 0.16 Mkm^2 of this area being croplands. Although the net changes in soil salinity/sodicity and the total area of salt-affected soils have been geographically highly variable, the continents with the highest salt-affected areas are Asia (particularly China, Kazakhstan, and Iran), Africa, and Australia. The proposed method can also be applied for quantifying the spatio-temporal variability of other dynamic soil properties, such as soil nutrients, organic carbon content, and pH.

Keywords: Soil salinization; Soil salinity; Soil sodicity; Global scale modelling; Machine Learning; Global trends.

Significance Statement

Land degradation due to soil salinization has detrimental impacts on vegetation, crops, and human livelihoods, leading to a need for a methodologically consistent analysis of the variability of different aspects of salt-affected soils. However, previous studies on the soil salinity issue have been primarily spatial and localised, leaving the large-scale spatio-temporal variations of soil salinity widely ignored. To address this gap, we present the first globally validated analysis quantifying the long-term variations (40 years) of the top soil salinity at high spatial resolutions using machine learning techniques. The results have significant implications for agro-ecological modelling, land assessment, crop growth simulation, and sustainable water management.

3.1 Introduction

Soil salinization is one of the main land-degrading threats influencing soil fertility, stability, and bio-diversity. Saline soils are the ones with excess accumulation of soluble salts in the root zone (Abrol et al. 1988). On the other hand, accumulation of high levels of sodium salt relative to other exchangeable cations is the main attribute of sodic soils (Bleam 2016). Wind, rainfall, and parent rock weathering are the main origins of these salts in “primary” soil salinization whereas in “secondary” soil salinization, excessive salt accumulation is human-induced (Rengasamy 2006). Saline and sodic soils, or in general salt-affected soils, mostly lie across arid and semi-arid climates where the dominance of evaporation over precipitation concentrates the salts in the root zone (Abrol et al. 1988; Ponnamperuma 1984), leading to undesirable alterations in the physical, chemical, and biological functions of the soil (Metternicht 2016; Daliakopoulos et al. 2016). Sodicity adversely influences the soil infiltration capacity (Wong et al. 2010), increases the susceptibility of water and wind-blown erosion (De la Paix et al. 2013), and exposes more soil organic matter to decomposing processes (Singh 2016). Soil salinity, on the other side, distresses the soil respiration, nitrogen cycle, and decomposing functionality of soil microorganisms (Singh 2016; Rath et al. 2015). Salinity stress affects the vegetation growth directly by reducing the plant water uptake (osmotic stress) and/or by deteriorating the transpiring leaves (specific ion effects) (Parihar et al. 2015), in turn reducing organic input to the soil and ultimately leading to desertification of lands (Sentis 1996; Perri et al. 2020). Under extreme conditions, dispersion of saline dust (De la Paix et al. 2013; Hassani et al. 2020), poverty, migration, and high costs of soil reclamation are long-term socio-economic consequences of soil salinization (Zaman et al. 2018).

Soil salinity and sodicity levels are spatially, vertically, and temporally dynamic (Zaman et al. 2018; Mulder et al. 2011), particularly at the top 0 - 30 cm soil layer which is substantially affected by governing climatic conditions. Naturally-occurring events, such as flash floods, El Niño and La Niña, alternative wet and dry years, and long periods of droughts can considerably affect soil salinization and accumulation/leaching of the salts in/from the root zone at daily to multi-year temporal resolutions. Similarly, anthropogenic activities like irrigation and dryland management can affect soil salinization at different temporal resolutions. Given the high dynamism in soil salinization processes, updated spatial and temporal information on the extent of salt-affected soils is indispensable for devising appropriate sustainable action programmes for managing land and soil resources (Daliakopoulos et al. 2016; Ivushkin et al. 2019; Oldeman et al. 2017; Fao/Iiasa/Isric/Isscas/Jrc 2012). This information can be also valuable for enhancing our understanding of terrestrial carbon dynamics (Wong et al. 2010; Setia et al. 2013), food security and agricultural modelling (Butcher et al. 2016; Shani et al. 2007), climate change impacts (Várallyay 1994; Suweis et al. 2010), water resources and irrigation management (Fan et al. 2012; Mateo-Sagasta et al. 2011), and efficiency of organic/inorganic reclamation practices (Amini et al. 2016; Mau et al. 2016). Several statistics on the global distribution of salt-affected soils (Ivushkin et al. 2019; Oldeman et al. 2017; Fao/Iiasa/Isric/Isscas/Jrc 2012; Szabolcs 1989; Squires et al. 2011; Ghassemi et al. 1995; FAO 2000; Schofield et al. 2003) have been generated based on data from soil surveys and statistical extrapolation (Abrol et al. 1988; Fao/Iiasa/Isric/Isscas/Jrc 2012); yet these estimations are mainly purely spatial (Ivushkin et al. 2019; Hengl et al. 2017a), not necessarily up-to-date (Ivushkin et al. 2019; Zaman et al. 2018), and in some cases incomparable (Montanarella et al. 2015; Rengasamy 2006). Therefore, there is still a need for a methodologically consistent dataset documenting long-term variations of the soil salinity and sodicity at high spatial resolutions (FAO Soils Portal 2020).

To address this need, we focused on two target variables: ground-derived measurements of soil EC_e (the ability of a water-saturated soil paste extract to conduct electrical current: representative of salinity severity) and ESP (exchangeable sodium percentage: representative of sodicity severity). We used 42,984 and 197,988 data, respectively, scattered over time from 1980 - 2018. We trained two-part predictive models for making 4-D predictions of soil salinity and sodicity as target variables (longitude, latitude, soil depth, and time; see Methods). Through mapping data-driven relations between soil EC_e/ESP observations and a collection of associated predictors generated from topographic, climatic, vegetative, soil, and landscape properties of the sampling locations (Appendix 2, Table A2-1), these two-part models enabled us to make long-term gridded predictions of soil salinity and sodicity at new locations with available predictors' values. Note that "prediction" refers to the estimation by the trained models of soil salinity/sodicity on a global scale from 1980 to 2018 even in locations where there is no measurement available rather than to future projection of soil salinity/sodicity on the basis of current trends. The first part of the models classified the soil into saline/sodic and non-saline/non-sodic classes (binary classification) and the second part predicted per-class severity of the salinity/sodicity issue (regression). Meaningful statistics derived from the EC_e and ESP predictions were then used to generate univariate thematic maps of the variability of different aspects of soil salinity/sodicity between 1980 and 2018 at ~1 km spatial resolution (30 arc-seconds; e.g. Figure 3-1). These were delimited to -55° and 55° latitudes, comprising tropics, subtropics, and temperate zones (see the Data Availability statement). We focused on the top soil layer (or surface soil) referring to the top 30 cm of the soil profile measured from the surface.

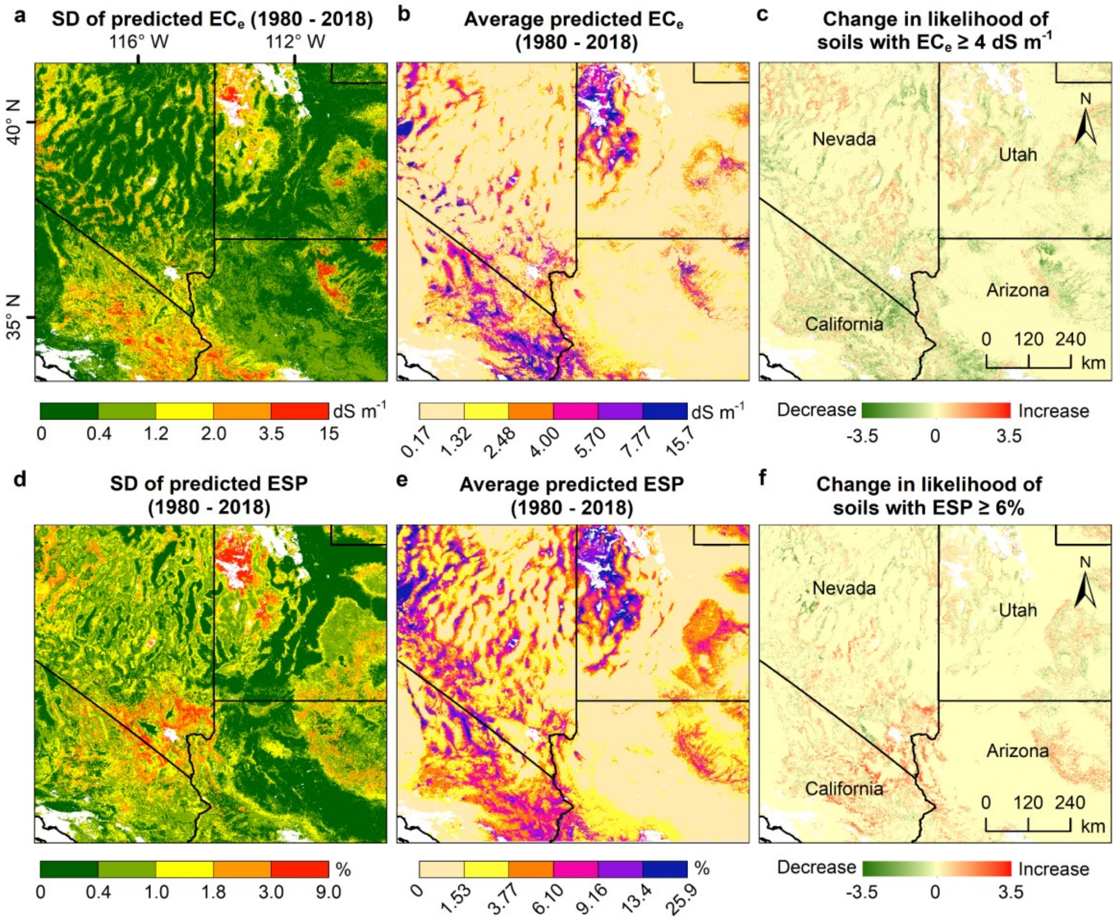


Figure 3-1: Variability of different aspects of soil salinity and sodicity in western USA. a and **d**: Standard deviation (SD) of annually predicted soil salinity (EC_e) and sodicity (ESP), respectively, between 1980 and 2018. **b** and **e**: Average of annually predicted EC_e and ESP, respectively (1980 - 2018). **c** and **f**: Change in the likelihood

(θ) of soils with an $EC_e \geq 4$ dS m^{-1} or ESP $\geq 6\%$ in the period 2000 - 2018 relative to 1981 - 1999 (the likelihood is dimensionelss, calculated by dividing the number of years with $EC_e \geq 4$ dS m^{-1} or ESP $\geq 6\%$ by the total number of years in the studied period). Positive θ indicates that the likelihood has increased and negative shows that it has decreased.

3.2 Validation of the predictive models

Appendix 2, Figure A2-1 (a to d) and Appendix 2, Table A2-2 illustrate the performance of the two-part fitted models in prediction of target variables. During the training of the classifier, any soil with $EC_e \geq 2$ dS m^{-1} and ESP $\geq 1\%$ was labelled as saline and sodic class, respectively. The overall accuracy for the saline/non-saline soil classifier evaluated by 10-fold cross validation (10-CV) was 89.65% (88.33% - 88.87%) and for the sodic/non-sodic soil classifier, it was 85.59% (85.05% - 85.24%); the values in parenthesises show the lower and upper bounds for 95% confidence intervals. The average per-class User's Accuracies (probability that predictions represent reality) for the salinity classifier was 88.3% and for the sodicity classifier 85.5%. The prediction errors evaluated by 10-CV normalised root-mean-square (normalised by range) was 8.82% (9.02% - 9.17%) for the regression model fitted to observations in the saline class and 6.94% (7.09% - 7.20%) for the regression model fitted to the sodic class.

To further evaluate the performance of our models, we compared our predicted soil surface EC_e/ESP with the corresponding EC_e/ESP outcomes of the often-cited global dataset of soil salinity/sodicity: Harmonised World Soil Database (Fao/Iiasa/Isric/Isscas/Jrc 2012) (HWSD; Appendix 2, Figure A2-1, e and f). To do so, we evaluated the outputs of our predictive models and HWSD surface estimations of EC_e and ESP against the available measured surface values of EC_e and ESP. Any available EC_e or ESP measurement from 1980 with zero upper sample's depth and a maximum lower sample's depth equal to 30 cm was used in this analysis. The coefficient of determination (R^2) between the predictions of our two-part model and 9,293 measured surface values of EC_e was 0.83, while for HWSD it was 0.12. Likewise, R^2 between 30,491 surface measurements of the ESP and our predictions was 0.86, while it was 0.26 for HWSD.

Moreover, we investigated the relationship between the catchment-level average of soil salinity estimations for three continents of Australia, Africa, and North America predicted by our trained models and the Dryness Index (the ratio of long-term potential evapotranspiration to rainfall) with the results presented in Appendix 2, Figure A2-2. This figure shows higher predicted salinities in drier climates (locations with higher Dryness Index) where excessive evapotranspiration leads to accumulation of the soluble salts in the soil root zone. The trend observed in Appendix 2, Figure A2-2 is in agreement with the physically-based modelling results reported in Porporato et al. (2015) for estimation of primary soil salinity in the soil root zone as a function of the Dryness Index. Appendix 2, Figure A2-2 provides additional verification of the validity our model predictions.

3.3 Importance of predictors

The importance of each predictor in the models developed in this study as well as how the predicted target variables depend partially on these predictors were investigated, which provided some mechanistic insights on possible influential parameters involved in soil salinization processes (Appendix 2, Figure A2-3; Appendix 2, Table A2-5). In general, soil classification, depth, Fraction of Absorbed Photo-synthetically Active Radiation (FAPAR) as a vegetation cover indicator, and temperature of different soil layers were the predictors highly correlated with target variables. Among 43 predictors, the most important predictors in estimation of EC_e values were FAPAR (10%), lower sample's depth (6.69%), soil's layer four (indicating the layer of soil lying between 100 - 289 cm below the surface) temperature (5.93%), soil clay content (5.68%), and the World Reference Base (WRB) soil classes (5.63%). From various WRB soil classes, the predicted salinity of Haplic Kastanozems and Haplic Leptosols was the highest. On the other hand, for prediction of ESP, the most significant predictors were WRB soil classes (15.96%), lower sample's depth (8.27%), upper sample's depth (7.18%), FAPAR (3.43%), and soil's layer three (indicating the layer of soil lying between 28 - 100 cm below the surface) temperature (2.69%). Also, Gleyic Podzols and Haplic Podzols showed the highest levels of predicted sodicity among the WRB soil classes. Our results suggest that FAPAR can be a better index for mapping soil salinity than Normalised Difference Vegetation Index (NDVI), which has been conventionally used as an indirect remote sensing indicator of soil salinity (Allbed et al. 2013; Daliakopoulos et al. 2016). Partial Dependency Plots (Appendix 2, Figure A2-3) show how the main individual parameters involved in soil salinization processes, e.g. climate, soil temperature, water table depth, and vegetation will affect the estimated values of the soil salinity/salinity, by marginalizing over the other predictors. These make the results suitable for evaluation of the risk of soil salinization in response to future change in key drivers of soil salinity, such as future climates and land cover.

3.4 Variability of soil salinity/sodicity

Traditionally, threshold values of EC_e and ESP have been used as primary indicators for distinguishing saline, sodic, and saline-sodic soils (showing properties of both saline and sodic soils) (Rengasamy 2006; Burt 2011). However, depending on the soil classification system, threshold values can be 4 (Abrol et al. 1988; Richards 1954), 15 (Gupta et al. 2008) (Solonchaks), or even 30 (Soil Survey Staff 2010) (salic) dS m^{-1} for EC_e and 6% (Northcote et al. 1972; Isbell 2016) or 15% (Richards 1954; Gupta et al. 2008; Soil Survey Staff 2010) (Solonetz or natric) for ESP. In addition, the distinguishing characteristics of saline and sodic soils are not limited only to the values of EC_e and ESP and other soil physio-chemical properties, such as pH, salt contents, SAR (Sodium Absorption Ratio), and permeability, should be taken into consideration (Abrol et al. 1988; Szabolcs 1989). For example, the Soil Science Society of America (SSSA) (Soil Science Glossary Terms Committee 2008) defines sodic soils as non-saline soils with enough concentrations of exchangeable sodium that can adversely affect crop productivity with a saturation extract SAR ≥ 13, rather than adopting any ESP threshold. Therefore, in the present study, we quantified variability in areas affected by salinity and sodicity by focusing only on soils' EC_e and ESP. An EC_e equal to 4 dS m^{-1} and an ESP equal to 6% were considered the critical thresholds, corresponding to the lower agronomic limits tolerable by crops (Fao/Iiasa/Isric/Isscas/Jrc 2012). Note that (re)occurrence of a soil with high salinity in a year means the salinity of that soil in that particular year is ≥ 4 dS m^{-1}. Similarly, (re)occurrence of a soil with high sodicity means the ESP of that soil in that particular year is ≥ 6%. Additionally, we assumed soils at a particular location are salt-affected if the annual predicted EC_e of that location is ≥ 4 dS m^{-1} and/or its predicted ESP is ≥ 6% in at least 75% of the years between 1980 and 2018. It should also be noted that all the statistics on salt-affected soils provided here were computed for the world's non-frigid zones, located in the latitudes between -55° and 55°.

Based on the calculated likelihood of annual reoccurrence of salt-affected soils (Figure 3-2 and Appendix 2, Figure A2-8 to Figure A2-12; ranges between zero and one), we estimated that an area of 5.9 Mkm^2 had an EC_e ≥ 4 dS m^{-1} in at least three-fourths of the 1980 - 2018 period. Assuming 2 dS m^{-1} as the lower tolerable limit of salinity, this area increases to 7.62 Mkm^2. During that period, however, an area of 9.18 Mkm^2 had an ESP ≥ 6% in at least three-fourths of the years; this area would reduce drastically to 0.13 Mkm^2 if the threshold value for sodicity were fixed at 15%. Globally, the likelihood of reoccurrence of soils with EC_e ≥ 4 dS m^{-1} in the 2000 to 2018 period was 0.94 of the 1981 to 1999 period (Appendix 2, Figure A2-

4). This value was 0.97 for the soils with ESP ≥ 6%. In total, we estimate that an area of 11.737 Mkm^2 was salt-affected in the 1980 - 2018 period. Note that this is ~25% higher than the often-cited approximation of Szabolcs (1989) and 41% greater than the FAO's estimation in 2000 (Martinez-Beltran 2005).

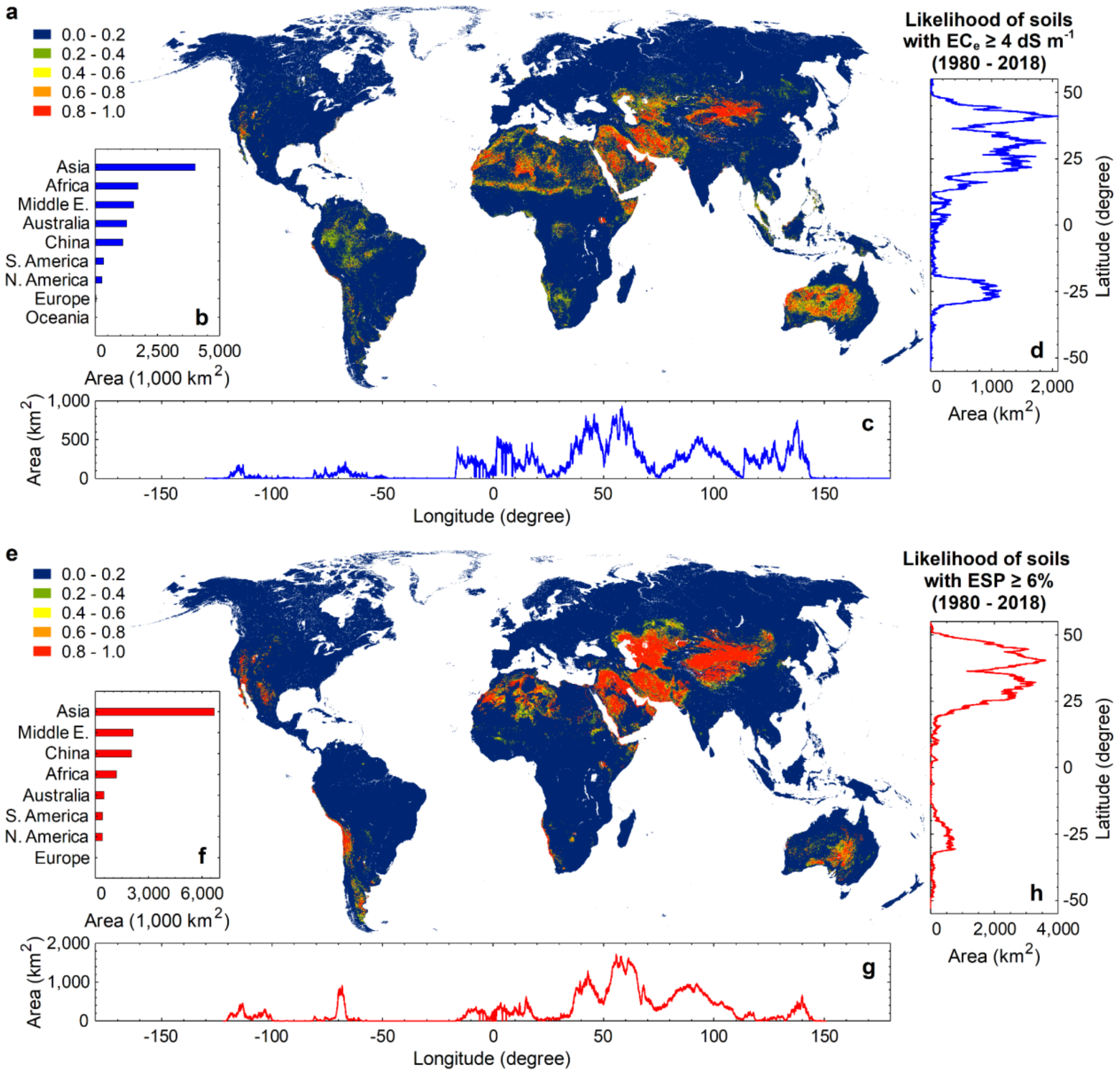


Figure 3-2: Global distribution of salt-affected soils (excluding the frigid zones). a and **e**: Likelihood of the surface soils with an $EC_e \geq 4$ dS m^{-1} and ESP ≥ 6% between 1980 and 2018, respectively (the likelihood is dimensionelss, calculated by dividing the number of years with $EC_e \geq 4$ dS m^{-1} or ESP ≥ 6% by the total number of studied years). The panels on the right (**d** and **h**) and below (**c** and **g**) the maps show the total area of soils with an annual predicted $EC_e \geq 4$ dS m^{-1} and ESP ≥ 6%, respectively, in at least 75% of the period between 1980 and 2018 for different longitudes and latitudes at 30 arc-second resolution (~1 km). **b** and **f**: Total area of the soils with an annual predicted $EC_e \geq 4$ dS m^{-1} and ESP ≥ 6%, respectively, in at least 75% of the 1980 - 2018 period at the continental level.

At the continental level, Asia (including the Middle East) had the largest area of salt-affected soils with 7.14 Mkm^2, followed by Africa with 2.292 Mkm^2, Australia and Oceania with 1.313 Mkm^2, South America with 0.527 Mkm^2, North America with 0.422 Mkm^2, and Europe with 0.024 Mkm^2. In terms of the area of salt-affected lands, the top ranking countries were China with 211.74 Mha, Australia with 131.40 Mha, Kazakhstan with 93.31 Mha, and Iran with 88.33 Mha (Appendix 2, Table A2-3).

Our analysis showed that globally 16.49 Mha of the salt-affected lands were located on croplands over the 1980 - 2018 period. This represents 0.88% of the global cultivated area in 2015, according to the GFSAD30CE V001 dataset (https://croplands.org/home). Cropland was considered here as any stretch of the land with at least 60% cultivated area from 1993 to 2018 and no distinction was made between irrigated and non-irrigated croplands. Our estimated value was 31.3% to 62.7% (7.52 to 28.25 Mha) lower than in the previous assessments (Sandra 1989; Ghassemi et al. 1995), although those focused on the world's irrigated lands. A large majority (536.1 Mha) of the salt-affected areas were located in barren areas (Appendix 2, Table A2-4). The next most salt-affected land cover types were open shrublands (144.12 Mha; dominated by woody perennials 1 - 2 m height, 10 - 60% cover) and grasslands (77.37 Mha). At 10.16 Mha, evergreen broadleaf forests had the largest salt-affected area among different forested land cover types. At the biome level, 928.23 Mha of the salt-affected lands were in deserts and xeric shrublands, followed by montane grasslands and shrublands (86.45 Mha). With respect to climatic conditions, 92% of the salt-affected areas were located in the regions with arid climate and 4.72% in polar tundra. The latter are mostly located in northwest China and north of Himalaya and have high levels of the sodicity.

Only South America with ~9,466 km^2 yr^{-1} had a statically significant increasing trend in the total area of soils with $EC_e \geq 4$ dS m^{-1} ($p < 0.05$; Figure 3-3 and Appendix 2, Table A2-20). However, all continents with a statistically significant trend in the area of soils with ESP $\geq 6\%$ showed an increasing trend; the highest rate of increase was found for Asia with ~5,616 km^2 yr^{-1} ($p < 0.05$; Appendix 2, Table A2-21). Although the strong regional variations are obscured by continental summaries, the overall observed trends and fluctuations may be related to the complex coupling between the surface soil salinity and multi-year climatic patterns or extreme environmental events. For instance, the substantial fluctuations of the salt-affected areas in Australia over relatively short time periods from 1998 to 2015 may be associated with continent-wide variations of the hydrology between dry and wet periods as a result of the El Niño-Southern Oscillation Cycle (Van Dijk et al. 2013) (Figure 3-3, c and i). Particularly in arid and semi-arid regions, the fluctuations in salinity levels can be confirmed by the stochastic salinization model of Suweis et al. (2010). Assuming constant soil and vegetation properties they concluded, for instance, that the probability of having a soil with root zone salinity > 4 dS m^{-1} with the rainfall frequency of 0.15 d^{-1} was approximately four times higher than the rainfall frequency of 0.2 d^{-1} (with mean rainfall depth of 1.79 cm).

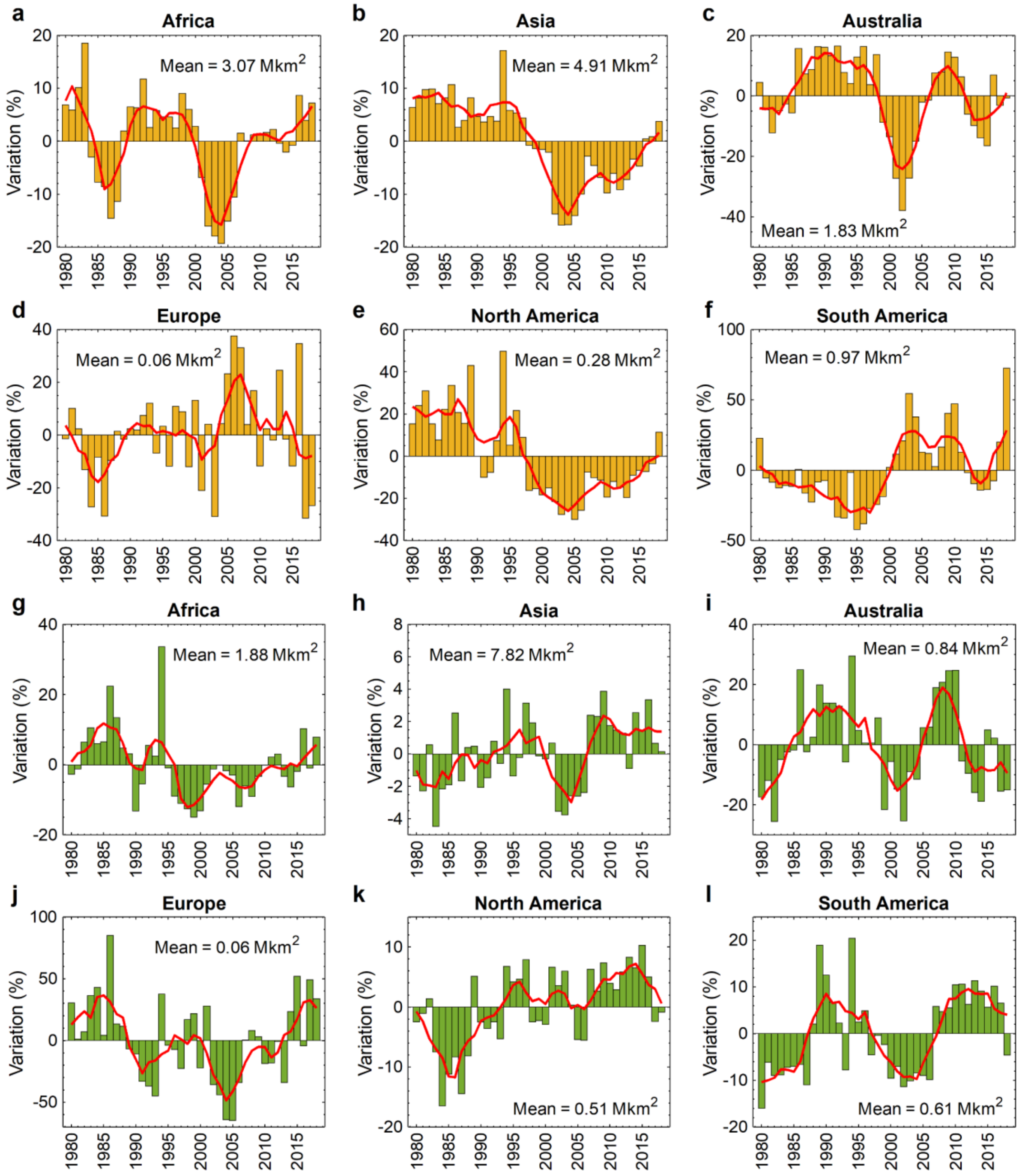


Figure 3-3: Variations in the total area of salt-affected soils between 1980 and 2018 at the continental level. **a** to **f**: Variations in the total area of soils with salinity of $EC_e \geq 4$ dS m^{-1}. **g** to **l**: Variations in the total area of soils with sodicity of ESP ≥ 6%. Red lines show the low-pass filtered (five-year running window) variation of the annual salt-affected areas. Mean values indicate the total area of salt-affected land on each continent averaged from 1980 to 2018.

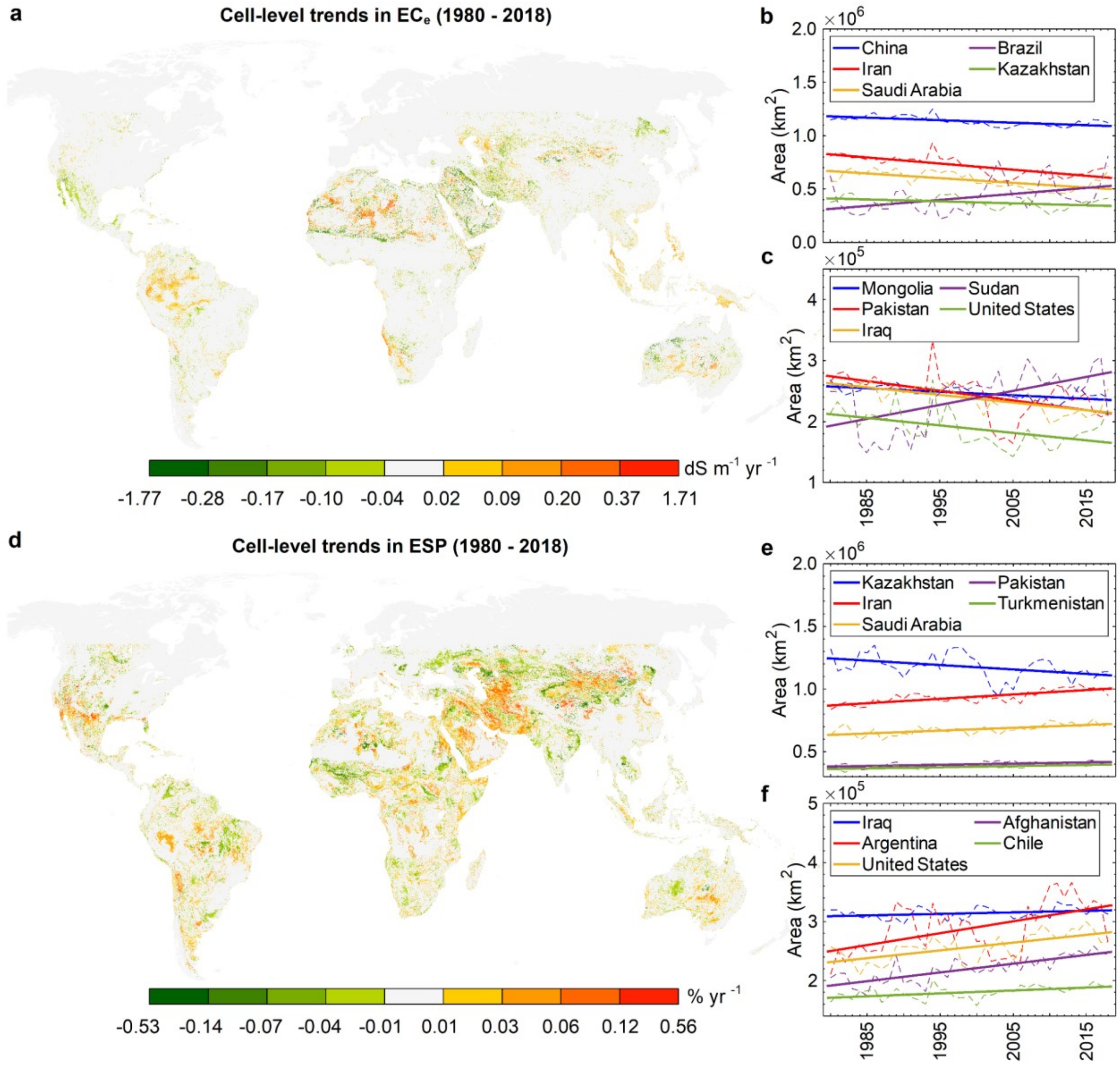


Figure 3-4: Variations in the soil cell-level salinity/sodicity and country-level area of the salt-affected soils ($p < 0.05$). **a** and **d**: Cell-level variations in EC_e and ESP between 1980 and 2018, respectively. Soil cell is any ~1 × 1 km stretch of the soil. Maps are delimited to -55 and 55 latitudes and higher latitudes are shown only for improving the visualization of the maps. **b** and **c**: Variations in the total area of soils with salinity of $EC_e \geq 4$ dS m^{-1} since 1980, at the country level. **e** and **f**: Variations in the total area of soils with sodicity of ESP ≥ 6% since 1980, at the country level (see Appendix 2, Table A2-22 and Appendix 2, Table A2-23 for annual gain or loss in the total area of salt-affected soils for all countries/states). Countries are sorted based on the mean annual area of soils with an $EC_e \geq 4$ dS m^{-1} or ESP ≥ 6% between 1980 and 2018, largest to smallest.

The trends in the total area of soils with $EC_e \geq 4$ dS m^{-1} were statistically meaningful ($p < 0.05$) for only 117 out of 256 countries/states (Figure 3-4), among which the following had the highest rate of annual increase: Brazil (~5,637 km^2 yr^{-1}), Peru (~2,308 km^2 yr^{-1}), Sudan (~2,294 km^2 yr^{-1}), Colombia (~2,007 km^2 yr^{-1}), and Namibia ~1,483 km^2 yr^{-1}). For sodicity (ESP ≥ 6%), the number of countries/states with a statistically significant trend of variation in the total area reduces to 70, with the highest values since 1980 estimated for Iran (~3,499 km^2

yr^{-1}), Saudi Arabia (~2,256 km^2 yr^{-1}), Argentina (~2,012 km^2 yr^{-1}), Afghanistan (~1,483 km^2 yr^{-1}), and the USA (~1,316 km^2 yr^{-1}).

In summary, the dataset, models, and analyses presented here quantified the global long-term variations of top soil EC_e and ESP as respective indicators of soil salinity and sodicity at a high spatial detail, given the limited availability of spatio-temporal data on soil salinity and sodicity. The proposed 4-D modelling approach for predicting soil EC_e and ESP provides new insights into the most influential environmental factors involved in soil salinization processes. Our findings indicate that the total area of salt-affected soils has been temporally and geographically highly variable in the studied period (1980 - 2018), showing both decreasing and increasing trends at the national to continental scales. This sheds new light on this topic, given that the general agreement in the literature is that the salt-affected areas are expanding (Ivushkin et al. 2019; Rozema et al. 2008). These data and the estimated statistics on salt-affected areas can support decision-making under current and future climate scenarios (Hengl et al. 2017a) and direct national and international land-reclamation efforts (Oldeman et al. 2017). Baseline estimates of the soil salinity and sodicity can also inform large-scale crop and agro-ecological models aimed at determining the impact of land degradation and climate change on the food production security (Folberth et al. 2016). These data can also be valuable for soil classification studies (Burt 2011) and development of a more robust response to climate change in soil salinization hotspots. Ultimately, existing models of terrestrial carbon cycling should benefit from the detailed data of soil salinity change (Wong et al. 2010) provided through this work.

3.5 Limitations of the models and recommendations for further research

From the map producers' standpoint, the reliability of the estimated soil surface EC_e and ESP might differ at the continent level and this can be attributed to an uneven spatial distribution of the input soil profiles data used for training the model (Appendix 2, Figure A2-5). Spatial heterogeneity of the soil profile/sample data is a major limitation and source of uncertainty in all Digital Soil Mapping techniques (Fao/Iiasa/Isric/Isscas/Jrc 2012; McBratney et al. 2003; Omuto et al. 2013). Spatial clustering of the training soil profile data is also reported as a major limitation by Ivushkin et al. (2019) and Hengl et al. (2017b) who have used Machine Learning (ML) algorithms for digital soil mapping. The majority of soil profiles are sampled from agricultural lands and areas, such as mountain tops, steep slopes, deserts, sand dunes, and dense tropical forests are considerably under sampled.

In the present study, to quantify how the spatial heterogeneity in the original training sets introduces biases in our analysis, we evaluated the performance of our model at the continental level. Comparisons between the measured surface values of soil EC_e and ESP and the values predicted by the two-part models developed in this study as well as the values presented by Harmonised World Soil Database (Fao/Iiasa/Isric/Isscas/Jrc 2012) are can be found in Appendix 2, Figure A2-21 and Appendix 2, Figure A2-22, at the continental level. Coefficients of determination between the measured values and predictions are provided for each region. As expected, predictions made for locations with a higher number of samples in the training set show higher accuracy suggesting that the reliability of the predictions made by our models is geographically variable. A large proportion of EC_e observations are from North America and Australia (> 90%), making them the most reliable zones of predictions. On the contrary, less than 1% of the ESP observations in the training datasets come from Australia, resulting in higher uncertainty in ESP predictions for Australia. Our investigation highlights the need for training datasets with more optimised spread patterns from unrepresented geographical locations. In addition, for the classification part of each predicted target variable, we produced 39-year mean of pixel-level scaled Shannon Entropy Index (H_s) (Shannon 1949) to identify the certainty of the classifier in binary prediction of classes (see Appendix 2, "Trend analysis" for calculation of H_s). The spatial distribution of H_s is shown in Appendix 2, Figure A2-23. H_s shows the certainty in model predictions; values close to zero indicate that the classifier is more certain about the results of binary classification while values close to one show higher uncertainty. Appendix 2, Figure A2-23 demonstrates that generally the salinity classifier is more certain about the predictions, compared to the sodicity classifier.

In addition to the challenges associated with the spatial heterogeneity in the original training sets discussed above, other limitations that could be addressed in future research include:

- The input data are not uniformly scattered through the time-domain: for EC_e, they are mostly gathered between 2000 and 2005, while the majority of ESP samples are related to the 1990s (Appendix 2, Figure A2-18).
- Despite recent progress in harmonization of the legacy soil profile data, the accuracy and methodology used by different laboratories for gathering and analysing soil samples has not been consistent. This may influence the results of the predictive models (Hengl et al. 2017a).

- Evaluating the propagation of uncertainty over the target variables introduced by each of the 43 predictors was not feasible due to the high computational load of ML algorithms. For the similar reason, we were not able to generate spatially explicit maps of the uncertainty for the predicted target variables and we could only estimate the global uncertainty using 10-fold cross validation.
- It was challenging to quantify the error propagation from the first part of the predictive models (classification) to the second part (regression).
- In this study, we predicted the variations of soil salinity and sodicity at a yearly time resolution, while lower temporal resolutions might be required in some cases. A flash flood or heavy rainfall event, for example, can alter the salinity/sodicity levels of a region within weeks or even days and the two-part models developed here cannot capture salinity/sodicity variations at those temporal resolutions.
- The spatial resolution of the generated maps (~1 km) is not suitable for farm-scale and local studies, so long-term mapping of soil salinity and sodicity at those resolutions remains an open research question.
- Although a fair portion of the available measured data were sampled before 1980, the collection of predictors used in the present study did not allow us to generate maps of EC_e and ESP before 1980s. In particular, remotely-sensed predictors are not available or accessible before 1980s, which makes it challenging to develop the salinity/sodicity map before 1980.
- Similar to the 1980 - 2018 time period, the developed methodology opens a possibility for projection of the soil salinity/sodicity, for example by the end of the 21st century, based on the current trends in soil salinization processes. For future projections, however, both the historical and projected values of the predictors are needed while not all of the 43 predictors used in the current analysis had projected values for the future.

3.6 Methods

Numerical methods have been used to provide the detailed predictions of soil salinization dynamics, mostly based on the solutions of Richard’s equation for water movement in soil unsaturated zone and convection-dispersion equations of solute transport, such as Saito et al. (2006) or Feddes et al. (1988). However, the application of these models remains constrained to localised and short-term simulations as numerical investigation of the interactions between water movement and solute transport in the root zone requires detailed knowledge of many parameters related to soil, climate, and vegetation (Suweis et al. 2010; Li et al. 2015) which are not available on a global scale. Another option for modelling long-term soil salinity is application of salt-balance equations as, for example, in the stochastic model of soil salinity proposed by Suweis et al. (2010) which takes a minimalistic approach to modelling the soil-plant-atmosphere interactions (Porporato et al. 2004). This approach requires long-term measurement of the root zone salt concentration for tuning the calibration parameters but such data are not available at large scales and in many places around the world. Moreover, although these vertically averaged salt balance models can provide mechanistic insights into the soil salinity response to fluctuations in key hydro-climatic drivers of soil salinity, they do not include information about the soil salinity originated from the parent material from which soil is formed.

Therefore, in the present investigation, we used the Digital Soil Mapping framework (Jenny 1994; McBratney et al. 2003) to characterise the spatio-temporal variability in soil salinity. In that framework, the soil characteristics are governed by soil-forming factors, including climate, organisms, relief, parent material, and time. If the relationship between soil profile characteristics (EC_e or ESP in this case), soil-forming factors and their distribution is known, the soil profile characteristics can be inferred/predicted depending on the distribution of the soil-forming factors (Hengl et al. 2017b).

Superior predictive performance of ML algorithms in characterising the relation between the soil profile characteristics and soil-forming factors has been demonstrated in recent studies (Hengl et al. 2017a; Padarian et al. 2020; Heung et al. 2016; Hengl et al. 2015). The procedure for estimation of soil salinity/sodicity involves (1) collection of measured soil salinity and sodicity data for training the model, (2) compiling and processing the predictors (covariates) and linking them to the measured soil salinity and sodicity, (3) mapping a relationship between measured soil profiles data and predictors through building supervised ML models, followed by the validation of the trained models, and (4) deployment of the trained models to predict the spatio-temporal variation of the soil EC_e and ESP at the global scale over the four-decade period considered in the study.

3.6.1 Data

The latest standardised soil dataset from WoSIS (World Soil Information Service) (Batjes et al. 2017) was used to obtain EC_e (dS m^{-1}) at the global scale and to train the models. For consistency, the electrical conductivity of other soil-to-water extract ratios (1:1, 1:2, 1:5, and

1:10) was ignored. This dataset contains 19,434 geo-referenced profile records. Depending on the number and depth of sampling, individual profiles may include information for one or more soil layers. Among 73,517 samples, the EC_e values of only 43,602 (11,303 profiles) samples were measured after January of 1980, the time after which the predictors required in our analysis were available; thus the rest of data points (29,915) were disregarded.

We complied the soil profiles data on soil exchangeable Na^+ (cmol kg^{-1}) and cation exchange capacity (CEC, cmol kg^{-1}) from the National Cooperative Soil Survey Characterization Database (NCSS, https://ncsslabdatamart.sc.egov.usda.gov/), Africa Soil Profiles Database (AfSP, ver. 1.2) (Leenaars et al. 2014), and ISIRC-WISE Harmonised Global Soil Profile Dataset (WISE, ver. 3.1) (Batjes 2008) and divided the exchangeable Na^+ by CEC to calculate ESP as the proposed criterion for evaluating the sodicity levels in soil samples (Burt 2011). Similar to EC_e, the values of ESP recorded before 1980 were excluded. This provided us with ESP values of 207,048 soil layers (36,578 profiles in total), which were used to train the models. The spatial distribution of the EC_e and ESP data used in training and validation of our models are illustrated in Appendix 2, Figure A2-5.

3.6.2 Predictors

We selected the predictors based on the relevance to soil salinization processes as follows: surface evaporation, plant transpiration, fertilisers, poor drainage, and a rising water-table depth (Zaman et al. 2018; Jambhekar et al. 2015; Shokri-Kuehni et al. 2020). In addition, the interactions of five main factors influencing soil formation processes, comprising climate, topography, living organisms, parent material, and hydrologic dynamics, were considered (Jenny 1994; Batjes et al. 2017). Based on these factors, 43 environmental predictors stacked from terrain's elevation data, climate datasets, atmospheric reanalysis, satellite-based remote sensing products, soil and lithological maps, and output of hydrological models were linked to the soil profiles data to develop predictive models of soil salinity/sodicity (Appendix 2, Table A2-1).

In a broad sense, the employed predictors could be categorised into two major groups: static (purely spatial) and dynamic (spatio-temporal). Static predictors were mainly soil texture, and topographic properties that were assumed to remain approximately constant in the period of the analysis (1980 - 2018). Soil texture data including clay, silt, and sand content (weight %) were collected from ISRIC global gridded soil information at 250 m spatial resolutions at five soil depths: 0, 15, 30, 60, and 100 cm (Hengl et al. 2017a). For each soil texture parameter, we generated the averages over the mentioned standard depths using trapezoidal rule (Hengl et al. 2017a). Topographic predictors comprised elevation (m), aspect (degrees), slope (degrees), plan and profile curvatures (calculated by a 10-parameter 3rd order polynomial method (Haralick 1983)), slope length (m), and Terrain Ruggedness Index (TRI) with a square cell radius of three. They were all derived from the SRTM Digital Elevation Database v4.1 (resampled to 250 m resolution) (Jarvis et al. 2008) and computed in the SAGA GIS Terrain Analysis-Hydrology and Morphometry libraries (except elevation and aspect) (Conrad et al. 2015). Other static predictors were: sample upper and lower depths from the surface (cm), soil classes based on the World Reference Base (WRB) (Hengl et al. 2017a) soil classification system, groundwater table depth at equilibrium (m) (Fan et al. 2013), the average of annual fertiliser input rate (1980 - 2018) for C3 annual and perennial crops (kg nitrogen ha^{-1} y^{-1} of crop season; for definition of C3 crops see Appendix 2, Table A2-1) (Hurtt et al. 2016), plant rooting depth (m) (Schenk et al. 2009), average soil and sedimentary thickness (m) (Pelletier et al. 2016), Topographic Index (Marthews et al. 2014), and parent material lithological classes (Hartmann et al. 2012).

Dynamic predictors, on the other hand, were mainly related to the climatic, hydrologic, and surface vegetative variables and were introduced to our model to account for the dynamic processes involved in soil salinization. At our targeted spatial resolution (~1 km at the equator), however, these processes can hardly influence the soil salinity on a daily or monthly basis. Therefore, the long-term averages of the dynamic predictors were applied. Depending on the predictor type, the averaging time window was different to capture the effect of seasonality and inter-annual variations on predictors' values. The dynamic predictors with decadal averaging time window were: annual potential evapotranspiration (mm yr^{-1}), annual precipitation (Harris et al. 2014) (mm yr^{-1}), and monthly minimum, maximum, mean, and diurnal temperature range (Harris et al. 2014) (°C). The dynamic predictors with five-year averaging window were: annual actual evapotranspiration (mm yr^{-1}), annual climate water deficit (mm yr^{-1}), monthly Palmer Drought Severity Index (Palmer 1965), and monthly root-zone soil moisture (mm), all derived from TerraClimate dataset (Abatzoglou et al. 2018). The dynamic predictors with annual averaging window were: remotely-sensed surface soil moisture (2 to 5 cm depth; percentage of total saturation) (Copernicus Climate Change Service 2018), evaporative stress factor (Martens et al. 2017), Leaf Area Index (LAI) (Claverie et al. 2014), the Fraction of Absorbed Photo-synthetically Active Radiation (FAPAR) (Claverie et al. 2014), Normalised Difference Vegetation Index (NDVI) (Didan et al. 2016), two-band Enhanced Vegetation Index

(EVI2) (Didan et al. 2016), wind speed (m s^{-1}) (ERA5 2017), as well as soil skin, layer one (0 - 7 cm), two (7 - 28 cm), three (28 - 100 cm), and four (100 - 289 cm) temperatures (˚K) (ERA5 2017). We generated a spatial layer of each dynamic predictor for each year from 1980 to 2018. The spatial resolution of dynamic variables was generally coarser than that of the static predictors. Additionally, we applied Land Cover Characteristics Data Base (LCCDB) (Loveland et al. 2000) to generate a layer of IGBP land cover classes (Belward et al. 1999) from 1980 to December 1996 as another dynamic predictor. For the 1997 to 2018 period, however, we adopted IGBP land cover classification data from Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) for years 2000, 2006, 2014, and 2018 (Sulla-Menashe et al. 2018).

The spatial resolution of some predictors, such as soil texture, soil classification, land cover, water table depth, and remotely-sensed products, was originally below ~1 km. These data layers were used directly to estimate the soil salinity/sodicity level. However, the spatial resolution of some predictors, mostly climatic ones, was above ~1 km. For those predictors, we used interpolation methods (Appendix 2, Table A2-1) to obtain the data layers at desired spatial resolution (~1 km) and the generated layers were used for prediction of soil salinity and sodicity. All predictors' layers were then projected to WGS 1984 spatial coordinates and saved as raster datasets, except elevation, slope, slope length, TRI, plan and profile curvatures, which were in the World Mercator coordinates system. To estimate the missing data, we filled the spatial gaps (pixels with null values) in data layers using the average of surrounding pixels. A circle with a radius of 4 was used to calculate the missing data using the mean from the neighbouring cells. Even after this procedure, some data were still missing. To resolve this issue, the observations corresponding to those missing cells in the rasters were disregarded, which were 618 (1.41%) observations for EC_e and 9,060 (4.37%) observations for ESP.

The values of cells from rasters of static predictors were directly extracted at locations of observations. For the predictors in the World Mercator projection, we first projected the coordinates of the observation points to World Mercator and then extracted the values of predictors. For the dynamic predictors, however, we binned the training datasets according to the year of acquisition of the observations. For each soil sample with a particular year and observation location, values of the dynamic predictors corresponding to that particular year and location of observation were extracted and attributed to the measured values of EC_e or ESP (all geo-referenced in the WGS 1984 coordinates system). Raster processing and data extractions were conducted in ArcGIS 10.7 (Esri 2011).

3.6.3 Training, validation, and statistical analysis

The final prepared matrices for training had 44 columns (43 representing predictors and one for the target variable) and the number of rows were equal to the number of observations for each target variable. Land cover, parent material lithological units, and WRB soil classes were the three categorical predictors.

In the final training matrices, a large proportion of the measured EC_e and ESP values were zero or close to zero (Appendix 2, Figure A2-18) and this could lead to fitting of the models with predictions biased towards the zero. Therefore, we investigated the patterns between predictors and target variables using a procedure similar to the one used in two-part models in statistics, which model the datasets featuring a large proportion of zeros (Frees 2009; Afifi et al. 2007). To that end, first we decomposed each training dataset into two classes: (1) non-saline ($0 \leq EC_e < 2$ dS m^{-1}; 28,635 observations or 66.6% of the whole training dataset) and saline ($2 \leq EC_e$ dS m^{-1}; 14,349 or 33.4% of the whole training dataset) for EC_e computation; and (2) non-sodic ($0 \leq ESP < 1$ %; 109,340 observations or 55.2% of the whole training dataset) and sodic ($1 \leq ESP$; 88,648 or 44.8% of the whole training dataset) for ESP computation. These thresholds were chosen with the aim of allowing us to divide the training sets into classes with approximately equal number of observations within each class. They should not be confused with the EC_e and ESP thresholds that are conventionally used for characterising saline and sodic soils. Then, a binary classification algorithm was trained to estimate the occurrence probability of each class determining whether the target was saline/sodic or non-saline/non-sodic class (we stress the difference between saline/sodic class and saline/sodic soil terms in our modelling procedure). In the next step, separate regression models were fitted to data in each class to predict the severity of the salinity/sodicity.

The training of the regression and classification models for predicting EC_e and ESP values was executed in the Statistics and Machine Learning toolbox of MATLAB (MATLAB, R2019b). The weight of observations in model trainings was assumed to be constant and equal to one. Based on a trade-off between speed, interpretability, and flexibility of different classification and regression ML algorithms, we used ensemble of regression and classification trees to train different parts of the two-part predictive models and produce the spatial-temporal maps of soil salinity/salinity. To do that, first we imported prepared training sets of salinity and sodicity into MATLAB and trained the classification and regression models for prediction of EC_e and ESP using different available ML algorithms with their default hyperparameter options. The results for classification and regression on saline/sodic classes for each target variable are presented in Appendix 2, Table A2-6. Models based on ensemble of regression/classification trees showed the highest speed, accuracy, and flexibility. Therefore, we chose them for the rest of the analysis.

For classification, MATLAB built-in "fitcensemble" function was used to train an ensemble of classification trees with "tree" type weak-learners. We employed automatic hyperparameter optimisation to find the hyperparameters that minimise the hold-out (with 25% being held out) cross-validation loss. The hyperparameters (Breiman 2001) were the ensemble aggregation method, learning rate, number of learning cycles, minimum leaf size, maximum number of splits, number of variables to sample, and split criterion. They were optimised by the Bayesian optimisation algorithm with the "expected-improvement-per-second-plus" acquisition function. We set the maximum number of objective function evaluations to 130 (there was no considerable variation in the observed minimum objective function after 100 evaluations). In ML classification problems, the class imbalance happens when the number of data in one class is considerably higher than in the other classes. This results in poor predictive power, especially for the class which is less represented. In our analysis, the number of samples in non-saline class was approximately two times higher than in the saline class. When there is a class imbalance in a binary classification problem, other accuracy metrics, such as the proportion of correct predictions to all predictions (accuracy) would have little use since the binary classifier scores a high accuracy if every prediction is assigned to the majority class. In such cases, Matthews Correlation Coefficient (MCC) (Matthews 1975) is a more reliable accuracy measure (Boughorbel et al. 2017) and we used this accuracy metric to evaluate the performance of the trained binary classifiers.

Likewise, we applied the MATLAB built-in "fitrenemble" function to fit a predictive model from the ensemble of regression trees for data within each separate class. With hyperparameter optimisation options similar to "fitcenemble", the candidate hyperparameters (Breiman 2001) for optimisation were the number of learning cycles, learning rate, minimum leaf size, maximum number of splits, and number of variables to sample; for regression, we used "LSBoost" (Least-squares Boosting) method for training the models. Logarithm transform was applied to normalise the right skewness in frequency distribution of the response variables (Appendix 2, Figure A2-18).

10-fold cross validation was used to estimate the performance of fitted models. In addition to the *MCC*, binomial deviance loss, misclassification accuracy, precision, and recall metrics were also calculated for the fitted classifier models. For regression predictions, root mean squared error (*RMSE*), mean absolute error (*MAE*), and Nash-Sutcliffe model efficiency coefficient (*NSE*) (Nash et al. 1970) in both logarithm-transformed and back-transformed scales were estimated. Since the hyperparameter optimisation was stochastic and it was not possible to regenerate the hyperparameter optimisation results of each training run, we repeated the training of each of these three models 30 times. Appendix 2, Table A2-7 to Table A2-12 show the results of hyperparameter optimisation and the 10-fold cross-validation for those 30 runs for each part of the developed two-part models. In total, there were two target variables, three models for each target variable, and 180 runs. Among the 30 trained models, we chose the one with the best performance (the lowest error, Appendix 2, Table A2-2). The trained classifiers with the highest *MCC* and regressions within each class with the highest *NSE* (in total six models) were selected for the rest of the analysis. Repeating the training process also gave us the opportunity to calculate the confidence intervals for the 10-fold cross-validation accuracy metrics (Appendix 2, Table A2-2). We generated 1,000 bootstrapped samples with replacement from validation metrics, and computed the 95% confidence intervals of the mean for each validation metric using the bias corrected and accelerated percentile method (MATLAB built-in “bootci” function).

3.6.4 Prediction of spatio-temporal evolution of soil salinity at the global scale

The trained models were applied to a global soil mask layer to make annual predictions of surface soil salinity at 30" resolution (0.008333°, ~1 km at the equator) since 1980. To generate the global soil mask layer, we re-projected/resampled the 2014 MODIS land cover map (Sulla-Menashe et al. 2018) to the WGS 1984 coordinates system/30" resolution using the nearest neighbour method and masked out the pixels labelled as water bodies, permanent wetlands, urban and built-up lands, and permanent snow and ice. Due to the unavailability of the topographic predictors’ values (as input of models) at frigid zones and higher latitudes, we focused on the pixels located between the -55° and 55° latitudes. The final raster layer was split to tiles to facilitate the subsequent data analysis. We converted the tiles to point feature layers, extracted the values of static and dynamic predictors to the points in each year, and exported the corresponding tables and points’ coordinates as text files to make predictions using the trained models in MATLAB. Predictions and x- y- coordinates (representative of longitude and latitude) defined in output tables were rasterised and mosaicked to generate the final maps of soil salinity for each year over the studied period. We divided the workflow of extraction of predictors' values to points between 16 processes on a machine with 16 cores through the multiprocessing Python module (Python.org) and the task was completed in six days. Exporting and saving the attribute tables as a text file and deployment of the trained models on the new data (~6 billion rows) was accomplished in nearly 60 days by running a parallel pool of 16 processes on the above-mentioned dedicated machine.

In total, for each target variable and location with x- y- coordinates, 39 predictions were made (each representing one year from 1980 to 2018). We calculated the intra-annual likelihood of saline/sodic soils occurring in each x- y- point following the approach proposed by Pekel et al. (2016). By dividing the number of years which had the EC_e values ≥ 4 dS m^{-1} and ESP values $\geq 6\%$ by the total number of studied years (39 years), the likelihood of surface soils with $EC_e \geq 4$ dS m^{-1} and ESP value $\geq 6\%$ was computed, respectively. To understand and quantify the variation in the likelihood of soils with $EC_e \geq 4$ dS m^{-1} and ESP $\geq 6\%$, we divided the study period into two 19-year periods: January 1981 to December 1999, and January 2000 to December 2018. Then, for each variable, we defined the parameter θ as $\theta = log_e$ *((Likelihood of the 2000 - 2018 period + 0.5) / (Likelihood of the 1981 - 1999 period + 0.5))* (Appendix 2, Figure A2-4). Due to the presence of zero frequency counts in either the 1981 - 1999 or 2000 - 2018 period, we added a "continuity correction" of 0.5 to the frequency counts for both periods (Higgins et al. 2011). We fitted a linear model to the predicted soil salinity and sodicity in each year since 1980 and the slope of the fitted models with $p < 0.05$ was considered as a soil salinity long-term trend for that location. We also generated two other layers from the soil cell-level mean (Appendix 2, Figure A2-6) and standard deviation of the annual predicted target variables (Appendix 2, Figure A2-7) between 1980 and 2018.

To estimate the annual soil area with $EC_e \geq 4$ dS m^{-1} or ESP $\geq 6\%$ at the land cover, biome, climate, and national/continental levels, first we discretised the annual predicted values for EC_e and ESP at each x- y- position into four classes: 0 - 4 dS m^{-1}, 4 - 8 dS m^{-1}, 8 - 16 dS m^{-1}, and > 16 dS m^{-1} for EC_e and 0 - 6%, 6 - 15%, 15 - 30%, and > 30% for ESP (each class includes its left class edge). Then, we directly derived the area of each x- y- point in the WGS 1984 coordinates system for salinity/sodicity classes (assuming each point represents a raster pixel with the size of 0.008333°), following the method presented in the Appendix 2 (Model deployment). The computed areas with the corresponding locations were converted to raster layers. Therefore, for each year and target variable, we produced four raster layers from the four salinity/sodicity classes representing the area of pixels (in WGS 1984). Finally using the

ArcGIS 10.7 “Zonal Statistics” tool, the sum of areas in each class and zone specified by biome (adopted form modified terrestrial ecoregions of the world, available at Nature Conservancy, Geospatial Conservation Atlas; https://geospatial.tnc.org/), climate zone (adopted from a world map of the climate classification after Kottek et al. (2006)), and country/continent border (adopted form global administrative areas, GADM (Hijmans et al. 2012)) datasets were calculated. For delineation of land cover zones, we compared the IGBP land cover classes of LCCDB (Loveland et al. 2000) in 1993 with MODIS generated land cover map of 2018 (Sulla-Menashe et al. 2018) and kept those pixels which were classified with the same land cover type in both years. The statistics on the trends and total areas of surface soils with $EC_e \geq 4$ dS m^{-1} and ESP ≥ 6% were calculated at different levels (land cover, biome, climate, country, and continent) by summing up the area of all salinity classes with $EC_e \geq 4$ dS m^{-1} and sodicity classes with ESP value ≥ 6%, respectively.

Data availability

Input training data (ground-measured values of EC_e and ESP), objects of the two-part predictive models, and thematic maps quantifying different aspects of surface soil salinity and sodicity (0 - 30 cm) are freely available at: https://data.mendeley.com/datasets/v9mgbmtnf2/1. The maps of surface soil salinity (EC_e) and sodicity (ESP) for each year between 1980 and 2018 are available at https://doi.org/10.6084/m9.figshare.13295918.v1. All statistics provided in this paper, in addition to further data on spatio-temporal variability of the salt-affected soils at the cell, land cover, biome, climate, country, and continental levels are available in a tabular format in Appendix 2 (Statistics on salt-affected regions).

Code availability

All computer codes and further details on methods required for regeneration of the main results presented in this paper can be found in Appendix 2 (Computer codes).

Acknowledgements

This study was funded by the UK Research Councils (grant no. EP/K011820/1) and the Presidential Doctoral Scholarship Award at The University of Manchester. The authors gratefully acknowledge this funding.

Author contributions. The contribution of all authors was equal and essential to preparation of the material provided/presented here.

Additional information. The authors declare no competing interests.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data, 5*, 170191.

Abrol, I., Yadav, J. S. P., & Massoud, F. (1988). *Salt-affected soils and their management*: Food & Agriculture Org.

Afifi, A. A., Kotlerman, J. B., Ettner, S. L., & Cowan, M. (2007). Methods for improving regression analysis for skewed continuous or counted responses. *Annu. Rev. Public Health, 28*, 95-111.

Allbed, A., & Kumar, L. (2013). Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: a review. *Advances in remote sensing, 2013*.

Amini, S., Ghadiri, H., Chen, C., & Marschner, P. (2016). Salt-affected soils, reclamation, carbon dynamics, and biochar: a review. *Journal of soils and sediments, 16*(3), 939-953.

Batjes, N. H. (2008). ISRIC-WISE Harmonized Global Soil Profile Dataset. *ISRIC-World Soil Information, Wageningen*.

Batjes, N. H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., & de Jesus, J. M. (2017). WoSIS: providing standardised soil profile data for the world. *Earth System Science Data, 9*(1), 1.

Belward, A. S., Estes, J. E., & Kline, K. D. (1999). The IGBP-DIS global 1-km land-cover data set DISCover: A project overview. *Photogrammetric Engineering and Remote Sensing, 65*(9), 1013-1020.

Bleam, W. F. (2016). *Soil and environmental chemistry*: Academic Press.

Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS one, 12*(6), e0177678.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Burt, R. (2011). Soil Survey Investigations Report, no. 45, version 2.0. *Natural Resources Conservation Service*.

Butcher, K., Wick, A. F., DeSutter, T., Chatterjee, A., & Harmon, J. (2016). Soil salinity: A threat to global food security. *Agronomy Journal, 108*(6), 2189-2200.

Claverie, M., & Vermote, E. (2014). NOAA Climate Data Record (CDR) of Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) Version 4. *NOAA National Centers for Environmental Information*.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., et al. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific Model Development Discussions, 8*(2).

Copernicus Climate Change Service. (2018). Soil moisture gridded data from 1978 to present. *Copernicus Climate Change Service, Climate Data Store (CDS)*.

Daliakopoulos, I., Tsanis, I., Koutroulis, A., Kourgialas, N., Varouchakis, A., Karatzas, G., & Ritsema, C. (2016). The threat of soil salinity: A European scale review. *Science of the Total Environment, 573*, 727-739.

De la Paix, M., Lanhai, L., Xi, C., Varenyam, A., Nyongesah, M., & Habiyaremye, G. (2013). Physicochemical properties of saline soils and aeolian dust. *Land Degradation & Development, 24*(6), 539-547.

Didan, K., & Barreto, A. (2016). NASA MEaSUREs vegetation index and phenology (VIP) vegetation indices monthly global 0.05 Deg CMG. *NASA EOSDIS Land Process. DAAC, 4*.

ERA5. (2017). Fifth generation of ECMWF atmospheric reanalyses of the global climate. *Copernicus Climate Change Service, Climate Data Store (CDS)*.

Esri, R. (2011). ArcGIS desktop: release 10. *Environmental Systems Research Institute, CA*.

Fan, X., Pedroli, B., Liu, G., Liu, Q., Liu, H., & Shu, L. (2012). Soil salinity development in the yellow river delta in relation to groundwater dynamics. *Land Degradation & Development, 23*(2), 175-189.

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science, 339*(6122), 940-943.

FAO. (2000). Extent and causes of salt affected soils in participating countries. *Global Network on integrated soil management for sustainable use of saltaffected soils.* Available at http://www.fao.org/ag/agl/agll/spush/topic2.htm.

FAO Soils Portal, S. R., Food and Agriculture Organization of the United Nations. (2020). Extent of salt-affected soils, http://www.fao.org/soils-portal/soil-management/management-of-some-problem-soils/salt-affected-soils/more-information-on-salt-affected-soils/en/.

Fao/Iiasa/Isric/Isscas/Jrc. (2012). Harmonized world soil database (version 1.2). *FAO, Rome, Italy and IIASA, Laxenburg, Austria*.

Feddes, R., Kabat, P., Van Bakel, P., Bronswijk, J., & Halbertsma, J. (1988). Modelling soil water dynamics in the unsaturated zone—state of the art. *Journal of hydrology, 100*(1-3), 69-111.

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., & Van Der Velde, M. (2016). Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nature communications, 7*(1), 1-13.

Frees, E. W. (2009). *Regression modeling with actuarial and financial applications*: Cambridge University Press.

Ghassemi, F., Jakeman, A. J., & Nix, H. A. (1995). *Salinisation of land and water resources: human causes, extent, management and case studies*: CAB international.

Gupta, R. K., Abrol, I. P., Finkl, C. W., Kirkham, M. B., Arbestain, M. C., Macías, F., et al. (2008). Solonchaks. In W. Chesworth (Ed.), *Encyclopedia of Soil Science* (pp. 737-738). Dordrecht: Springer Netherlands.

Haralick, R. M. (1983). Ridges and valleys on digital images. *Computer Vision, Graphics, and Image Processing, 22*(1), 28-38.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International journal of climatology, 34*(3), 623-642.

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems, 13*(12).

Hassani, A., Azapagic, A., D'Odorico, P., Keshmiri, A., & Shokri, N. (2020). Desiccation crisis of saline lakes: A new decision-support framework for building resilience to climate change. *Science of the Total Environment, 703*, 134718.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one, 12*(2).

Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., et al. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS one, 10*(6), e0125814.

Hengl, T., Leenaars, J. G., Shepherd, K. D., Walsh, M. G., Heuvelink, G. B., Mamo, T., et al. (2017). Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutrient Cycling in Agroecosystems, 109*(1), 77-102.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma, 265*, 62-77.

Higgins, J., & Wells, G. (2011). Cochrane handbook for systematic reviews of interventions.

Hijmans, R., Garcia, N., Kapoor, J., Rala, A., Maunahan, A., & Wieczorek, J. (2012). Global Administrative Areas (Boundaries). *Berkeley: Museum of Vertebrate Zoology and the International Rice Research Institute, University of California*.

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Calvin, K., Fujimori, S., et al. (2016). Harmonization of global land use change and management for the period 850–2100. *Geoscientific Model Development*.

Isbell, R. (2016). *The Australian soil classification*: CSIRO publishing.

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., & De Sousa, L. (2019). Global mapping of soil salinity change. *Remote Sensing of Environment, 231*, 111260.

Jambhekar, V., Helmig, R., Schröder, N., & Shokri, N. (2015). Free-flow–porous-media coupling for evaporation-driven transport and precipitation of salt in soil. *Transport in Porous Media, 110*(2), 251-280.

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database.

Jenny, H. (1994). *Factors of soil formation: a system of quantitative pedology*: Courier Corporation.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift, 15*(3), 259-263.

Leenaars, J. G., Van OOstrum, A., & Ruiperez Gonzalez, M. (2014). Africa soil profiles database version 1.2. A compilation of georeferenced and standardized legacy soil profile data for Sub-Saharan Africa (with dataset). *ISRIC–World Soil Information, Wageningen, the Netherlands*.

Li, H., Yi, J., Zhang, J., Zhao, Y., Si, B., Hill, R. L., et al. (2015). Modeling of soil water and salt dynamics and its effects on root water uptake in Heihe arid wetland, Gansu, China. *Water, 7*(5), 2382-2401.

Loveland, T., Reed, B., Brown, J., Ohlen, D., Zhu, Z., Yang, L., & Merchant, J. (2000). Global land cover characteristics database (GLCC) version 2.0. *USGS Natl. Cent. for Earth Resour. Obs. and Sci., Sioux Falls, SD*.

Martens, B., Gonzalez Miralles, D., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development, 10*(5), 1903-1925.

Marthews, T., Dadson, S., Lehner, B., Abele, S., & Gedney, N. (2014). A high-resolution global dataset of topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences Discussions, 11*, 6139-6166.

Martinez-Beltran, J. (2005). *Overview of salinity problems in the world and FAO strategies to address the problem.* Paper presented at the Managing saline soils and water: science, technology and social issues. Proceedings of the international salinity forum, Riverside, California, 2005.

Mateo-Sagasta, J., & Burke, J. (2011). Agriculture and water quality interactions: a global overview. *SOLAW Background Thematic Report-TR08, 46*.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure, 405*(2), 442-451.

Mau, Y., & Porporato, A. (2016). Optimal control solutions to sodic soil reclamation. *Advances in water resources, 91*, 37-45.

McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1-2), 3-52.

Metternicht, G. (2016). Soils: salinization. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, 1-10.

Montanarella, L., Badraoui, M., Chude, V., Costa, B., Dos Santos, I., Mamo, T., et al. (2015). Status of the World's Soil Resources Main report.

Mulder, V., De Bruin, S., Schaepman, M. E., & Mayr, T. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma, 162*(1-2), 1-19.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology, 10*(3), 282-290.

Northcote, K. H., & Srene, J. (1972). *Australian soils with saline and sodic properties*. (No. R&D Rpt).

Oldeman, L. R., Hakkeling, R., & Sombroek, W. G. (2017). *World map of the status of human-induced soil degradation: an explanatory note*: International Soil Reference and Information Centre.

Omuto, C., Nachtergaele, F., & Rojas, R. V. (2013). *State of the Art Report on Global and regional Soil Information: Where are we? Where to go?* : Food and Agriculture Organization of the United Nations, Rome.

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. *Soil, 6*(1), 35-52.

Palmer, W. C. (1965). *Meteorological drought* (Vol. 30): US Department of Commerce, Weather Bureau.

Parihar, P., Singh, S., Singh, R., Singh, V. P., & Prasad, S. M. (2015). Effect of salinity stress on plants and its tolerance strategies: a review. *Environmental Science and Pollution Research, 22*(6), 4056-4075.

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature, 540*(7633), 418.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G. Y., et al. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems, 8*(1), 41-65.

Perri, S., Suweis, S., Holmes, A., Marpu, P. R., Entekhabi, D., & Molini, A. (2020). River basin salinization as a form of aridity. *Proceedings of the National Academy of Sciences, 117*(30), 17635-17642.

Ponnamperuma, F. (1984). Role of cultivar tolerance in increasing rice production on saline lands.

Porporato, A., Daly, E., & Rodriguez-Iturbe, I. (2004). Soil water balance and ecosystem response to climate change. *The American Naturalist, 164*(5), 625-632.

Porporato, A., Feng, X., Manzoni, S., Mau, Y., Parolari, A. J., & Vico, G. (2015). Ecohydrological modeling in agroecosystems: Examples and challenges. *Water resources research, 51*(7), 5081-5099.

Rath, K. M., & Rousk, J. (2015). Salt effects on the soil microbial decomposer community and their role in organic carbon cycling: a review. *Soil Biology and Biochemistry, 81*, 108-123.

Rengasamy, P. (2006). World salinization with emphasis on Australia. *Journal of experimental botany, 57*(5), 1017-1023.

Richards, L. A. (1954). *Diagnosis and improvement of saline and alkali soils. Handbook No. 60*: US Department of Agriculture, Washington, DC.

Rozema, J., & Flowers, T. (2008). Crops for a salinized world. *Science, 322*(5907), 1478-1480.

Saito, H., Šimůnek, J., & Mohanty, B. P. (2006). Numerical analysis of coupled water, vapor, and heat transport in the vadose zone. *Vadose Zone Journal, 5*(2), 784-800.

Sandra, P. (1989). Water for agriculture: facing the limits. *Worldwatch paper (USA). no. 93.*

Schenk, H., Jackson, R., Hall, F., Collatz, G., Meeson, B., Los, S., et al. (2009). Islscp ii ecosystem rooting depths. *ORNL DAAC.*

Schofield, R., & Kirkby, M. (2003). Application of salinization indicators and initial development of potential global soil salinization scenario under climatic change. *Global Biogeochemical Cycles, 17*(3).

Sentis, I. (1996). Soil salinization and land desertification. *Soil degradation and desertification in Mediterranean environments. Logroño, Spain, Geoforma Ediciones*, 105-129.

Setia, R., Gottschalk, P., Smith, P., Marschner, P., Baldock, J., Setia, D., & Smith, J. (2013). Soil salinity decreases global soil organic carbon stocks. *Science of the Total Environment, 465*, 267-272.

Shani, U., Ben-Gal, A., Tripler, E., & Dudley, L. M. (2007). Plant response to the soil environment: An analytical model integrating yield, water, soil type, and salinity. *Water resources research, 43*(8).

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE, 37*(1), 10-21.

Shokri-Kuehni, S. M., Raaijmakers, B., Kurz, T., Or, D., Helmig, R., & Shokri, N. (2020). Water Table Depth and Soil Salinization: From Pore-Scale Processes to Field-Scale Responses. *Water resources research, 56*(2), e2019WR026707.

Singh, K. (2016). Microbial and enzyme activities of saline and sodic soils. *Land Degradation & Development, 27*(3), 706-718.

Soil Science Glossary Terms Committee. (2008). *Glossary of soil science terms 2008*: Soil Science Society of America, ASA-CSSA-SSSA.

Soil Survey Staff. (2010). *Keys to soil taxonomy*: United States Department of Agriculture, Soil Conservation Service: Washington, DC.

Squires, V. R., & Glenn, E. P. (2011). Salination, desertification and soil erosion. *The role of food, agriculture, forestry and fisheries in human nutrition, 3*, 102-123.

Sulla-Menashe, D., & Friedl, M. A. (2018). User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. *USGS: Reston, VA, USA*, 1-18.

Suweis, S., Rinaldo, A., Van der Zee, S., Daly, E., Maritan, A., & Porporato, A. (2010). Stochastic modeling of soil salinity. *Geophysical Research Letters, 37*(7).

Szabolcs, I. (1989). *Salt-affected soils*: CRC Press, Inc.

Van Dijk, A. I., Beck, H. E., Crosbie, R. S., de Jeu, R. A., Liu, Y. Y., Podger, G. M., et al. (2013). The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water resources research, 49*(2), 1040-1057.

Várallyay, G. (1994). Climate change, soil salinity and alkalinity. In *Soil responses to climate change* (pp. 39-54): Springer.

Wong, V. N., Greene, R., Dalal, R. C., & Murphy, B. W. (2010). Soil carbon dynamics in saline and sodic soils: a review. *Soil use and management, 26*(1), 2-11.

Zaman, M., Shahid, S. A., & Heng, L. (2018). *Guideline for salinity assessment, mitigation and adaptation using nuclear and related techniques*: Springer.

Chapter 4 Climate change and primary soil salinization: A global scale perspective for the 21st century

Amirhossein Hassani[a], Adisa Azapagic[a*], and Nima Shokri[b*]

[a] Department of Chemical Engineering and Analytical Science, The University of Manchester, Sackville Street, Manchester M13 9PL, UK.
[b] Hamburg University of Technology, Institute of Geo-Hydroinformatics, Am Schwarzenberg-Campus 3 (E), 21073 Hamburg, Germany

*Corresponding authors: adisa.azapagic@machester.ac.uk; nima.shokri@tuhh.de

Abstract

Soil salinization has become one of the major environmental and socioeconomic issues globally and this is expected to be exacerbated further with projected climatic change. Determining how climate change influences the dynamics of naturally-occurring soil salinization has scarcely been addressed due to highly complex processes influencing salinization. This paper sets out to address this long-standing challenge by developing data-driven models capable of predicting primary (naturally-occurring) soil salinity and its variations in the world's drylands up to the year 2100 under changing climate. Analysis of the future predictions made here identifies the dryland areas of South America, southern and Western Australia, Mexico, southwest United States, and South Africa as the salinization hotspots. Conversely, we project a decrease in the soil salinity of the drylands in the northwest United States, the Horn of Africa, Eastern Europe, Turkmenistan, and west Kazakhstan in response to climate change over the same period.

4.1 Introduction

The Soil Science Society of America (Soil Science Glossary Terms Committee, 2008) defines saline soil as a non-sodic soil containing sufficient amount of soluble salt which could adversely influence most crop plants. Conventionally, electrical conductivity of a saturated soil paste extract (EC_e) has been used as a measure of the soil salinity (Burt, 2011). Soil salinization is a land degradation process that results in excessive accumulation of soluble salts in the soil (Abrol et al., 1988; Bleam, 2016). In naturally-occurring or primary soil salinization, the predominant origins of soluble salts are rainfall (wet deposition of oceanic salts), aeolian processes (dry deposition of oceanic salts), and physical or chemical weathering of parent rock materials (Rengasamy, 2006; Zaman et al., 2018). Transport of the accumulated salts from saline geological depositions by streamflow or shallow underground waters is an additional source of primary salinization (I. Daliakopoulos et al., 2016). In anthropogenic or secondary soil salinization, however, the main sources of salinization are human interventions, such as irrigation with brackish or saline water, rising water tables due to poor land and water management, surface or subsurface sea water intrusion into coastal aquifers as a result of rising sea levels or over-exploitation of the fresh underground waters, and overuse of fertilizers (I. Daliakopoulos et al., 2016; Pannell & Ewing, 2006; Zaman et al., 2018).

Excessive accumulation of the soluble salts in the root zone may go beyond the salt tolerance of plants, affecting adversely the growth rate of the plants (Ayub et al., 2020). A soil with salinity of $EC_e \geq 2$ dS m^{-1} (at 25 °C) is traditionally considered as a saline soil (Fao/Iiasa/Isric/Isscas/Jrc, 2012); however, depending on the plant type, climatic conditions, and soil-water balance properties, the salt-tolerance of sensitive crops and plants can be different (Maas & Grattan, 1999). Salinity stress deteriorates the plants' transpiring leaves which is known as specific ion effects (Greenway & Munns, 1980) or directly reduces the plant water uptake from the rooting zone, resulting in osmotic stress on the plant (Munns & Tester, 2008; Parihar et al., 2015). Soil salinity also imposes nutritious imbalances in plants (Rengasamy, 2006). Soil salinity between 2 and 4 dS m^{-1} can negatively impact the yields of sensitive plants and at salinity levels higher than 8 dS m^{-1}, the growth of most of crops and plants shows a severe decrease in response to excessive soil salinity (Hillel, 2000; Soil Science Glossary Terms Committee, 2008). Vegetation loss in turn reduces the soil stability and exposes the soil to wind or water erosion (De la Paix et al., 2013). In addition to deleterious effects on vegetation, excessive soil salinity decreases the biological functioning of the soil micro-organisms to a level that disturbs the soil nitrogen cycle, respiration, and organic matter input (Rath & Rousk, 2015; Singh, 2016). Reduced environmental health due to aeolian dispersion of saline dust originated from the saline soils (De la Paix et al., 2013; Hassani, Azapagic, D'Odorico, et al., 2020), land abandonment and desertification (Perri et al., 2020; Sentis, 1996), worsening of economic welfare, and human migration are other detrimental consequences of excessive soil salinity (Hassani, Azapagic, D'Odorico, et al., 2020; Rengasamy, 2006).

Accurate and reliable data on spatial distribution of salt-affected soils are important to develop action plans for management of soil, water, and vegetation and will contribute toward data-driven policy making (Oldeman et al., 2017; Omuto et al., 2013; Pannell, 2001). These data have also implications for tuning large-scale agro-ecological models (Folberth et al., 2016) and planning sustainable reclamation practices (Amini et al., 2016). With varying levels of accuracy and spatial coverage, from the local (Paz et al., 2020; Scudiero et al., 2014; Taghizadeh-Mehrjardi et al., 2014) to the global scale (FAO ITPS, 2015; Ghassemi et al., 1995; Hassani et al., 2020; I. Szabolcs, 1989), defining the spatial distribution and location of salt-affected soils has been under focus of various studies. According to the global-scale studies, salt-affected soils lie across all climate zones and continents with an estimated global area of ~8.31 - 11.73 Mkm^2, depending on the methods used for estimation of area of the salt-affected soils. Nevertheless, the general consensus is that the saline and salt-affected soils (including sodic soils) are particularly found in drylands where the excess of evaporation over water input to the soil accumulates salts in the upper soil layer (Abrol et al., 1988; Fischer et al., 2008; Richards, 1954).

Drylands, including hyper-arid, arid, semi-arid, and dry sub-humid lands, are characterised by a multi-annual Aridity Index (AI) of less than 0.65 mm mm^{-1}, computed as the ratio of total precipitation to potential evapotranspiration (Middleton & Thomas, 1997; UNEP-WCMC, 2007). Drylands occupy a total of ~45% of the Earth's surface (Prăvălie, 2016; Schimel, 2010). With the advance of proximal/remote sensors and digital soil mapping techniques, there is a rising interest in spatio-temporal mapping and monitoring of the soil salinity (Gorji et al., 2017; Ivushkin et al., 2019; Mulder et al., 2011). Due to the temporal and vertical variability in salinity levels of the salt-affected soils (Mulder et al., 2011; Zaman et al., 2018), updated predictions on long-term variations of soil salinity can provide a clearer understanding of the dynamics of the terrestrial carbon sink (Wong et al., 2010), climate change impacts (Gy Várallyay, 1994), and alterations in the land, vegetation, and water resources (National Land and Water Resources Audit, 2001). Even though the above-mentioned purely spatial or spatio-temporal studies have substantially advanced our understanding of the current status of the salt-affected soils and processes involved in salinization, predictions of the future extent and dynamics of soil salinization at the global scale are still missing, partly due to the complex processes and many parameters influencing soil salinization at the global scale. This makes the future prediction of soil salinization in the face of future climate uncertainties a grand challenge, which is precisely one of the key objectives of the present investigation.

The projected hydrological consequences of climate change may result in physical, biological, biochemical, and chemical degradation of the soils (G Várallyay, 2010). As one of the major threats to soil stability, fertility, and biodiversity, it is expected that the soil salinity will be a significant and growing concern in a warmer world (Talat, 2020; Tomaz et al., 2020).

To formulate appropriate plans for sustainable management of soil, water, and vegetation, reliable predictions on the probable occurrence and expansion or shrinkage of the salt-affected soils in response to the threat of climate change are crucial. Compared to other dynamic soil properties, such as P, N, and organic matter content, prediction of soil-salinity responses to climate variability on a global scale has received much less attention (D. Corwin, 2020). The available literature on the effect of climate change as a source of soil salinization is mainly descriptive and quantitative predictions of the future status of salt-affected soils on the basis of current trends are rare. The IPCC report (Intergovernmental Panel on Climate Chang, 1996) predicts that climate change will likely impact all the primary mechanisms for soil salinization, including soluble salts accumulation due to a change in hydrological balance, sea salt intrusion, and wind-born salt deposition. An increase in the rate of evapotranspiration and alteration in precipitation patterns, particularly in arid and semi-arid areas, is expected to reduce the soil leaching efficiency and consequently, increase the salt concentrations in top-soil horizons (Bates, 2009; Karmakar et al., 2016; Szabolcs, 1990). Expansion of irrigated areas and the higher demand for water use under rising global temperatures, in combination with poor drainage/irrigation practices, are expected to result in the spread of secondary salinization (Yeo, 1998). Land use modifications and occurrence of more extreme climate events, such as prolonged droughts followed by severe floods, have the potential to release and redistribute large amount of salts from the geological substrates with high concentration of salts and may put new areas at risk of soil salinization (Van Weert et al., 2009). In addition, rising sea levels and unsustainable extraction of fresh water resources from coastal aquifers can worsen the issue of sea water-induced soil salinization in coastal regions (Dasgupta et al., 2015; Karmakar et al., 2016).

A few studies investigated some aspects of the relationship between projected climate change and soil salinization. Szabolcs (1990) was among the first who estimated that the salt-affected areas in North Mediterranean regions will be doubled by 2050 in response to 1 °C increase in the average annual temperature. Similarly, National Land and Water Resources Audit (2001) estimated that Australia’s drylands at risk of soil salinity imposed by dryland management actions may expand to 170,000 km^2 in 2050, relative to approximately 57,000 km^2 in 2000. Schofield and Kirkby (2003) developed a set of soil salinization indicators including low relief, high two-way annual moisture flux, and local flow deficit in large catchments to identify the current and future (2079 - 2099) locations with salinization potential across the globe and concluded that areas at risk of soil salinity are expanding. Although these studies provide an understanding of the salinization potential and limitations of the methods used for projecting the soil salinity, they are not based on up-to-date datasets and they mainly highlight the areas at risk; no quantitative and spatially explicit predictions are provided. Other studies on predicting impacts of climate change on soil salinization are mainly focused on predicting secondary salinization processes imposed by unsustainable irrigation practices (I. N. Daliakopoulos et al., 2016; Martín-Rosales et al., 2007; Zanchi & Cecchi, 2010) or sea water intrusion (Chen & Mueller, 2018; Colombani et al., 2015; Oude Essink et al., 2010) at local scales. Thus, there is a need for a quantitative global-scale analysis, characterising the geographical distribution and projecting the long-term variations in soil salinity in the face of future climate fluctuations and uncertainties, which motivated the present investigation.

This study is among the first attempts for addressing the need for a quantitative tool capable of predicting long-term primary soil salinity on a global scale with a high spatial and temporal resolution. These models and the results will be of interest to local authorities, land managers, and policy makers, helping to plan mitigation of and adaptation to soil salinization. In particular, we performed comprehensive data-driven modelling and analyses to reveal how the projected or hypothesised variations in the key drivers may influence primary soil salinity on the global scale, in both mid- (2031 - 2060) and long-term (2071 - 2100) futures. We only focus on soil salinity in the top-soil horizon (0 - 1 m), quantified by the concentration of soluble salts which is expressed by the extent of EC_e. Other aspects of salt-affected soils, such as sodicity (which is traditionally measured by the soil exchangeable sodium percentage) or alkalinity, are not within the scope of this analysis. The potential soil salinity caused by sea level rise, saline groundwater, or irrigation is also excluded from the study. Note that modelling the salinity intrusion in coastal areas in response to sea rise needs a relatively precise estimation of the future groundwater extraction from the coastal aquifer. Similarly, projected data of groundwater level and salinity change (either natural or anthropogenic) are needed for predicting the ground water-induced soil salinity which is not currently available. As mentioned in Yeo (1998), it is difficult to generate a clear prediction of the impacts of climate change on the extent of salinization caused by irrigation as this requires reliable estimations of irrigation expansion and the quality of irrigation water in future. Therefore, this study can be deemed as projection of the primary soil salinization under future climate uncertainty.

Several numerical methods have been developed to simulate the soil salinization by considering different modes of mass transfer mechanisms transporting solute in unsaturated soil (such as D. L. Corwin et al. (2007); Schoups et al. (2006)); however, the application of these models remains limited to small-scale simulations where the detailed soil characteristics data are available. Moreover, employing analytical approaches, such as the stochastic model of soil salinity (Perri et al., 2018; Suweis et al., 2010) or the developed frameworks for mechanistic modelling of the climate, vegetation, and soil salinity interactions (Mau & Porporato, 2015; Porporato et al., 2015; Runyan & D'Odorico, 2010), would be applicable for projecting soil salinity only if the initial soil salinity or required calibration parameters for tuning were available; currently, such data are not available, particularly on a global scale.

Recent studies demonstrated the great potential of Machine Learning (ML) algorithms in digital soil mapping and predicting spatio-temporal properties of the soil (Padarian et al., 2020). In the present study, we used supervised ML algorithms for projecting the long-term (up to year 2100) variations in soil salinity. In summary, the methodology included exposure of a known set of input data (predictors) and a set of known responses (soil salinity profiles) to ML models to develop trained models based on the relations between the two sets. The trained models were later applied to a new set of known input data (with unknown responses) to generate predictions for the response (see Methods).

Dryland areas are generally known as the regions with the highest vulnerability to hydro-climatic consequences of climate change (I. Daliakopoulos et al., 2016). For this reason, the majority of our measured input soil profiles data were sampled from the dryland areas of the world. We made predictions of soil EC_e only for the dryland areas with an AI ≤ 0.65 (UNEP-WCMC, 2007) as extrapolation of the ML results to other areas is a matter of uncertainty (Hengl et al., 2017). The rest of this paper discusses the significance of the predictors and global variation in primary soil salinization at the grid-cell level, followed by the country-level analysis. Changes in the total area of drylands with an $EC_e \geq 2$ dS m^{-1} (and $EC_e \geq 4$ dS m^{-1}) at the country and continental levels are also presented. Finally, methods and their limitations are discussed.

4.2 Results

4.2.1 Predictors' significance and their relation to the predicted soil salinity

Appendix 3, Table A3-1 shows the estimates of the predictor importance for the trained models based on the output of the GCMs used for spatio-temporal prediction of the EC_e (see Methods for details of predictors and trained models). The percentage values reported in Appendix 3, Table A3-1indicate the relative importance of each predictor in the final trained model in each input dataset. Among the 14 applied predictors, the long-term (five-year average) annual precipitation frequency is relatively the most influential soil predictors with an overall importance of 14% for all 16 best fitted models. WRB soil classes and daily evapotranspiration are respectively the second and the third influential environmental predictors in estimation of the soil EC_e with the overall importance of 13.07% and 9.26%, respectively.

The effect of each of the 13 non-categorical predictors (see Methods) on the predicted outcome of the trained models is shown in Appendix 3, Figure A3-1 (Partial Dependency Plots, PDPs). The effect of long-term daily wet and dry deposition rates of sea salts are presented in Appendix 3, Figure A3-2. Appendix 3, Figure A3-1, a and b suggest that shallower depths are not necessarily associated with higher EC_e in soil under natural conditions. However, in many previous experimental, analytical and numerical investigations (Guglielmini et al., 2008; Huinink et al., 2002; Rad et al., 2015; Shokri-Kuehni et al., 2017; Shokri, 2014), higher solute concentrations and solute precipitation close to the evaporation surface were observed when the Peclet number (quantifying the relative importance of chemical diffusion and advection) was greater than the one during saline water evaporation from porous media. It must be noted that under natural environmental conditions (which is the case in our investigation), many parameters influence the complex dynamics of solute transport and deposition in soil, including the vegetation and land cover, rainfall, microorganisms' activities, depth of water tables, soil chemical compositions and heterogeneity, human interventions, and land-atmosphere interactions. These parameters, which could not be included in the majority of the previous experiments conducted under well-controlled laboratory conditions or numerical simulations, could induce significant impacts on solute distribution in soil under natural conditions (Abrol et al., 1988; Scudiero et al., 2014).

Fine-textured soils (soils with the higher clay content) show higher Water Holding Capacity (WHC, the difference between field capacity and wilting point) and lower saturated hydraulic conductivity. Overall, the predicted EC_e values provided by each of the 16 trained models show a reverse relation with the soil clay content and WHC which is in line with previous experimental results and a literature review (Li et al., 2014) (Appendix 3, Figure A3-1, c, g, and h). Similarly, based on numerical, experimental, and field-scale investigations, Shokri-Kuehni et al. (2020) concluded that soil salinity for coarse-textured soils is greater than for medium and fine-textured soils when the water table is shallow and hydraulically connected to the evaporation surface. Our predicted results regarding the effects of soil texture on soil salinity are generally in agreement with the above mentioned physically-based determined trends and behaviour.

Moreover, the analysis of PDPs shows that the effective plant rooting depth influences the predicted EC_e approximately up to the depth of 4 m. The PDPs also demonstrate a strong negative correlation between soil salinity and terrain elevation, topographic slope, and precipitation frequency (Appendix 3, Figure A3-1, d, e, and i). These correlations can be explained by the prior pedologic knowledge: the lower hillslope and the higher precipitation frequency result in more efficient leaching of the salts accumulated in the root zone (Suweis et al., 2010), resulting in lower salinity. The relationship between the predicted soil EC_e values and other predictors, however, is more complicated and deriving general trends remains a challenge.

4.2.2 Projected soil salinity in drylands up to the year 2100

The trained models based on the output of Global Circulation Models (GCMs) were applied to new input predictor data to estimate the annual soil salinity for each grid-cell (0.5° spatial resolution) of the global soil base map of the drylands between 1904 and 2100 (see Methods for details of GCMs, predictors, and trained models). Figure 4-1 shows the spatial distribution of the change in primary soil EC_e projected by the multi-model ensembles in the mid- (2031 - 2060) and long-terms (2071 - 2100), relative to the reference period (1961 - 1990) at the 0.5° spatial resolution. The RCP 4.5 and RCP 8.5 scenarios (Representative Concentration Pathways which result in a respective radiative forcing of 4.5 and 8.5 W m^{-2} in year 2100, relative to pre-industrial conditions) are related to CMIP5 (Coupled Model Inter-comparison Project Phase 5 (Taylor et al., 2012)) data project, while the SSP 2-4.5 and SSP 5-8.5 scenarios (projections forced by RCP 4.5 and RCP 8.5 global forcing pathways for the Shared Socio-economic Pathways 2 and 5) refer to CMIP6 (CMIP Phase 6 (Eyring et al., 2016)).

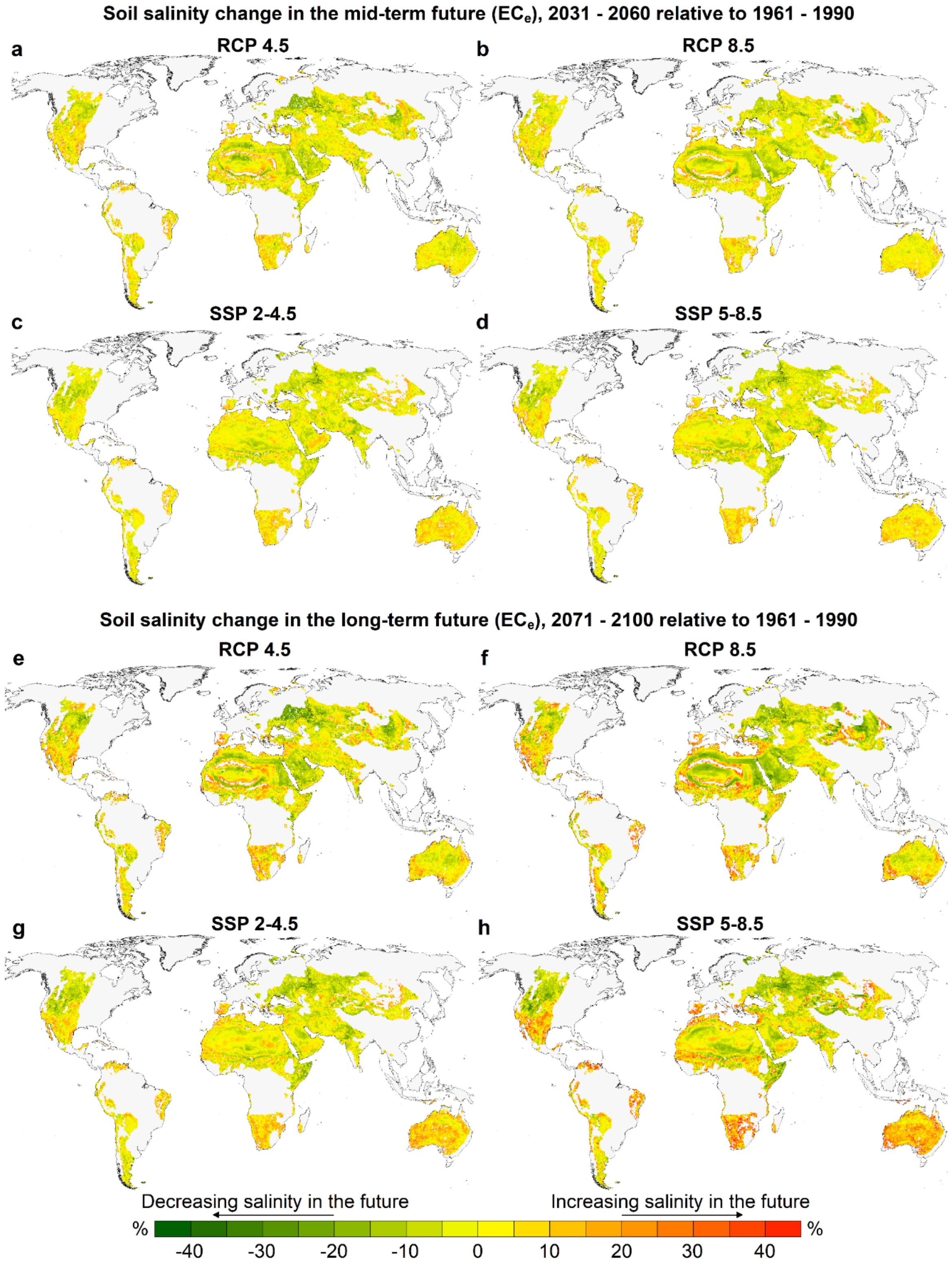


Figure 4-1: Multi-model ensemble mean of the change in predicted soil salinity represented by saturated paste electrical conductivity (EC$_e$) in the mid- and long-term futures, relative to the reference period (1961 - 1990) under different greenhouse gas concentration trajectories. **a** to **d**, mid-term prediction of changes in EC$_e$ (2031 - 2060). **e** to **h**, long-term prediction of changes in EC$_e$ (2071 - 2100). The average of the predictions to the depth of 1 m were used for calculations of salinity change. At each map cell (pixel) and based on each GCM, we calculated the mean of soil salinity for the reference, mid-, and long-term future periods and then computed the relative change as: *(Future mean - Reference mean)/Reference mean*; the percentage value of each

cell represents the multi-GCM mean of the calculated relative changes presented by the colour map. Positive values indicate an increase in soil salinity while the negative values are indicative of a decreasing trend.

Our results reveal that the sign (positive: indicative of a higher EC_e and negative: indicative of a lower EC_e) and intensity of changes in primary soil salinity are geographically highly variable; the variations are more extreme at the end of the 21th century compared to the mid-term future. Generally, the relative changes in soil salinity are more severe for the GHG emission rates which result in higher radiative forcing scenarios (RCP 8.5 and SSP 5-8.5). However, the intensity and spatial distribution of the projected changes based on the CMIP5 models are not necessarily the same as the CMIP6-based models predictions. Although our aim was to include all available projections in the analysis, in the case of discrepancy between CMIP5 and CMIP6 models, the predictions made based on the CMIP6 GCMs should be prioritised as they are more recent, forced by more updated data, and generally of higher spatial resolutions (Eyring et al., 2016).

According to our long-term predictions based on all multi-model ensembles, the drylands areas of South America, southern Australia, Mexico, southwest United States, and South Africa are generally at the highest risk of increased soil salinity, compared to the reference period. The threat of climate-induced soil salinity is also projected to increase in drylands of Spain, Morocco, and northern Algeria. To a lesser extent, western and southern Sahara and central Indian drylands, in addition to the desert soils of southeast Mongolia and north of China, are estimated to become saltier in response to the projected climate change by 2100 for different GHG concentration trajectories. On the other hand, our results indicate that the extent of soil salinity will remain constant or decrease relative to the reference period in the drylands located across the northwest United States, the Horn of Africa, Eastern Europe, Turkmenistan, and west Kazakhstan.

Additionally, Appendix 3, Figure A3-3 shows the long-term future relative change in the five-year moving averages of daily dry and wet deposition rates of the sea salts (the 1971 - 2100 mean minus the 1961 - 1990 mean) projected by the multi-GCM ensemble means, as the two predictors used for training the models. Overall, the CMIP6 models predict a more severe increase or decrease in dry and wet deposition rates; however, all ensemble means are in agreement on an increasing trend in the dry deposition rate of sea salts in coastal regions, particularly in the southern hemisphere. All models also project a decreasing trend in dry deposition rates in north-western United States, west Canada, and central Asian regions; however, for these locations, the projected sign of the change in wet deposition rates is different between the CMIP5 and CMIP6 models. To some extent, the projections of these deposition rate can explain why soil salinity decreases in some regions, e.g. central Asia and Kazakhstan, where there is less certainty on the projected sign of changes in precipitation and evapotranspiration (Miao et al., 2020).

Not all of the predictions generated based on the CMIP5 and CMIP6 GCMs used in this study are in agreement on the extent and sign of the soil salinity by the end of the century. Figure 4-2, in particular, shows the multi-model ensemble agreements on the sign of the predicted change in soil salinity in the long-term future under different trajectory scenarios of GHG concentration. A cell value close to 100% indicates a complete agreement of the ensemble members on the sign of the salinity change. For the RCP 4.5 ensemble, as an example, an ensemble agreement of 100% of a grid-cell shows that all seven models in the ensemble are predicting an increase or a decrease in soil salinity in the long-term future relative to the reference period (depending on the sign of change). Especially under the SSP 2-4.5 and SSP 5-8.5 scenarios, the multi-GCM certainty of the predictions for a great proportion of the drylands of southern/eastern Australia, South America, and southern Africa indicate the southern hemisphere is at a higher risk of salinity caused by climate change. The projected increase in soil salinity in south-west and southern Australia induced by rising shallow groundwater tables as a result of dryland resource management and activities (National Land and Water Resources Audit, 2001) can exacerbate the climate-induced soil salinization projected here. However, the certainty of the predictions made for drylands located in the Middle East, Russia, and Sahara is seemingly lower than for the other zones. For those dryland regions, the uncertainty is also recognisable through the difference in the predictions made based on the CMIP5 and CMIP6 models in Figure 4-1. For example, the CMIP5 models predict an increase in soil salinity in Russian drylands, while the CMIP6 models show the opposite trend in those regions.

4.2.3 Country-level projected changes in soil salinity

At the country level, we calculated descriptive statistics for the relative changes in soil salinity estimated at each grid-cell (in the mid- and long-term futures compared to the reference period) based on the multi-model ensemble mean, including grid-cells mean, 95% confidence intervals of the mean, standard error of the mean, and variance (Appendix 3, Table A3-2 - 9). We did not calculate these descriptive statistics at the continental level as there was no noticeable difference between the results for various continents due to the high number of grid-cells within each continent.

Although the country-level results mask the majority of the local-scale variabilities of the soil salinity, the provided statistics help to have a better understanding of the countries with the highest risk of salinization. We ranked the countries based on the total number of grid-cells located in each country and calculated all aforementioned statistics only for the 30 countries with the highest number of grid-cells (Appendix 3, Table A3-10 shows the top 30 countries and the total estimated area of their drylands).

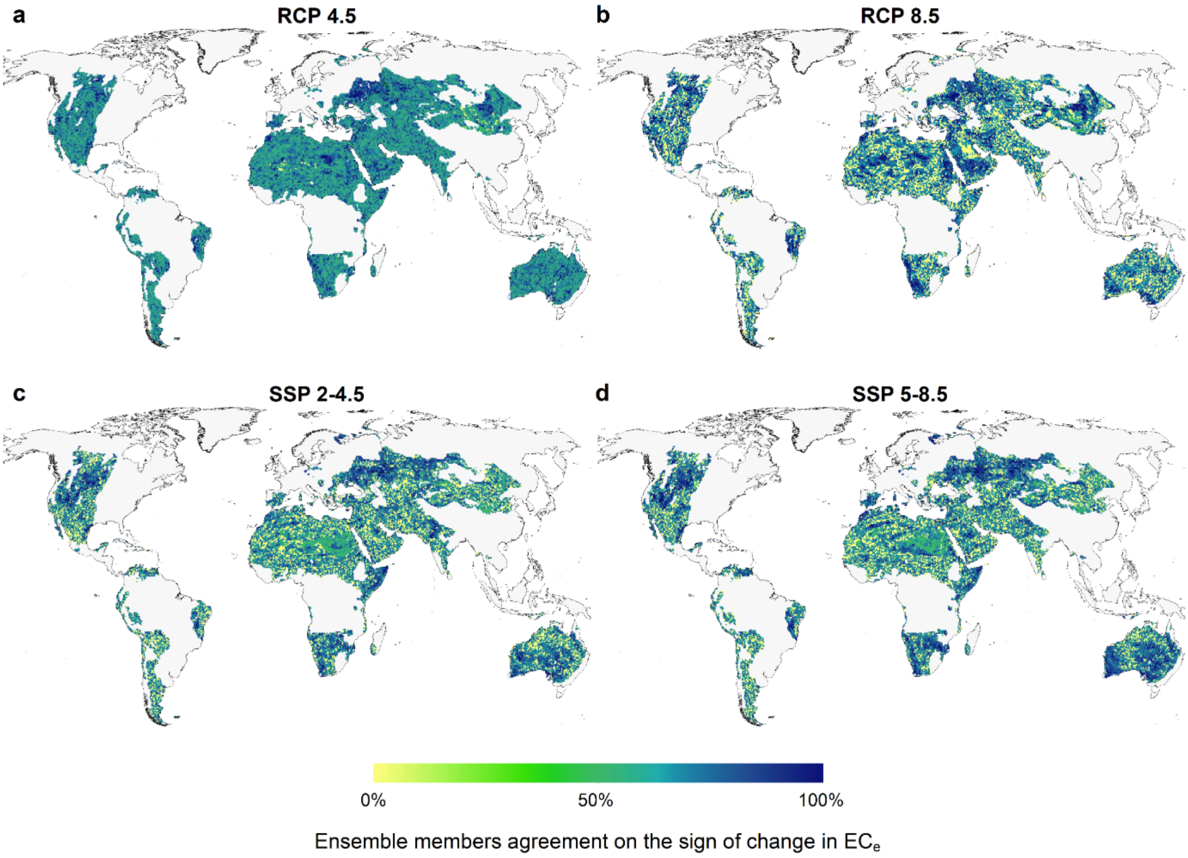


Figure 4-2: Multi-GCM ensemble agreement on the sign of change in predicted values of soil EC$_e$ in the long-term future (2071 - 2100), relative to the reference period (1961 - 1990) under different greenhouse gas concentration trajectories. 100% shows the full agreement of the models on the sign of change, while zero indicates inconsistency among the models' predictions.

For the 2071 - 2100 period relative to 1961 - 1990 and under RCP 8.5 as the worst case scenario, the countries with the highest relative increase in the soil salinity were Brazil (with a mean grid-cell increase in EC$_e$ of 15.1% and the 95% confidence intervals of 13.25% - 16.95%), Namibia (13.57%; 12.1% - 15.04%), South Africa (11.2%; 9.41% - 13%), and Mexico (6.38%; 4.96% - 7.8%). The increase in soil salinity for Australia was much lower (3.31% and 2.88% - 3.73%). Under SSP 5-8.5, the countries with the highest relative increase in grid-cell means of soil salinity in the same period were Botswana (24.94%; 22.71% - 27.16%), South Africa (21.35%; 19.84% - 22.85%), Namibia (17.69%; 16.14% - 19.24%), and Brazil (16.21%; 14.77% - 17.66%). Overall, our calculated statistics suggest that the soil salinity will be increased more extensively by the climate change impacts in the regions spread across the southern latitudes, specifically below -20°.

4.2.4 Change in the total area of salt-affected soils in drylands

Additionally, based on our predictions for soil salinity extent in each grid-cell, we estimated the total area of salt-affected soils up to year 2100. Currently, no unique definition is available for the salt-affected soils. Contingent on the soil classification system, different values of EC_e, ranging from 2 dS m^{-1} to even 30 dS m^{-1}, are adopted as the minimum threshold of salinity for characterising the saline soils (Gupta et al., 2008; Richards, 1954; Soil Survey Staff, 2010). Accordingly, here we quantified the areal variation of the soils exposed to the threat of primary salinization assuming an EC_e equal to 2 dS m^{-1} as the critical threshold, corresponding to the upper salinity limit tolerable by sensitive crops (Maas & Grattan, 1999). The results were computed at the country (Appendix 3, Table A3-11), continental (Table 4-1, Figure 4-3, Appendix 3, Figure A3-4), and global levels (Appendix 3, Figure A3-5). Additionally, Appendix 3, Figure A3-6 to 8 and Appendix 3, Table A3-12 and 13 show the projected variation in the total area of naturally-occurring salt-affected soils assuming 4 dS m^{-1} as the critical threshold at the continent and country levels. As before, at the country level, only the top 30 countries with the highest number of the grid-cells were included. This analysis could be an indicator of the spatial expansion of the soil salinity in drylands in response to climate change.

Table 4-1: Continental-level predicted change in the total area of soils with $EC_e \geq 2$ dS m^{-1} in the mid- and long-term futures relative to the average of the 1904 - 1999 period under different greenhouse gas concentration trajectories.

Scenarios	Continent					
	Africa	Asia	Australia	North America	Europe	South America
RCP 4.5, mid-term (%)	0.00	-1.03	0.02	-0.23	-6.58	2.35
RCP 4.5, long-term (%)	0.17	-2.02	0.70	-0.33	-9.13	1.84
RCP 8.5, mid-term (%)	0.02	-1.36	0.79	0.13	-2.55	2.21
RCP 8.5, long-term (%)	-0.02	-3.05	0.60	0.83	-5.35	4.88
SSP 2-4.5, mid-term (%)	0.41	-0.05	1.59	-3.32	-2.09	2.56
SSP 2-4.5, long-term (%)	0.61	-0.25	2.40	-2.89	-2.68	3.04
SSP 5-8.5, mid-term (%)	0.51	0.02	1.36	-2.28	-1.90	3.60
SSP 5-8.5, long-term (%)	1.45	-0.28	3.38	-2.45	-0.92	6.70

Overall, under emission rates resulting in the radiative forcing of 8.5 W m^{-2}, all CMIP5 and CMIP6-derived predictions indicate an increasing trend in the total area of dryland soils with an $EC_e \geq 2$ dS m^{-1} for Australia and South America and a decreasing trend for Asia and Europe relative to the average of 1904 - 1999 period. For Australia and South America, we estimate the respective increases of 3.4% and 6.7% in the total area of dryland soils with $EC_e \geq 2$ dS m^{-1} between 2071 - 2100 relative to 1904 - 1999 period according to the multi-GCM ensemble means under the SSP 5-8.5 scenario. The CMIP5 and CMIP6-derived predictions of the total area of dryland soils with $EC_e \geq 2$ dS m^{-1}, however, are not in agreement on the sign and extent of the change for Africa and North America. The multi-model ensemble means under the SSP 5-8.5 scenario predict an increase of 1.5% and a decrease of -2.5% for the total area of dryland soils with a salinity ≥ 2 dS m^{-1} located in Africa and North America (between 2071 - 2100 relative to 1904 - 1999), respectively. Brazil (with 43%), Mexico (14.5%), and Mongolia (8%) had the highest estimated expansion in the total area of dryland soils with a salinity ≥ 2 dS m^{-1} between 2071 - 2100 relative to 1904 - 1999 periods under SSP 5-8.5 at the country level. On the opposite side of the continuum, Canada (with -10%), Somalia (-8.5%), and Ethiopia (-5%) had the largest predicted shrinkage of saline soils under SSP 5-8.5 (among the top 30 counties with the highest number of grid-cells in our analysis).

4.3 Discussion

The results obtained here do not agree with the global scale predictions of Schofield and Kirkby (2003) who estimated that Australia and western North America would be the areas with lower salinization potential in the 2070 - 2099 period, while they predicted a high potential for salinization in lands across Eastern Europe and Kazakhstan. In addition to the difference between the methodologies used for the projections of soil salinity, this discrepancy is due to various other reasons. For example, unlike the current study, Schofield and Kirkby (2003) only used one GCM (HadCM3GGa), developed before 2000, to specify their salinization indicators. Furthermore, they estimated the future potential evapotranspiration as an empirical function of air temperature to calculate AI as an indicator of soil salinity, while we used total evapotranspiration derived from the more physically-based GCMs.

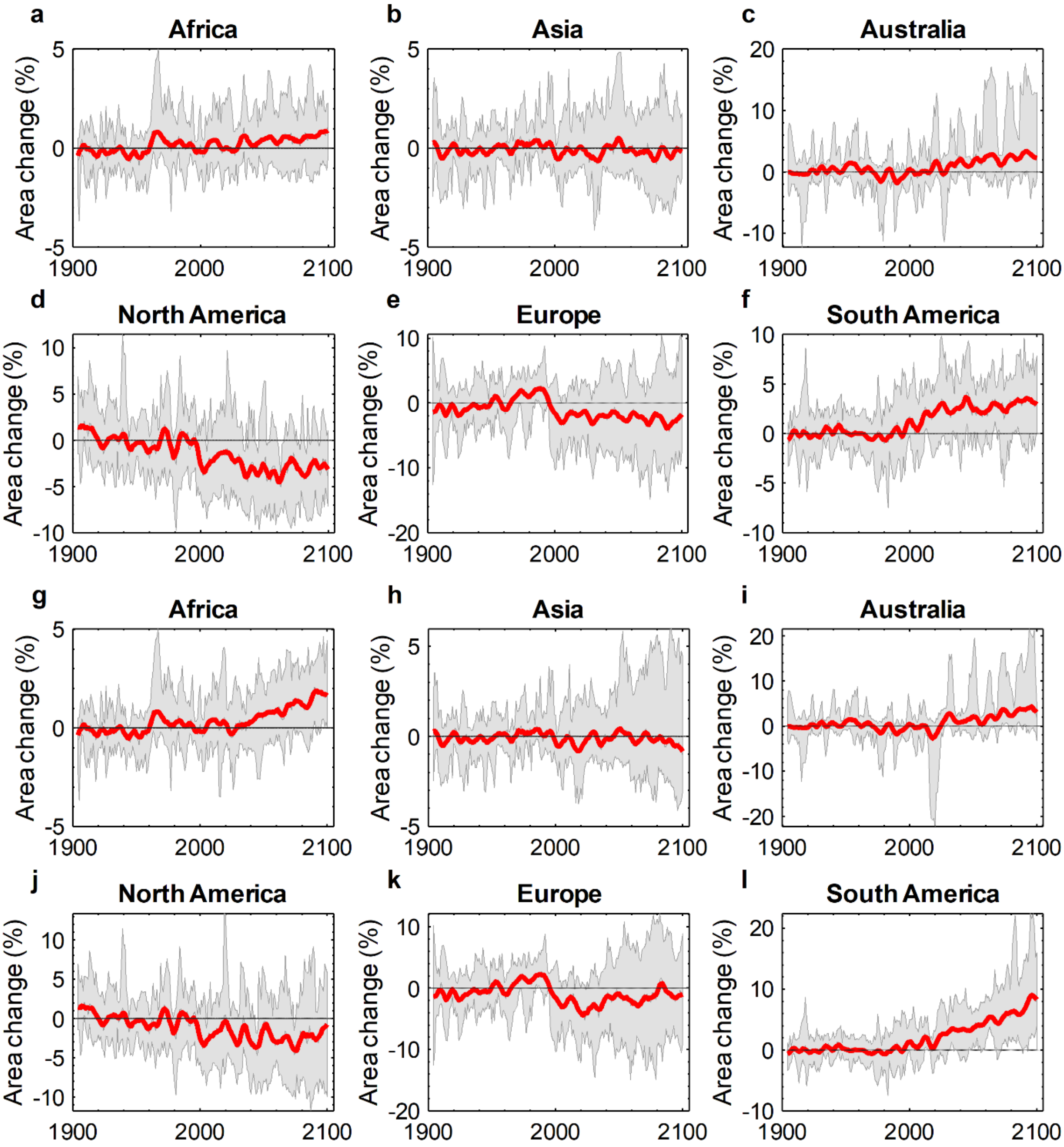


Figure 4-3: Continental-level predicted annual change in the total area of soils with an $EC_e \geq 2$ dS m^{-1} relative to the 20th century average (1904 - 1999) for the models obtained from the CMIP6 data project. a to **f**: Relative change under SSP 2-4.5 greenhouse gas concentration trajectory. **g** to **l**: Relative change under SSP 5-8.5 greenhouse gas concentration trajectory. Shaded areas show the minimum and maximum range of the relative changes predicted by multi-model ensemble members. Red lines show the low-pass filtered (5-year running window) of the multi-model ensemble mean of the predicted variations; since all spatio-temporal predictors are five-year moving averages, 1904 is the beginning of the period.

The results of ML models are primarily based on the trends they capture from the input data used for training. Therefore, projected changes of EC_e in the hotspots of climate-induced soil salinization can be mainly attributed to the variations in spatio-temporal input data projected by GCMs. As mentioned before, precipitation frequency and evapotranspiration were the most influential spatio-temporal predictors for the predictions of the trained models. According to the analytical salt mass balance, higher evapotranspiration rate and precipitation with a lower frequency and intensity accumulate more salts in the root zone (Suweis et al., 2010). By the end of the century, an ensemble mean decrease in precipitation (under RCP 8.5) of up to 40% was reported by Giorgi et al. (2019) for the southern hemisphere, particularly southern and Western Australia, Namibia, and Brazil for the June-July-August months, which are also the salinization hotspots according to our results. Similarly, in the northern hemisphere, they predicted a more severe decrease in precipitation for Mexico, West Africa, and Mediterranean coasts for December-January-February. At smaller spatial scales, other studies projected an increase in the number and duration of drought events, higher potential and actual evapotranspiration, decreasing trends in frequency and intensity of precipitation, and in general drier conditions by the mid and end of the century.

Using 34 GCMs under the two different emission scenarios of RCP 4.5 and RCP 8.5, Shi et al. (2020) predicted that potential evapotranspiration tends to increase in south-eastern Australia. Likewise, using 22 CMIP5 models, a substantial increase in the number of warm temperature extremes and periods of dryness was projected by Alexander and Arblaster (2017) for Australia, one of the predicted salinization hotspots in the current study. Similar trends for Australia were projected by Grose et al. (2020) by analysing the available CMIP6 multi-model ensemble. By analysis of 14 GCMs under the RCP 4.5 and RCP 8.5 future scenarios, a substantial decrease in precipitation during the summer (up to 1.5 mm day^{-1}) is expected by Colorado-Ruiz et al. (2018) in southern Mexico, also a projected salinization hotspots in the present study. A decrease in the frequency of precipitation during winter and spring in south-western United States is projected by Easterling et al. (2017), as also found in this study to be a hotspot. An increase in the number of consecutive dry days in West Sahara (Klutse et al., 2018) and actual evapotranspiration in arid areas across north-western China (Ma et al., 2018) under the 1.5 °C and 2.0 °C global warming scenarios reported in the literature is congruent with the findings of the current study.

To conclude, lack of reliable predictive tools and data to assist land managers and policy makers for understanding the land cover dynamics is one of the main obstacles to long-term sustainable land and environment management. In the present study, we used legacy soil-profiles data and a set of purely spatial and spatio-temporal predictors to develop some predictive ML models for projection of the primary soil salinity (represented by electrical conductivity) as one of the major threats to the soil fertility, stability, and bio-diversity in world drylands. Our analysis provides long-term gridded (at 0.5° spatial resolution) predictions of primary soil salinity change in drylands globally in response to projected key climatic drivers of soil salinity, which is currently missing in the soil and land management literature. In the face of projected future climatic uncertainties, the developed predictive models and generated data in the present investigation can help with decision-making regarding land and water resources management to recognise the hotspots of soil salinization, devise the necessary action plans, and implement those plans towards sustainable land and water resources management.

Under different GHG concentration trajectories, our predictions suggest that by the late 21th century the drylands areas of South America, southern Australia, Mexico, southwest United States, and South Africa are at the risk of higher soil salinity caused by climate change, compared to the reference period (1961 - 1990). In addition, increase in climate-induced soil salinity threatens the drylands of Spain, Morocco, and northern Algeria by the end of the century. On the other hand, our results project a decreasing trend in primary soil salinity of the drylands located in the northwest United States, the Horn of Africa, Eastern Europe, Turkmenistan, and west Kazakhstan, relative to the reference period. The reliability of the predictions made here are different: the projected soil salinities for the drylands located in North America and Australia are of the highest level of reliability while the drylands of central Asia, Middles East, and the Great Sahara have the highest uncertainty in predictions for soil salinity. Other zones such as India, South America, and South Africa are in the middle in terms of the reliability of predictions.

4.4 Methods

In a previous study (Hassani et al., 2020), we developed tree-based two-part predictive ML models for determining annual surface (referring to top 30 cm of the soil) soil salinity and sodicity (represented by exchangeable sodium percentage) over the past four decades (1980 - 2018) at ~1 km^2 spatial resolution on a global scale. In the present study, however, we aimed to predict the future dynamics of soil salinization up to the year 2100 under changing climate. In the present investigation, we focused on primary salinization and the trained tree-based ML models were only regressive models. The next sections explain the details of the workflow for predicting soil salinity (EC_e) including: (1) collection of the measured soil-salinity profiles, (2) collection and processing of salinity predictors, (3) exposing the salinity profiles and predictors data to ML models, training the models, and validation of the trained models, and (4) employing the trained models to project the spatio-temporal variation of the soil EC_e up to the year 2100 under different greenhouse gas (GHG) concentration trajectories. Finally, we discuss the accuracy of the trained models for prediction of EC_e.

4.4.1 Soil-salinity profiles

We obtained the geo-referenced soil profiles (points) with measured values of EC_e from the soil profile dataset of World Soil Information Service (WoSIS) (Batjes et al., 2017). The spatial distribution of the profiles data used as an input into the ML models is presented in Figure 4-4, a. The WoSIS EC_e database includes 19,434 soil profiles and each individual profile (with a unique profile ID) may include one or more samples for various depths below the soil surface. The data cover the sampling period from 1950 to 2014. Since the date of sampling was an essential parameter in model training, we removed the EC_e profiles without sampling dates. This reduced the total number of EC_e samples from 73,517 to 59,649, with the number of samples per year shown in Figure 4-4, b. In addition, we dropped the soil EC_e profiles sampled from the croplands to remove the effects of human interventions from the analysis. As a result, a total 44,708 samples (11,517 profiles) remained in our analysis for model training and accuracy assessment.

Global land cover data provided by Earth-Observation Satellites before 1997 were scarce. Accordingly, we divided the profiles into two categories based on the date of sampling: before 1997 and after 1997. For the period before 1997, we identified the profiles located in croplands using the Global Land Cover Characteristics Database, Version 2.0 at ~1 km resolution (Belward et al., 1999). Due to a lack of historical land cover data, we assumed that the land cover/land use did not change considerably before the 1980s. For profiles sampled after January 1997, however, we identified the samples/profiles located in croplands using land cover maps for years 2000, 2006, 2014, and 2018 with similar International Geosphere-Biosphere Programme (IGBP) land cover legend adopted from the MODIS Data Collection (MCD12Q1 and MCD12C1) (Sulla-Menashe & Friedl, 2018). We selected the IGBP land cover legend as it was available in both datasets. Each profile sampling date was attributed to the layer with the nearest year of acquisition. The MODIS land cover layers were first re-projected to the World Geodetic System (WGS 1984) spatial coordinates at 0.004° (~500 m) using the nearest neighbour method.

4.4.2 Predictors

We used two types of predictor to train the models for predicting EC_e as the target variable: purely spatial and spatio-temporal. Purely spatial predictors included the land and soil attributes which were relatively constant during the period of the analysis, while spatio-temporal predictors were the large-scale hydro-climatic variables derived from the output of selective Global Circulation Models (GCMs). In total, 14 predictors were used, of which nine purely spatial and the rest spatio-temporal. The pre-processing details, projection, extent, and resolution of the predictors' layers are summarised in Table 4-2. These predictors were primarily selected to represent the main factors affecting the salt balance in the root zone in non-irrigated soils (Suweis et al., 2010). In addition, we included in our model training additional soil formation factors, including topography and parent material (weathered rock or deposit from which the soil is formed) (Batjes et al., 2017; Jenny, 1994).

The purely spatial predictors comprised:

- soil classes based on the World Reference Base (WRB) classification (IUSS Working Group WRB, 2015);
- soil texture represented by the percentage of clay content, obtained from the ISRIC global gridded soil information at ~250 m spatial resolution (Hengl et al., 2017);
- soil wilting point in mm (Global Soil Data Task Group, 2000);
- soil field capacity in mm (Global Soil Data Task Group, 2000);
- effective plant rooting depth in m (Yang et al., 2016);
- topographic slope in degrees; and
- terrain elevation in m.

Slope and terrain elevation layers were derived from the World Elevation Terrain data adopted from ArcGIS Living Atlas of the World (Esri, 2020) and were re-projected to the WGS 1984 coordinates system at 0.002° (~250 m) spatial resolution using the cubic convolution method.

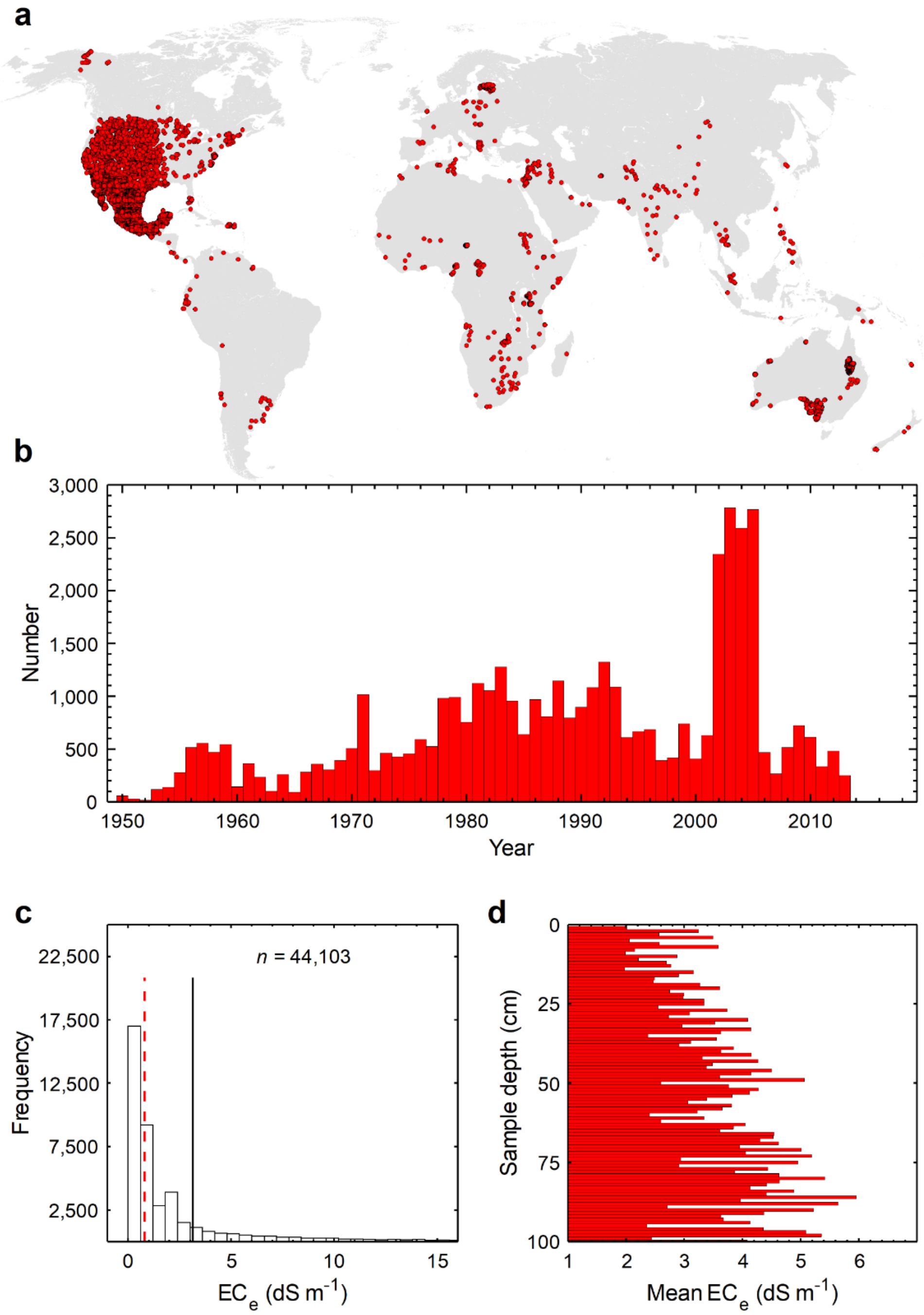


Figure 4-4: General properties of the EC$_e$ profiles used for training the models. a, spatial distribution of the soil salinity profiles used for model training and prediction of the soil salinity. Each profile includes one or more soil samples. **b**, temporal distribution of the samples used for training the predictive models of soil salinity. Each bar shows the number of samples within one year. **c**, frequency distribution of the measured values of EC$_e$. The solid and dashed vertical lines represent the mean and median values, respectively. **d**, average of the measured soil salinity values at 1 cm intervals to the depth of 1 meter below the surface.

Table 4-2: Purely spatial and spatio-temporal predictors used for model training and prediction of soil salinity.

Purely spatial predictors					
Predictor	**Pre-processing**	**Source**	**Projection**	**Source layer spatial range**	**Spatial resolution**
Sample upper depth (cm)	-	Original soil dataset	-	-	-
Sample lower depth (cm)					
Elevation (m)	Resampling the original DEM to ~ 250 m resolution by the cubic convolution method	World Elevation Terrain service (Imagery Layer from Esri)	WGS 1984 Web Mercator (Auxiliary Sphere)	Left: -20,037,507.84 m Right: 20,037,507.90 m Bottom: -20,037,508.41 m Top: 20,037,508.34 m	0.25 m
Slope (degrees)	Resampling the original DEM to ~ 250 m resolution by the cubic convolution method and then calculating the slope using ArcGIS "slope" function				
World Reference Base soil classes (120 classes)	-	ISRIC-SoilGrids250	GCS WGS 1984	180W-180E, 62S-87.37N	0.00208°
Soil clay content (%)	Per-cell average of five standard soil depths: 0, 15, 30, 60, and 100 cm was calculated using the trapezoidal rule and ArcGIS "cell statistics" tool				
Field capacity (mm)	Raster datasets for different continents were merged into a single global one	Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS)	GCS WGS 1984	180W-180E, 56.49S-90N	0.00833°
Wilting point (mm)					
Effective plant rooting depth (m)	The original dataset was geo-referenced to the GCS WGS 1984 coordinates system by the nearest neighbor method	Yang et al. (2016)	GCS WGS 1984	180.25W-179.75E, 90.25S-89.75N	0.5°
Spatio-temporal predictors					
Predictor	**Pre-processing**		**Projection**	**Extent**	**Source and spatial resolution**
Five-year moving average of annual precipitation frequency (day^{-1})	Precipitation frequency (λ) was calculated by dividing the number of wet days (daily precipitation > 1 mm) by the total number of days of a year (T). Precipitation fluxes (kg m^{-2} s^{-1}) were transformed to daily sums by multiplying by a factor of 86,400		GCS WGS 1984	180W-180E, 90S-90N	Global Circulation Models (GCMs) presented in Table 4-3
Five-year moving average of annual precipitation intensity (cm)	Precipitation intensity was calculated by $\alpha\lambda T$ = Annual accumulative precipitation, where α was the precipitation intensity, λ was precipitation frequency, and T was the total number of days of a year (365)				
Five-year moving average of daily evapotranspiration (cm day^{-1})	First an annual average was calculated from monthly evapotranspiration fluxes (kg m^{-2} s^{-1}). Then the annual average flux was transformed to daily sum by multiplying by a factor of 8,640				
Five-year moving average of daily dry deposition rate of sea salts (mg day^{-1} m^{-2})	First an annual average was calculated from monthly dry deposition rates (kg m^{-2} s^{-1}). Then the annual average flux was transformed to daily sum by multiplying by a factor of 86,400				
Five-year moving average of daily wet deposition rate of sea salts (mg day^{-1} m^{-2})	First an annual average was calculated from monthly wet deposition rates (kg m^{-2} s^{-1}). Then the annual average flux was transformed to daily sum by multiplying by a factor of 86,400				

We filled the missing grid-cells (or cells with no data values) in purely spatial predictor layers with an average from the cells surrounding the missing grid-cell. We used a circle with a radius of four cells from the neighbouring cells to calculate the average and fill the data gap. All purely spatial predictors were assumed to be vertically constant. Rasters processing was conducted in ArcGIS 10.7 (Desktop, 2011). Then, we obtained the values of grid-cells of purely spatial predictors at the locations of EC_e profiles (Figure 4-4, a) to later train predictive models of soil salinity (see "model training for prediction of soil salinity"). The upper and lower depths of the measured EC_e samples derived from the original WoSIS database were the additional purely spatial predictors used for model training; these were introduced to account for the effect of depth on soil salinization processes.

The spatio-temporal predictors considered here were precipitation intensity, precipitation frequency, daily evapotranspiration, and sea salts wet and dry deposition rates (Table 4-2). To make predictions for future periods, we needed the projected values of the predictors. Therefore, we derived the values of spatio-temporal predictors from the outputs of the GCMs under different GHG concentration trajectories.

For training the models, we used the GCMs available in both CMIP5 and CMIP6 data projects to consider the uncertainty in GCMs predictions and to cover all available projections for dry and wet sea salt deposition rates. Additionally, this gave us the opportunity to analyse the differences between the CMIP5 and CMIP6 model outputs in terms of the derived predictors' values and their effects on the projected soil salinity. The historical outputs of GCMs, including precipitation, evapotranspiration, and dry and wet deposition rates of sea salts, were used for training the predictive ML models (CMIP5: 1900 - 2005; CMIP6: 1900 - 2014). The projected outputs of GCMs for the same parameters were used to make future predictions of soil salinity (CMIP5: 2006 - 2100, CMIP6: 2015 - 2100). For the CMIP5 models, predictors were calculated based on the future projections forced by the RCP 4.5 and RCP 8.5 scenarios. Likewise, for the GCMs models of CMIP6, predictors were computed using future projections forced by RCP 4.5 and RCP 8.5 global forcing pathways for the Shared Socio-economic Pathways (SSP) 2 and 5, respectively. These medium (4.5) and high (8.5) radiative forcing pathways were chosen because they respectively represent the most plausible (or stabilization) and worst case scenarios of emissions by the end of the 21th century.

Since the total number of wet days and the total annual precipitation values were calculated from the daily precipitation fluxes, the GCMs with precipitation data at daily resolution were required. Additionally, not all of the available GCMs in the CMIP5 and CMIP6 projects had the dry and wet deposition rates of the sea salts. Accordingly, our analysis was narrowed down to a total of 16 GCMs outputs under different GHG concentration trajectories from both CMIP5 and CMIP6 projects. For the GCMs with different ensemble members (MIROC5 and CESM2-WACCM-gn, in particular), we computed an ensemble mean to avoid a bias in the results of final multi-GCM ensembles toward the GCMs with the higher number of participating ensemble members. In total, data of 16 GCMs were downloaded from the CMIP5 and CMIP6 data (Cinquini et al., 2014), respectively. Details on the final chosen GCMs, their spatial resolution, and their used ensemble members are presented in Table 4-3.

The original longitude values of netCDF files were set in the range -90° and 90°, referenced to the Greenwich Prime Meridian, to be in the same spatial extent as the purely spatial predictors. Then, using the bilinear interpolation method, all were interpolated to 0.5° × 0.5° WGS 1984 longitude-latitude regular grid to be able to generate multi-GCM ensemble from the outputs of our predictive models. Calculation of the spatio-temporal predictors and processing of the original netCDF files were conducted in the Climate Data Operators (Schulzweida, 2019) environment. The prepared netCDF data based on the outputs of GCMs were then converted to multi-band rasters, after which we obtained the values of spatio-temporal predictors at locations of EC_e profiles. These values combined with the values of purely spatial predictors were used to train the predictive models of soil salinity. It was not practical to use these spatio-temporal predictors at annual temporal resolutions because the salt level reaches a steady state condition usually at much longer time scales (Suweis et al., 2010). Therefore, we used a five-year moving average instead to better capture the effect of intra-annual trends in these predictors on soil salinity variations. Finally, the five-year moving averages of the spatio-temporal predictors were attributed to each observation of EC_e according to the year of sampling.

4.4.3 Model training for prediction of soil salinity

The measured values of EC_e (target or response variable) and the values of each of the 14 predictors (each represented by one column of data), attributed to the measured values of EC_e, were then imported to MATLAB for model training and validation. For each GCM, a separate matrix of data was prepared, with a total of 16 matrices. The WRB soil classes (as the only categorical predictor) were represented by a vector of positive integers that contained values assigned to different soil classes. The other 13 predictors were non-categorical represented by a set of real numbers. In spite of employing the method explained earlier for estimation of the missing cells in predictors' layers, the values of some purely spatial predictors were still missing in the final imported matrices. Therefore, the corresponding EC_e values (each represented by a row of data) were eliminated and not used for model training. As a result, 1.28% of the sample rows were excluded from the analysis.

We applied MATLAB Statistics and ML toolbox (MATLAB, R2019b) for building and validating the predictive models of EC_e. Here, we used an ensemble of regression trees for training and projecting the soil salinity based on the predictor datasets obtained from each of the 16 GCMs shown in Table 4-3. We chose tree-based models due to their relatively higher accuracy and computational speed compared to other ML algorithms (Breiman, 2001; Hassani et al., 2020). Additionally, tree-based predictive models are highly flexible in mapping non-linear relations between the known predictors and known responses (Hengl et al., 2015; Kuhn & Johnson, 2013) and are robust in handling outliers and collinearity concerns in environmental modelling (De'Ath, 2007; Elith & Leathwick, 2017). The MATLAB built-in "fitrenemble" function was applied for training the regression ensembles.

The model hyperparameters, or parameters that should be set before launching the training process of a ML algorithm, were tuned using MATLAB automatic hyperparameter optimiser. These comprised ensemble aggregation method, number of learning cycles, learn rate, minimum leaf size, maximum number of splits, and number of variables to sample (Breiman, 2001). By varying the hyperparameters, the optimiser attempts to find a combination of their values which minimises the *log (1 + cross-validation loss)*. Holdout cross-validation method (with 25% of data being held out) was used for optimisation and the cross-validation loss was quantified using mean squared error. The optimiser used the Bayesian optimisation algorithm with the "expected-improvement-per-second-plus" acquisition function. The maximum number of objective function evaluations was 100 since there was no notable decrease in the value of the observed minimum objective function after 100 evaluations. We repartitioned the cross-validation at every iteration and assumed the weight of all observation rows to be equal to one. We applied the log-transform to address the issue of right skewness in frequency distribution of the target variable; however, the log-transformation and back-transform of the predicted responses had a negligible impact on the accuracy of the trained modes.

Table 4-3: Global Circulation models (GCMs) used for calculation of the spatio-temporal predictors.

Model name	Ensemble member(s)[a]	Scenario(s)	Spatial resolution (latitude × longitude)	Source
Coupled Model Inter-comparison Project Phase 5 (CMIP5)				
GISS-E2-H	r6i1p3	RCP 4.5	2° × 2.5°	NASA Goddard Institute for Space Studies (Schmidt et al., 2006).
GISS-E2-R	r6i1p3	RCP 4.5	2° × 2.5°	NASA Goddard Institute for Space Studies (Schmidt et al., 2006).
MIROC5	r1i1p1, r2i1p1, r3i1p1	RCP 4.5, RCP 8.5	1.4008° × 1.40625°	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (M. Watanabe et al., 2010).
MIROC-ESM-CHEM	r1i1p1	RCP 4.5, RCP 8.5	2.7906° × 2.8125°	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies (S. Watanabe et al., 2011).
MIROC-ESM	r1i1p1	RCP 4.5, RCP 8.5	2.7906° × 2.8125°	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies (S. Watanabe et al., 2011).
MRI-CGCM3	r1i1p1	RCP 4.5, RCP 8.5	1.12148° × 1.125°	Meteorological Research Institute (Yukimoto et al., 2012)
NorESM1-M	r1i1p1	RCP 4.5, RCP 8.5	1.8947° × 2.5°	Norwegian Climate Centre (Bentsen et al., 2013).
MRI-ESM1	r1i1p1	RCP 8.5	1.8947° × 2.5°	Meteorological Research Institute (Yukimoto, 2011)
Coupled Model Inter-comparison Project Phase 6 (CMIP6)				
CESM2-WACCM-gn	r1i1p1f1, r2i1p1f1, r3i1p1f1	SSP 2-4.5, SSP 5-8.5	0.94240838° × 1.25°	Community Earth System Model Contributors (Danabasoglu et al., 2020).
CNRM-ESM2-1-gr	r1i1p1f2	SSP 2-4.5, SSP 5-8.5	1.4003477° × 1.40625°	National Centre for Meteorological Research, Météo-France and CNRS laboratory (Séférian et al., 2019).
GFDL-ESM4-gr1	r1i1p1f1	SSP 2-4.5, SSP 5-8.5	1° × 1.25°	NOAA Geophysical Fluid Dynamics Laboratory (Dunne et al., 2019).
INM-CM4-8-gr1	r1i1p1f1	SSP 2-4.5, SSP 5-8.5	1.5° × 2°	Institute for Numerical Mathematics (Volodin et al., 2019a).
INM-CM5-0-gr1	r1i1p1f1	SSP 2-4.5, SSP 5-8.5	1.5° × 2°	Institute for Numerical Mathematics (Volodin et al., 2019b).
MIROC-ES2L-gn	r1i1p1f2	SSP 2-4.5, SSP 5-8.5	2.7889823° × 2.8125°	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies (Hajima et al., 2019).
MRI-ESM2-0-gn	r1i1p1f1	SSP 2-4.5, SSP 5-8.5	1.8645104° × 1.875°	Meteorological Research Institute (Yukimoto et al., 2019).
NorESM2-LM-gn	r1i1p1f1	SSP 2-4.5, SSP 5-8.5	1.8947368° × 2.5°	Norwegian Climate Centre (Seland et al., 2020).

[a] Four indices defining an ensemble member: “r” for realization, “i” for initialization, “p” for physics, and “f” for forcing.

The Bayesian optimisation algorithm could return different results since its chosen acquisition function depends on the runtime of the objective function; the optimiser avoids the regions with extremely high runtimes. According to the non-reproducibility of the tuned set of hyperparameters, the model training and hyperparameter tuning jobs on each of 16 datasets were repeated 30 times (480 models in total). The maximum number of learning cycles was limited to 500 to keep the runtime for each training task below 10 minutes. High runtime and computational costs did not allow us to repeat the trainings more than 30 times. We accelerated the model training process by running the computations on a machine with 48 cores using the MATLAB Parallel Computing Toolbox. The goodness-of-fit of the trained models was evaluated by 10-fold cross-validation R^2 (the extent of variation explained by the model (Moriasi et al., 2015)), root mean squared error (*RMSE*), mean absolute error (*MAE*), and Nash-Sutcliffe model efficiency coefficient (*NSE* (Nash & Sutcliffe, 1970)). Then we used the bias corrected and accelerated percentile method to calculate the 95% confidence intervals of the mean for each validation metric based on 1,000 bootstrap samples (with replacement) derived from the results of the 30 runs performed for each of the 16 datasets. Among the 30 trained models for each input training set, the one with the lowest *RMSE* was selected; we chose *RMSE* as it is more sensitive to large errors (Hyndman & Koehler, 2006). In total, 16 models remained in our analysis for soil salinity projections.

4.4.4 Model implementation and soil salinity projection

We converted the world drylands layer delineated by the United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC, 2007) to a raster layer at 0.5° spatial resolution for generation of a global soil base map of the drylands. From that layer, we constrained our analysis to areas with an AI ≤ 0.65 and masked out the grid-cells (pixels) with an AI > 0.65 to keep only the drylands in our analysis (UNEP-WCMC, 2007). The remained raster had 24,045 grid-cells and we used it as the global soil base map of the drylands.

Similar to input training profiles data, we extracted the values of purely spatial and spatio-temporal predictors to the location of the base map grid-cells and then a five year moving average from the values of spatio-temporal predictors was computed. We applied the best chosen trained models to these new locations (cells) and the corresponding values of the predictors. As mentioned before, the degree of soil salinity and solute concentration change along the soil depth. Usage of the upper and lower depths of the samples as predictors in the model training enabled us to make predictions of EC_e at different depths below the soil surface.

In this regard, the trained models can be considered as four-dimensional predictive models of soil salinity that make predictions for different longitudes, latitudes, depths, and times. For each pixel and each year, we predicted the values of soil salinity at five depths: 0, 10, 30, 60, and 100 cm. We used the trapezoidal rule to compute an average of the EC_e (dS m^{-1}) to the depth of 1 meter as follows (Hengl et al., 2017):

$$EC_{e,\,ave} = [(10 - 0) \times (EC_e\,(10) + EC_e\,(0)) + (30 - 10) \times (EC_e\,(30) + EC_e\,(10)) + \ldots$$
$$(60 - 30) \times (EC_e\,(60) + EC_e\,(30)) + (100 - 60) \times (EC_e\,(100) + EC_e\,(60))] \,/\, (100 \times 2)$$

where EC_e is the predicted salinity at the corresponding depth. The outlier that is more than three scaled Median Absolute Deviations (MAD) away from the median of all predictions of a year were removed by the MATLAB “isoutlier” built-in function; this was the most robust method for removing outliers according to the user guide (see MATLAB “isoutlier” documentation for further details). In total, for each grid-cell of the global soil base map of the drylands, 197 predictions of EC_e were made in the period between 1904 and 2100 (one prediction for each year); since all spatio-temporal predictors are five-year moving averages, 1904 is the beginning of the period.

To compare the future state of the drylands soil salinity against the past conditions, we considered three time periods in our analysis: reference period (1961 - 1990), mid-term future (2031 - 2060), and long-term future (2071 - 2100). We used 30-year periods and 1961 - 1990 as the reference period based on the recommendations of the World Meteorological Organization for evaluations of the long-term changes in climatic variables (World Meteorological Organization, 2017). Soil salinity predictions for years in the future periods were averaged and compered to the average of the predictions for years in the reference period.

We calculated the area of each grid-cell of the global soil base map of the drylands in the WGS 1984 spatial coordinates using the computer code presented in the Supplementary Information. We estimated the total annual area of salt-affected soils between 1904 and 2100 and then computed the annual percentage change in the area of those soils by dividing the total area at each year by the average area of salt-affected soils over the period. We assumed an average of 95 years would be enough to remove the potential noise introduced by the spatio-temporal predictors. We used global administrative areas dataset (GADM, 2020) to estimate the total area of salt-affected soils at the national and continental levels. Numerical values representing the countries and continents were attributed to each cell of the base soil map.

4.4.5 Accuracy assessment of the trained models

The results of hyperparameter tuning and the 10-fold cross-validation accuracy metrics of the best fitted models are summarised in Appendix 3, Table A3-14. Appendix 3, Table A3-15 also presents the calculated lower and upper limits of 95% confidence intervals of the 10-fold cross-validation accuracy metrics, calculated for the trained models. For all 16 models, the MATLAB ensemble aggregation method of "LSBoost" was superior in fitting the models, compared to the "Bagged" method.

For the best fitted models, the lowest R^2 was 71.72% (with the 95% confidence intervals of 67.62% - 69.89%) related to the GISS-E2-R model, while the highest R^2 between the measured and predicted values of EC_e was 73.95% (67.34% - 70.32%), calculated for the CNRM-ESM2-1 model (see Table 4-3 for the details of GCMs). For all 16 models, the average calculated 10-fold cross-validation R^2 was 72.79%. Likewise, GISS-E2-R and CNRM-ESM2-1 were the models with the highest and lowest calculated values of *RMSE*, respectively. The average of 10-fold cross-validation *RMSE* for all 16 best fitted models was 3.6, ranging from 3.52 (3.78 - 3.93) to 3.67 (3.76 - 3.95). This represents a normalised *RMSE* equal to ~6% (normalised to the observed range of the EC_e values).

To understand better how well the best fitted models predict the response values, the relation between the measured (values sampled from the soil profiles) and predicted values of EC_e is visualised in Figure 4-5 via bin scatter plots. Taking a conservative approach, Figure 4-5 shows only the validation plots for the six (out of the 16 best-fitted) models with the worst performance (i.e. with highest *RMSE* values). The models predictions are fairly concentrated around the $y = x$ line, suggesting a good agreement of the modelled values with measured data. The accuracy of predictions increases with EC_e values, with a tendency for over-estimations for $EC_e \leq 1$ dS m^{-1}. Overall, the relatively high R^2 (>70%) values indicate a satisfactory model fitting, particularly as such values are not common in digital soil mapping (Malone et al., 2009).

Additionally, we evaluated the accuracy of the vertical prediction of the 16 best fitted models, i.e. the prediction accuracy at various depths from the soil surface. To do so, we categorised the measured and predicted (by 10-fold cross-validation) values of EC_e into six bins of 0 - 20 cm, 20 - 40 cm, 40 - 60 cm, 60 - 80 cm, 80 - 100 cm, and 100 - 200 cm based on an average from the lower and upper depths of the samples (each bin included its left edge); the bins edges were chosen so that the number of samples available for each bin stayed roughly equal and the deeper depths were not considered due to lack of data. The calculated R^2 values for each bin and each of the 16 models are reported in Appendix 3, Table A3-16. The averages of the 16 models R^2 values for the shallowest to deepest soil layers (bins) were 63.59%, 72.99% 77.39%, 77.31%, 79.59%, and 72.51%, respectively. These accuracies are in line with the reported R^2 values of Taghizadeh-Mehrjardi et al. (2014) who developed separate regression tree-based models to predict soil salinity (78% for 0 - 15 cm soil layer). However, their analysis was purely spatial and was only focused on the saline soils located in a local area in central Iran (72,000 ha), while the current analysis projects the spatio-temporal variability in soil salinity on the global scale. We did not observe a decrease in predictive accuracy of the digital soil models at the higher depths reported in other studies, such as Malone et al. (2009), Minasny et al. (2006), and Minasny et al. (2006).

In addition to global accuracy assessment of the trained models, we evaluated the predictive power of the best fitted models at the country and continental levels (Figure 4-6, a and b). We grouped the measured sample values of EC_e according to the continent or the country where the samples were acquired and compared the mean of each group with the mean of the 10-fold cross-validated predictions for each group. Only 87 countries had measured input profiles data of EC_e required for our analysis. At the country level, the R^2 between the mean of predictions of the 16 models and the mean of measured values of EC_e was 80.41% while at the continental level, this value was 99.64%. The reason for such a high accuracy at the continental level is the high number of data points within each continent which makes the predicted and estimated averages close to each other.

Similarly, we compared the predictions of our models with other available gridded datasets on soil EC_e, including HWSD (Harmonised World Soil Database (Fao/Iiasa/Isric/Isscas/Jrc, 2012)) and WISE (World Inventory of Soil Emission Potentials) which derived soil properties on a 30 × 30 arc-seconds global grid (WISE-30; (Batjes, 2015)), at the country and continental levels. Since these two datasets provide data for different soil layers (HWSD: two layers at 0 - 30 cm and 30 - 70 cm; WISE-30: seven layers, with five fixed depth intervals of 20 cm up to the depth of 100 cm and two 50 cm depth intervals between 100 and 200 cm), we only focused on surface measurements. For comparison with HWSD, any soil sample with the upper sample depth of 0 cm and a lower sample depth ≤ 30 was chosen as the surface measurement (a total of 8,995 samples) while for WISE-30, any EC_e sample with the lower sample depth of 20 cm was chosen as the surface measurement (a total of 7,535 samples).

At the location of each particular surface measurement, we predicted the soil salinity for 0 - 20 or 0 - 30 cm (depending on the target dataset for comparison) soil layers using the purely spatial and spatio-temporal values of predictors corresponding to the year of sampling of that particular surface measurement. Then we grouped the predictions and surface measurements based on the country and continent of sampling. At the country level, the R^2 between the mean of our models predictions and the mean of surface measured values (0 - 30 cm) of EC_e for 74 countries was 68.55%, while this value for HWSD was 13.6%. At the continental level, these values were 91.48% and 74.98%, respectively (Figure 4-6, c and d). Compared to the WISE-30 predictions, the R^2 between the mean of our models predictions and the mean of surface measured values (0 - 20 cm) of EC_e was 69.33% and 87.99% at the country (71 countries) and continental levels, respectively whereas the WISE-30 values were 17.22% and 5.53% (Figure 4-6, e and f). Although HWSD and WISE-30 datasets are purely spatial and they do not include information on the temporal variability of the soil salinity, comparing the predictions made by the models developed here against the predictions of those datasets can provide a better quantitative understanding of the improved predictive performance of our models.

4.4.6 Model limitations, uncertainties, and perspectives for future research

ML models are one of the solutions suggested for time series projection challenges (Ye et al., 2019). However, unlike the analytical models, ML models do not enable consideration of the mechanistic insights in the predictive algorithms of soil properties (Hengl et al., 2017). As mentioned earlier, no harmonised dataset is currently available quantifying the concentration of the soluble salts in salt-affected soils and, to a great extent, quantification of the severity of soil salinity in the field is limited to EC_e measurements. Provision of such dataset can be a baseline for developing more mechanistic approaches in projections of soil salinity. Although it would be very challenging, projecting large-scale soil salinity driven by groundwater table, irrigation practices, and sea level rise are important areas for future research.

Captured trends and projections in this study depend on the input data used for training the models. Inconsistency in accuracy and methods applied by different laboratories for measuring soil properties can negatively impact the trends captured by the trained models. As we go towards the past, the number of available samples and their accuracy decreases (Figure 4-4, b); this in turn may influence the validation procedures applied to the predictions made by ML models (Hengl et al., 2017). It may also generate predictions biased towards the recent periods when more data samples are available. Additionally, more care should be given to application of the predictions made here at locations underrepresented by input data for training the ML models. In the current study, the majority of soil profiles used for training were sampled from North America and Australia due to a greater data availability. Thus, there is a possibility that the results are biased towards the soil and hydro-climatic conditions of these two continents. One solution to address this issue can be to develop more regional ML models; yet, this is challenging in the locations with the low number of sample data. Decrease in the number of available input data reduces the efficiency of the model training, resulting in less accurate and unsatisfactory validation outcomes. More updated and geographically scattered profile data are required in future studies to address the issue of inconsistency in the legacy soil-profile data.

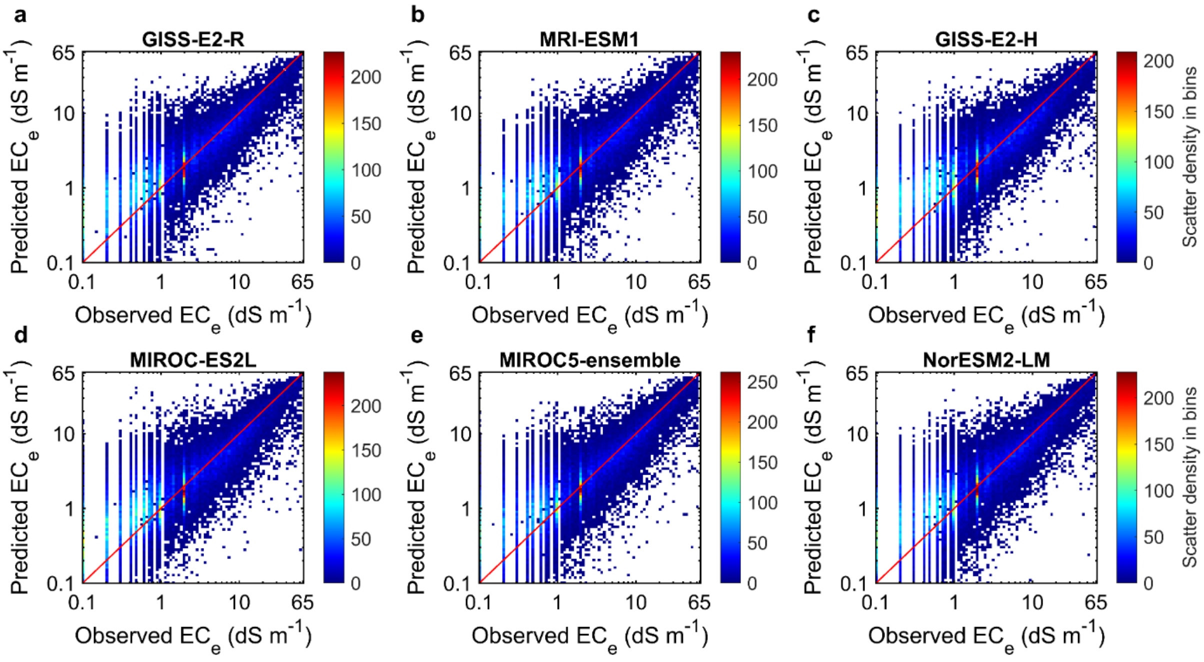


Figure 4-5: 10-fold cross-validation plots for the six trained models with the highest root mean squared error (*RMSE*) values out of the final 16 best fitted models. The *RMSE* decreases from **a** to **f.** The colour maps show the scatter density in each bin. The red lines represent the y = x line.

More importantly, the extent of uncertainty in the predictors used for training the models is not spatially constant. All the predictors used here are large-scale estimations of other models which inherently include some degrees of uncertainty. Particularly, purely spatial predictors including the wilting point, field capacity, and effective plant rooting depth, are less certain in large deserts where observations are scarce for tuning and validation of the models. One way to address this issue is to provide spatially explicit maps of uncertainty for the predictions of the ML algorithms. However, this needs spatially explicit uncertainty maps of the predictors or their probability distributions. In the case of our study, such data were not available for the predictors. Additionally, ML algorithms are highly computationally demanding and estimation of the outputs uncertainty ranges by methods such as Monte Carlo simulations was not feasible by our computational resources (assuming hypothetical distributions of uncertainty in the predictors and input profiles data). Thus, we did not quantify the posterior distribution and uncertainty of the predictions and instead we estimated the global accuracy of the projected results via the 10-fold cross-validation method. A less computationally intensive framework is needed in the future for provision of the spatially explicit estimations of uncertainties in outputs of the ML models.

The number of GCMs with projected wet and dry sea salt deposition rates (which are also necessary for mechanistic approaches) were rather limited in both CMIP5 and CMIP6 data projects. More ensemble members could improve the certainty of the projected soil salinity. Furthermore, the spatial resolution of our salinity projections was relatively coarse (0.5°); although the purely spatial predictors were of the adequate resolution, there was no point in prediction of the soil salinity values at finer resolutions since the spatio-temporal resolution of the GCMs grids was roughly between 1 and 3°. Such issues might be addressed with improvement of the spatial resolution of GCMs and the number of GCMs with sea salt aerosols projections in upcoming years.

Data availability. Input data for training the predictive models, objects of the predictive models, annual predictions made by the models for each location, and spatially-explicit maps quantifying the change in predicted soil salinity in the mid- (2031 - 2060) and long-term futures (2071 - 2100), relative to the reference period (1961 - 1990) are freely available at: www.figshare.com.

Code availability

Computer codes required for regeneration of the main results presented in this paper can be found in Supplementary Information appendix (computer codes section).

Acknowledgements

The first author's PhD study was supported by Presidential Doctoral Scholarship Award at The University of Manchester, which is gratefully acknowledged.

Author contributions. The contribution of all authors was equal and essential to preparation of the material presented here.

Competing Interests. The authors declare no competing interests.

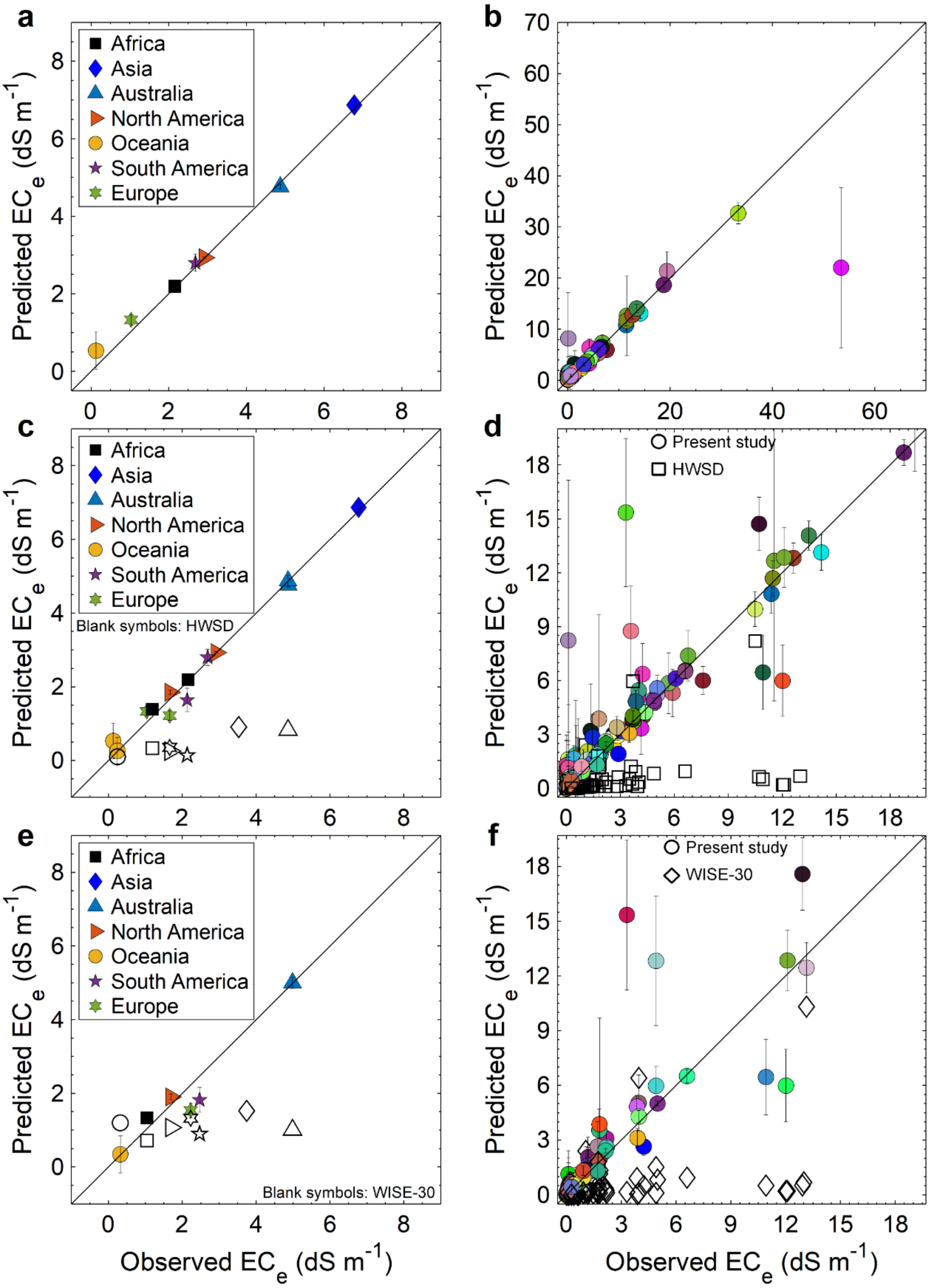


Figure 4-6: Comparison of the predicted values of soil salinity (EC$_e$) in the present study and the measured values as well as the soil EC$_e$ predicted in other datasets (i.e. HWSD and WISE-30) at the continental (a, c, e) and country levels (b, d, f). a and **b**, average predicted values versus average measured values at the continental and country levels (87 countries), respectively. **c** and **d**, average of the surface (0 - 30 cm) salinity (EC$_e$) values predicted by the present study and Harmonised World Soil Database (HWSD) versus the average of measured surface salinity at the continental and country levels (74 countries), respectively. **e** and **f**, average of the surface (0 - 20 cm) salinity predicted by the present study and WISE-30 (World Inventory of Soil Emission Potentials derived soil properties) dataset versus the average of measured surface salinity at the continental and country levels (71 countries), respectively. The error bars represent the minimum and maximum of average values calculated for the 29 models used in the study.

References

Abrol, I., Yadav, J. S. P., & Massoud, F. (1988). *Salt-affected soils and their management*: Food & Agriculture Org.

Alexander, L. V., & Arblaster, J. M. (2017). Historical and projected trends in temperature and precipitation extremes in Australia in observations and CMIP5. *Weather and Climate Extremes, 15*, 34-56.

Amini, S., Ghadiri, H., Chen, C., & Marschner, P. (2016). Salt-affected soils, reclamation, carbon dynamics, and biochar: a review. *Journal of soils and sediments, 16*(3), 939-953.

Ayub, M. A., Ahmad, H. R., Ali, M., Rizwan, M., Ali, S., ur Rehman, M. Z., & Waris, A. A. (2020). Salinity and its tolerance strategies in plants. In *Plant Life Under Changing Environment* (pp. 47-76): Elsevier.

Bates, B. (2009). *Climate change and water: IPCC technical paper VI*: World Health Organization.

Batjes, N. H. (2015). *World soil property estimates for broad-scale modelling (WISE30sec)*. Retrieved from

Batjes, N. H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., & de Jesus, J. M. (2017). WoSIS: providing standardised soil profile data for the world. *Earth System Science Data, 9*(1), 1.

Belward, A. S., Estes, J. E., & Kline, K. D. (1999). The IGBP-DIS global 1-km land-cover data set DISCover: A project overview. *Photogrammetric Engineering and Remote Sensing, 65*(9), 1013-1020.

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., et al. (2013). The Norwegian earth system model, NorESM1-M—Part 1: Description and basic evaluation of the physical climate. *Geosci. Model Dev, 6*(3), 687-720.

Bleam, W. F. (2016). *Soil and environmental chemistry*: Academic Press.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Burt, R. (2011). Soil Survey Investigations Report, no. 45, version 2.0. *Natural Resources Conservation Service*.

Chen, J., & Mueller, V. (2018). Coastal climate change, soil salinity and human migration in Bangladesh. *Nature Climate Change, 8*(11), 981-985.

Cinquini, L., Crichton, D., Mattmann, C., Harney, J., Shipman, G., Wang, F., et al. (2014). The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data. *Future Generation Computer Systems, 36*, 400-417.

Colombani, N., Mastrocicco, M., & Giambastiani, B. M. S. (2015). Predicting salinization trends in a lowland coastal aquifer: Comacchio (Italy). *Water Resources Management, 29*(2), 603-618.

Colorado-Ruiz, G., Cavazos, T., Salinas, J. A., De Grau, P., & Ayala, R. (2018). Climate change projections from Coupled Model Intercomparison Project phase 5 multi-model weighted ensembles for Mexico, the North American monsoon, and the mid-summer drought region. *International journal of climatology, 38*(15), 5699-5716.

Corwin, D. (2020). Climate change impacts on soil salinity in agricultural areas. *European Journal of Soil Science*, 1-21.

Corwin, D. L., Rhoades, J. D., & Šimůnek, J. (2007). Leaching requirement for soil salinity control: Steady-state versus transient models. *Agricultural Water Management, 90*(3), 165-180.

Daliakopoulos, I., Tsanis, I., Koutroulis, A., Kourgialas, N., Varouchakis, A., Karatzas, G., & Ritsema, C. (2016). The threat of soil salinity: A European scale review. *Science of the Total Environment, 573*, 727-739.

Daliakopoulos, I. N., Pappa, P., Grillakis, M. G., Varouchakis, E. A., & Tsanis, I. K. (2016). Modeling soil salinity in greenhouse cultivations under a changing climate with SALTMED: model modification and application in Timpaki, Crete. *Soil Science, 181*(6), 241-251.

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., et al. (2020). The Community Earth System Model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems, 12*(2), e2019MS001916.

Dasgupta, S., Hossain, M. M., Huq, M., & Wheeler, D. (2015). Climate change and soil salinity: The case of coastal Bangladesh. *Ambio, 44*(8), 815-826.

De'Ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology, 88*(1), 243-251.

De la Paix, M., Lanhai, L., Xi, C., Varenyam, A., Nyongesah, M., & Habiyaremye, G. (2013). Physicochemical properties of saline soils and aeolian dust. *Land Degradation & Development, 24*(6), 539-547.

Desktop, E. A. (2011). Release 10. *Redlands, CA: Environmental Systems Research Institute, 437*, 438.

Dunne, J., Horowitz, L., Held, I., Krasting, J., John, J., Malyshev, S., et al. (2019). The GFDL Earth System Model version 4.1 (GFDL-ESM4. 1): Model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems, 11*, 3167-3211.

Easterling, D. R., Kunkel, K., Arnold, J., Knutson, T., LeGrande, A., Leung, L. R., et al. (2017). Precipitation change in the United States.

Elith, J., & Leathwick, J. (2017). Boosted Regression Trees for ecological modeling. *R Documentation. Available online: https://cran. r-project. org/web/packages/dismo/vignettes/brt. pdf (accessed on 12 June 2011).*

Esri. (2020). World Elevation Terrain data, ArcGIS Living Atlas of the World *avaialble at www.arcgis.com/home/item.html?id=58a541efc59545e6b7137f961d7de883.*

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development, 9*(5), 1937-1958.

FAO ITPS. (2015). Status of the world's soil resources (SWSR)–main report. *Food and agriculture organization of the United Nations and intergovernmental technical panel on soils, Rome, Italy, 650.*

Fao/Iiasa/Isric/Isscas/Jrc. (2012). Harmonized world soil database (version 1.2). *FAO, Rome, Italy and IIASA, Laxenburg, Austria.*

Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuizen, H., Verelst, L., & Wiberg, D. (2008). Global agro-ecological zones assessment for agriculture (GAEZ 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy, 10.*

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., & Van Der Velde, M. (2016). Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nature communications, 7*(1), 1-13.

GADM. (2020). Database of Global Administrative Areas. *https://gadm.org/.*

Ghassemi, F., Jakeman, A. J., & Nix, H. A. (1995). *Salinisation of land and water resources: human causes, extent, management and case studies*: CAB international.

Giorgi, F., Raffaele, F., & Coppola, E. (2019). The response of precipitation characteristics to global warming from climate projections. *Earth System Dynamics, 10*(1), 73-89.

Global Soil Data Task Group. (2000). Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS). Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme - Data and Information System). *Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/569.*

Gorji, T., Sertel, E., & Tanik, A. (2017). Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecological Indicators, 74*, 384-391.

Greenway, H., & Munns, R. (1980). Mechanisms of salt tolerance in nonhalophytes. *Annual review of plant physiology, 31*(1), 149-190.

Grose, M. R., Narsey, S., Delage, F., Dowdy, A. J., Bador, M., Boschat, G., et al. (2020). Insights from CMIP6 for Australia's future climate. *Earth's Future, 8*(5), e2019EF001469.

Guglielmini, L., Gontcharov, A., Aldykiewicz Jr, A. J., & Stone, H. A. (2008). Drying of salt solutions in porous materials: Intermediate-time dynamics and efflorescence. *Physics of fluids, 20*(7), 077101.

Gupta, R. K., Abrol, I. P., Finkl, C. W., Kirkham, M. B., Arbestain, M. C., Macías, F., et al. (2008). Solonchaks. In W. Chesworth (Ed.), *Encyclopedia of Soil Science* (pp. 737-738). Dordrecht: Springer Netherlands.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M., Abe, M., et al. (2019). Description of the MIROC-ES2L Earth system model and evaluation of its climate—Biogeochemical processes and feedbacks. *Geoscientific Model Development Discussions, 2019*, 1-73.

Hassani, A., Azapagic, A., D'Odorico, P., Keshmiri, A., & Shokri, N. (2020). Desiccation crisis of saline lakes: A new decision-support framework for building resilience to climate change. *Science of the Total Environment, 703*, 134718.

Hassani, A., Azapagic, A., & Shokri, N. (2020). Predicting long-term dynamics of soil salinity and sodicity on a global scale. *Proceedings of the National Academy of Sciences, 117*(52), 33017-33027.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one, 12*(2).

Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., et al. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS one, 10*(6), e0125814.

Hillel, D. (2000). *Salinity management for sustainable irrigation: integrating science, environment, and economics*: The World Bank.

Huinink, H., Pel, L., & Michels, M. v. A. (2002). How ions distribute in a drying porous medium: A simple model. *Physics of fluids, 14*(4), 1389-1395.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting, 22*(4), 679-688.

Intergovernmental Panel on Climate Chang. (1996). Climate Change 1995: Impacts, Adaptations and Mitigation of Climate Change: Scientific Technical Analyses: Special Report of Working Group II. *Cambridge Univ. Press, New York.*

IUSS Working Group WRB. (2015). World reference base for soil resources 2014, update 2015: International soil classification system for naming soils and creating legends for soil maps. *FAO, Rome.*

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., & De Sousa, L. (2019). Global mapping of soil salinity change. *Remote Sensing of Environment, 231*, 111260.

Jenny, H. (1994). *Factors of soil formation: a system of quantitative pedology*: Courier Corporation.

Karmakar, R., Das, I., Dutta, D., & Rakshit, A. (2016). Potential effects of climate change on soil properties: a review. *Science international, 4*(2), 51-73.

Klutse, N. A. B., Ajayi, V. O., Gbobaniyi, E. O., Egbebiyi, T. S., Kouadio, K., Nkrumah, F., et al. (2018). Potential impact of 1.5 C and 2 C global warming on consecutive dry and wet days over West Africa. *Environmental Research Letters, 13*(5), 055013.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26): Springer.

Li, X., Chang, S. X., & Salifu, K. F. (2014). Soil texture and layering effects on water and salt dynamics in the presence of a water table: a review. *Environmental reviews, 22*(1), 41-50.

Ma, X., Zhao, C., Tao, H., Zhu, J., & Kundzewicz, Z. W. (2018). Projections of actual evapotranspiration under the 1.5 C and 2.0 C global warming scenarios in sandy areas in northern China. *Science of the Total Environment, 645*, 1496-1508.

Maas, E. V., & Grattan, S. (1999). Crop yields as affected by salinity. *Agricultural drainage, 38*, 55-108.

Malone, B. P., McBratney, A., Minasny, B., & Laslett, G. (2009). Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma, 154*(1-2), 138-152.

Martín-Rosales, W., Pulido-Bosch, A., Vallejos, Á., Gisbert, J., Andreu, J. M., & Sánchez-Martos, F. (2007). Hydrological implications of desertification in southeastern Spain/Implications hydrologiques de la désertification dans le sud-est de l'Espagne. *Hydrological Sciences Journal/Journal des Sciences Hydrologiques, 52*(6), 1146-1161.

Mau, Y., & Porporato, A. (2015). A dynamical system approach to soil salinity and sodicity. *Advances in water resources, 83*, 68-76.

Miao, L., Li, S., Zhang, F., Chen, T., Shan, Y., & Zhang, Y. (2020). Future Drought in the Dry Lands of Asia Under the 1.5 and 2.0° C Warming Scenarios. *Earth's Future, 8*(6), e2019EF001337.

Middleton, N., & Thomas, D. (1997). *World atlas of desertification.. ed. 2*: Arnold, Hodder Headline, PLC.

Minasny, B., McBratney, A. B., Mendonça-Santos, M., Odeh, I., & Guyon, B. (2006). Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Soil Research, 44*(3), 233-244.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE, 58*(6), 1763-1785.

Mulder, V., De Bruin, S., Schaepman, M. E., & Mayr, T. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma, 162*(1-2), 1-19.

Munns, R., & Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol., 59*, 651-681.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology, 10*(3), 282-290.

National Land and Water Resources Audit. (2001). Australian Dryland Salinity Assessment 2000: extent, impacts, processes, monitoring and management options. *Commonwealth of Australia, Canberra*.

Oldeman, L. R., Hakkeling, R., & Sombroek, W. G. (2017). *World map of the status of human-induced soil degradation: an explanatory note*: International Soil Reference and Information Centre.

Omuto, C., Nachtergaele, F., & Rojas, R. V. (2013). *State of the Art Report on Global and regional Soil Information: Where are we? Where to go?* : Food and Agriculture Organization of the United Nations Rome.

Oude Essink, G., Van Baaren, E. S., & De Louw, P. G. (2010). Effects of climate change on coastal groundwater systems: A modeling study in the Netherlands. *Water resources research, 46*(10).

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. *Soil, 6*(1), 35-52.

Pannell, D. J. (2001). Dryland salinity: economic, scientific, social and policy dimensions. *Australian Journal of Agricultural and Resource Economics, 45*(4), 517-546.

Pannell, D. J., & Ewing, M. A. (2006). Managing secondary dryland salinity: options and challenges. *Agricultural water management, 80*(1-3), 41-56.

Parihar, P., Singh, S., Singh, R., Singh, V. P., & Prasad, S. M. (2015). Effect of salinity stress on plants and its tolerance strategies: a review. *Environmental Science and Pollution Research, 22*(6), 4056-4075.

Paz, A. M., Castanheira, N., Farzamian, M., Paz, M. C., Gonçalves, M. C., Santos, F. A. M., & Triantafilis, J. (2020). Prediction of soil salinity and sodicity using electromagnetic conductivity imaging. *Geoderma, 361*, 114086.

Perri, S., Suweis, S., Entekhabi, D., & Molini, A. (2018). Vegetation controls on dryland salinity. *Geophysical Research Letters, 45*(21), 11,669-611,682.

Perri, S., Suweis, S., Holmes, A., Marpu, P. R., Entekhabi, D., & Molini, A. (2020). River basin salinization as a form of aridity. *Proceedings of the National Academy of Sciences, 117*(30), 17635-17642.

Porporato, A., Feng, X., Manzoni, S., Mau, Y., Parolari, A. J., & Vico, G. (2015). Ecohydrological modeling in agroecosystems: Examples and challenges. *Water resources research, 51*(7), 5081-5099.

Prăvălie, R. (2016). Drylands extent and environmental issues. A global approach. *Earth-Science Reviews, 161*, 259-278.

Rad, M. N., Shokri, N., Keshmiri, A., & Withers, P. J. (2015). Effects of grain and pore size on salt precipitation during evaporation from porous media. *Transport in Porous Media, 110*(2), 281-294.

Rath, K. M., & Rousk, J. (2015). Salt effects on the soil microbial decomposer community and their role in organic carbon cycling: a review. *Soil Biology and Biochemistry, 81*, 108-123.

Rengasamy, P. (2006). World salinization with emphasis on Australia. *Journal of experimental botany, 57*(5), 1017-1023.

Richards, L. A. (1954). *Diagnosis and improvement of saline and alkali soils. Handbook No. 60*: US Department of Agriculture, Washington, DC.

Runyan, C. W., & D'Odorico, P. (2010). Ecohydrological feedbacks between salt accumulation and vegetation dynamics: Role of vegetation-groundwater interactions. *Water resources research, 46*(11).

Schimel, D. S. (2010). Drylands in the earth system. *Science, 327*(5964), 418-419.

Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., et al. (2006). Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *Journal of Climate, 19*(2), 153-192.

Schofield, R., & Kirkby, M. (2003). Application of salinization indicators and initial development of potential global soil salinization scenario under climatic change. *Global Biogeochemical Cycles, 17*(3).

Schoups, G., Hopmans, J., & Tanji, K. (2006). Evaluation of model complexity and space–time resolution on the prediction of long-term soil salinity dynamics, western San Joaquin Valley, California. *Hydrological Processes: An International Journal, 20*(13), 2647-2668.

Schulzweida, U. (2019). CDO user's guide *Climate data operators, Version 1.9.8.* http://doi.org/10.5281/zenodo.3539275.

Scudiero, E., Skaggs, T. H., & Corwin, D. L. (2014). Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Regional, 2*, 82-90.

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems, 11*(12), 4182-4227.

Seland, Ø., Bentsen, M., Seland Graff, L., Olivié, D., Toniazzo, T., Gjermundsen, A., et al. (2020). The Norwegian Earth System Model, NorESM2–Evaluation of theCMIP6 DECK and historical simulations. *Geoscientific Model Development Discussions*, 1-68.

Sentis, I. (1996). Soil salinization and land desertification. *Soil degradation and desertification in Mediterranean environments. Logroño, Spain, Geoforma Ediciones*, 105-129.

Shi, L., Feng, P., Wang, B., Li Liu, D., Cleverly, J., Fang, Q., & Yu, Q. (2020). Projecting potential evapotranspiration change and quantifying its uncertainty under future climate scenarios: A case study in southeastern Australia. *Journal of hydrology*, 124756.

Shokri-Kuehni, S. M., Raaijmakers, B., Kurz, T., Or, D., Helmig, R., & Shokri, N. (2020). Water Table Depth and Soil Salinization: From Pore-Scale Processes to Field-Scale Responses. *Water resources research, 56*(2), e2019WR026707.

Shokri-Kuehni, S. M., Vetter, T., Webb, C., & Shokri, N. (2017). New insights into saline water evaporation from porous media: Complex interaction between evaporation rates, precipitation, and surface temperature. *Geophysical Research Letters, 44*(11), 5504-5510.

Shokri, N. (2014). Pore-scale dynamics of salt transport and distribution in drying porous media. *Physics of fluids, 26*(1), 012106.

Singh, K. (2016). Microbial and enzyme activities of saline and sodic soils. *Land Degradation & Development, 27*(3), 706-718.

Soil Science Glossary Terms Committee. (2008). *Glossary of soil science terms 2008*: Soil Science Society of America, ASA-CSSA-SSSA.

Soil Survey Staff. (2010). *Keys to soil taxonomy*: United States Department of Agriculture, Soil Conservation Service: Washington, DC.

Sulla-Menashe, D., & Friedl, M. A. (2018). User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. *USGS: Reston, VA, USA*, 1-18.

Suweis, S., Rinaldo, A., Van der Zee, S., Daly, E., Maritan, A., & Porporato, A. (2010). Stochastic modeling of soil salinity. *Geophysical Research Letters, 37*(7).

Szabolcs. (1990). Impact of climatic change on soil attributes: influence on salinization and alkalinization. In *Developments in soil science* (Vol. 20, pp. 61-69): Elsevier.

Szabolcs, I. (1989). *Salt-affected soils*: CRC Press, Inc.

Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., & Malone, B. (2014). Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma, 213*, 15-28.

Talat, N. (2020). Alleviation of soil salinization and the management of saline soils, climate change, and soil interactions. In *Climate Change and Soil Interactions* (pp. 305-329): Elsevier.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society, 93*(4), 485-498.

Tomaz, A., Palma, P., Alvarenga, P., & Gonçalves, M. C. (2020). *Soil salinity risk in a climate change scenario and its effect on crop yield*: Elsevier.

UNEP-WCMC, L. (2007). A spatial analysis approach to the global delineation of dryland Areas of relevance to the CBD programme of work on dry and sub-humid lands, Dataset based on spatial analysis between WWF terrestrial ecoregions (WWF-US, 2004) and aridity zones (CRU/UEA; UNEPGRID, 1991). Dataset checked and refined to remove many gaps, overlaps and slivers (July 2014). *UK, Cambridge*.

Van Weert, F., Van der Gun, J., & Reckman, J. (2009). Global overview of saline groundwater occurrence and genesis. *International Groundwater Resources Assessment Centre*.

Várallyay, G. (1994). Climate change, soil salinity and alkalinity. In *Soil responses to climate change* (pp. 39-54): Springer.

Várallyay, G. (2010). The impact of climate change on soils and on their water management. *Agronomy Research, 8*(Special Issue II), 385-396.

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019a). INM-CM4-8 model output prepared for CMIP6 PMIP. https://doi.org/10.22033/ESGF/CMIP6.2295

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019b). INM INM-CM5-0 model output prepared for CMIP6 CMIP. https://doi.org/10.22033/ESGF/CMIP6.1423

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., et al. (2010). Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *Journal of Climate, 23*(23), 6312-6335.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development, 4*(4), 845.

Wong, V. N., Greene, R., Dalal, R. C., & Murphy, B. W. (2010). Soil carbon dynamics in saline and sodic soils: a review. *Soil use and management, 26*(1), 2-11.

World Meteorological Organization. (2017). WMO guidelines on the calculation of climate normals. *World Meteorological Organization*.

Yang, Y., Donohue, R. J., & McVicar, T. R. (2016). Global estimation of effective plant rooting depth: Implications for hydrological modeling. *Water resources research, 52*(10), 8260-8276.

Ye, L., Gao, L., Marcos-Martinez, R., Mallants, D., & Bryan, B. A. (2019). Projecting Australia's forest cover dynamics and exploring influential factors using deep learning. *Environmental Modelling & Software, 119*, 407-417.

Yeo, A. (1998). Predicting the interaction between the effects of salinity and climate change on crop plants. *Scientia Horticulturae, 78*(1-4), 159-174.

Yukimoto, S. (2011). *Meteorological research institute earth system model version 1 (MRI-ESM1): model description*: Meteorological Research Institute.

Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., et al. (2012). A new global climate model of the Meteorological Research Institute: MRI-

CGCM3—Model description and basic performance—. *Journal of the Meteorological Society of Japan. Ser. II, 90*, 23-64.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Ser. II.*

Zaman, M., Shahid, S. A., & Heng, L. (2018). *Guideline for salinity assessment, mitigation and adaptation using nuclear and related techniques*: Springer.

Zanchi, C., & Cecchi, S. (2010). *Soil salinisation in the Grosseto plain (Maremma, Italy): an environmental and socio-economic analysis of the impact on the agro-ecosystem.*: Springer.

Chapter 5 Conclusions and recommendations

The overall objective of this research has been to help decision-making in addressing the shrinkage of saline lakes and soil salinization issues as part of the two water and land management-related SDGs for water-scarce areas. The following specific objectives have been achieved:

- an eco-hydrological land-based framework is proposed as a solution for sustainable restoration of shrinking saline lakes, in view of major socio-economic, hydrologic, climatic, and agronomic affecting parameters in saline lakes' basins;
- a remote sensing data-based method has been developed which enables decision- and policy-makers to identify the parts of a saline lake with priority for restoration;
- the extent and severity of soil salinity/sodicity and soil salinization trends have been evaluated at the different geographical levels over the past four decades; and
- the soil salinization hotspots have been predicted by the end of the 21st century.

The following sections explain the particular findings of this work and the insights discussed in the previous chapters. This is followed by a discussion on the limitations of the work and recommendations for future research.

5.1 Sustainable solutions for restoration of saline lakes

The proposed framework presented in chapter 2 can be used for sustainable restoration of saline lakes, under different socio-economic and climatic scenarios. It can help decision- and policy-making to recognise the optimal land-based solutions for restoration of the desiccating saline lakes. The application of the framework for restoration of Lake Urmia revealed that:

- an increase in the Urmia basin air temperature close to 1.95˚C is projected by an ensemble of multi-GCMs by 2050 under RCP 4.5, compare to the base-line period (1960 - 1995); this increase is projected to be more extreme and close to 2.47 °C under RCP 8.5;
- overall, a decrease in precipitation is projected for the Urmia basin. Under RCP 4.5 and RCP 8.5 GHGs emission scenarios, the average annual precipitation in the Urmia basin is expected to decrease by 2.7% and 11.6%, in the 2030 - 2050 period relative to the base-line period (1960 - 1995), respectively.
- it is estimated that annually 4 cm salt has precipitated on the lake bed between 1997 and 2010;

- the lake's sub-surface gain has been 12.8% to 15.4% of the total inflow to the lake — nearly 2,861.3 Mm^3 yr^{-1};
- under RCP 4.5 and RCP 8.5 scenarios, the results of water balance over the lake's area show that annually 3,648 Mm^3 and 3,692 Mm^3 surface water inflow is required for the complete restoration of the lake (to the depth of 1274.1 m) by 2050. These environmental water requirements are inputs for the third step of the framework;
- to achieve the aims of inter-basin restoration, 78,700 ha and 114,826 ha reduction in the total irrigated area (438,900 ha) is required under RCP 4.5 and RCP 8.5 scenarios, respectively;
- in the case of intra-basin restoration, these respective reductions are 95,600 ha and 133,687 ha;
- assuming that the reductions in irrigated lands will be converted to rain-fed systems or grasslands and depending on the type, speed, and amount of conversion, it is estimated that the proposed land use change strategy leads to a reduction in cumulative GHGs emissions ranging between 0.07 (0.05-0.09) Mt CO_2 eq. and 1.55 (0.65-2.44) Mt CO_2 eq.
- it is estimated that there is a potential for the annual release of 36,391 t of saline dust (particulate matters smaller than 10 μm) originated from the lake bed if its moisture falls under 5%;
- if that soil moisture increases to 15%, the dust emissions would reduce to 3,342 t yr^{-1};
- restoring the northern areas of Lake Urmia (areas above the highway) can mitigate 22% of the annual dust emissions originated from the lake bed while the areas below the main southern islands can prevent 39% of the dust emissions;
- accordingly, restoration of the southern half of the lake is proposed as a short-term solution since it can prevent near to 77% of saline dust emissions originated from the Urmia lake bed; and
- for restoration of the southern part to the target level of 1274.1 m, annually 1.83 km^3 (under RCP 4.5) and 1.86 km^3 (under RCP 8.5) would be enough and the majority of these water requirements are currently provided by the southern supplying rivers.

5.2 Extent and severity of soil salinity/sodicity

The following conclusions can be drawn from the analysis of the generated dataset on the extent and severity of the soil salinity/sodicity:

- among the 43 environmental predictors, soil classes, depth from the surface, Fraction of Absorbed Photo-synthetically Active Radiation (FAPAR), and temperature of different soil layers were the most significant predictors for prediction of the soil salinity and sodicity;
- globally 11.73 Mkm^2 of soils have been salt-affected, i.e. their EC_e is ≥ 4 dS m^{-1} and/or their ESP is $\geq 6\%$ in at least three-fourths of the years between 1980 and 2018. Soil EC_e of an area close to 5.9 Mkm^2 have been ≥ 4 dS m^{-1} in at least three-fourths of the 1980 - 2018 period while soil ESP of an area close to 9.18 Mkm^2 have been $\geq 6\%$ in at least three-fourths of the years of that period;
- FAPAR is a better vegetation index compared to the Normalised Difference Vegetation Index (NDVI) for monitoring the salt-affected soils;
- between 1980 and 2018, it is estimated that 164,900 km^2 of the salt-affected soils are located in croplands;
- globally, the outputs of the developed models reveal that the possibility of observing salt-affected soils in the 2000 - 2018 period is lower than the 1981 - 1999 period;
- globally, the likelihood of reoccurrence of soils with $EC_e \geq 4$ dS m^{-1} in the 2000 to 2018 period was 0.94 of the 1981 to 1999 period; for the soils with ESP $\geq 6\%$, this likelihood was 0.97;
- between 1980 and 2018 and at the continental level, Asia (including the Middle East), Africa, Australia and Oceania, South America, North America, and Europe had the largest area of salt-affected soils, respectively;
- at the country level, China, Australia, Kazakhstan, and Iran had the largest area of salt-affected soils between 1980 and 2018;
- in the same period, barren areas, open shrublands, and grasslands were the land cover types with the largest area of salt-affected soils;
- among different forest types, evergreen broadleaf forests had the largest salt-affected area between 1980 and 2018;
- deserts and xeric shrublands, followed by montane grasslands and shrublands, were the biome types with the largest area of salt-affected soils between 1980 and 2018;

- regions with arid climate and polar tundra were the climate zones with the highest level of the salt-affected soils in the same period;
- annually ~9,466 km^2 were added in the same period to the total area of soils with an $EC_e \geq 4$ dS m^{-1} in South America; the trend was not statically significant for the rest of the continents ($p \geq 0.05$);
- over the same period and at the continental level, the highest rate of increase in the total area of soils with an ESP ≥ 6% was estimated for Asia with ~5,616 km^2 yr^{-1}; Asia was followed by South America and North America with respective increase rates of ~1,813 km^2 yr^{-1} and ~1,652 km^2 yr^{-1}; the trend was not statically significant for the rest of the continents ($p \geq 0.05$);
- at the country level, the highest rates of annual increase in the total area of soils with an $EC_e \geq 4$ dS m^{-1} was estimated for Brazil, Peru, Sudan, Colombia, and Namibia; and
- for sodicity (ESP ≥ 6%), the highest increase rates since 1980 were estimated for Iran, Saudi Arabia, Argentina, Afghanistan, and the USA.

The provided statistics on spatio-temporal variability of salinity and sodicity emphasise the high dynamism involved in soil salinization processes and necessity for implementation of effective policy responses to control the human-induced portion of those processes. Additionally, the results show that soil salinity and sodicity was highly variable at different geographical levels between 1980 and 2018. The uses of these data include, but are not limited to, food production and security, climate impacts, and soil classification studies.

5.3 Future of the soil salinity in drylands

The following specific conclusions can be inferred from the outputs of the models predicting future soil salinity in drylands:

- among the 14 predictors applied for developing the models, long-term annual precipitation frequency, the WRB soil classes, and long-term daily evapotranspiration are the most affecting predictors;
- in the mid- and long-term futures, compared to the reference period (1961 - 1990), and under different scenario pathways including RCP 4.5, RCP 8.5, SSP 2-4.5, and SSP 5-8.5, the analysis of the predictions indicate that the dryland areas of Mexico, southwest United States, South America, South Africa, and southern and Western Australia, and to a lesser degree, drylands of Morocco, Spain, and northern Algeria are the salinization

hotspots in response to variation in the key drivers of primary salinization used in this research;

- in the mid- and long-term futures, relative to the reference period (1961 - 1990), a reduction is projected in the soil salinity of the drylands spread across the north western United States, the Horn of Africa, Eastern Europe, Turkmenistan, and west Kazakhstan;
- overall, the findings show that in response to the projected climate change, the increase in soil salinity will be more severe in the southern latitudes, especially below -20°;
- under RCP 8.5 as the worst case scenario, Brazil, Namibia, and South Africa have the highest relative increase in the grid-cell mean values of soil salinity (1971 - 2100 relative to 1961 - 1990);
- under SSP 5-8.5, the countries with the highest relative increase in the grid-cell means of soil salinity are Botswana, South Africa, and Namibia, respectively;
- at the continental level, relative to the average of 1904 - 1999 period and under the RCP 8.5 or SSP 5-8.5 scenarios, predictions for the 2071 - 2100 period show an increasing trend in the total area of dryland soils with an $EC_e \geq 2$ dS m^{-1} for Australia and South America and a decreasing trend for Asia and Europe.

5.4 Methodological limitations and future research directions

5.4.1 Sustainable restoration of saline lakes

The present study used the case of Lake Urmia in Iran to investigate the applicability of the conceptual proposed framework for eco-hydrological and sustainable restoration of desiccating saline lakes. The limitations and the need for further future research here can be categorised into two groups, the ones related to the case study (Lake Urmia) used for examining the applicability of the proposed framework and the ones related to the framework itself. In the following, first the data and methodological limitations encountered during the application of the framework for restoration of the Lake Urmia are discussed and then the future research directions for improvement of the proposed framework are presented.

One of the major limitations during the application of the proposed framework for restoration of Lake Urmia was lack of reliable data for predictive calculations. For calculation of the available surface water within each basin, the run-off ratio was used while there are more advanced physical-based models for calculation of the run-off. However, these models needs a lot of detailed data of soil and landscape which were not available for Lake Urmia river basins. High resolution data of water uptakes by municipal and industrial sectors were not also available and these were calculated based on the average water usage per capita. There was no reliable data on the sewage water that returns to the system and accordingly, grey water was not considered in the calculations. A key parameter for application of the framework was the evaporation rate from the surface area of Lake Urmia which was calculated indirectly in the present study. Field measurements of the evaporation rate from the lake area can considerably improve the downscaling of the GCMs outputs and decrease the uncertainty in the estimates of the annual water inflow required for restoration of Lake Urmia. For the case of partial restoration, the vertical dust fluxes were calibrated using the field or experimental data available from other saline lakes as such data were not available for Lake Urmia bed.

Only 14 major crops in Urmia basin were considered in the land use/cropping optimisation and subject to data availability, other crops can be added to the analysis. Additionally, the required annual water for restoration of Lake Urmia was calculated assuming that target water level is 1274.1 m above mean sea level. This is based on the analysis of Abbaspour and Nazaridoust (2007) who estimated that the water salinity of the lake reduces to 240 g L^{-1} at this water level, and this salinity is tolerable by Lake Urmia's brine shrimp. However, there are a lot of uncertainties in the salinity that brine shrimp can tolerate (Agh 2007). Also it is very difficult to ensure that the salinity of Lake Urmia falls below 240 g L^{-1} at that water level as there are a huge amount of salt already precipitated on the lake bed which can dissolve in the lake water and increase the salinity.

Subject to data availability, the proposed framework is mainly applicable to the basins where human actions are the main cause of desiccation. The proposed framework for sustainable restoration of saline lakes also can be improved in the future research. The temporal resolution considered in the proposed framework was annual which means that seasonal variations are not taken into account. Also the analysis was conducted at the river basin level and subject to data availability, finer spatial resolutions, e.g. sub-basin can be considered for restoration. The proposed framework was based on single constant water level for estimation of the annual water required for restoration by supplying rivers. In a more realistic situation, a range of target water levels can be determined based on the social, economic, and environmental services provided by the saline lake. Also in the proposed framework, reduction of the upstream water uptake was majorly based on land use change in agricultural sector and other initiatives for decreasing the water consumption were not considered. Reducing the water evaporation from the soil/lake surface or transferring more refined wastewater into the lake are other solutions for reducing the water usage and can be considered in the framework. Also the framework can involve more socio-economic aspects for improving the agriculture-dependent livelihoods in saline lakes basins. For example, investing in and empowering the industrial activities in the saline lake basins can be a long-term substitution to compensate the revenue loss as result of the proposed land use change activities. However, how this can be strategically planned needs fundamental investigation.

5.4.2 Predicting the extent, severity, and trends in soil salinity and sodicity

The spatial and temporal resolution of the predictions made here can be improved. This needs environmental predictors with higher spatial and temporal resolutions.

In the present study, ML algorithms were used for developing predictive models of soil salinity and sodicity. One of the limitations of the ML algorithms is extrapolation of the results to locations where enough representative data are not available for training the models. The data used for training in the current study are biased towards the agricultural lands and developing countries. Addition of more and more data sampled from all geographical regions can improve the certainty of the predictions made by models. Additionally, the accuracy and methods used by different laboratories for measuring the soil salinity and sodicity have not been constant over the time and this can negatively impact the accuracy of the predictions. Erroneous samples and noises can result in serious mistakes in the generated maps. This heterogeneity also exists in the vertical direction and more samples are available for the top soil horizons compared to the deeper soil layers. Developing more analytical models and approaches for estimation of the soil salinity and salinity at large spatial scales can eliminate the need for developing data-driven models and improve the accuracy of the results.

In the present study, geographically explicit maps of uncertainty were not developed due to the computational load of the ML algorithms. In addition, the uncertainty that can be imposed by the lack of available data in the past periods is required to be quantified. As we move back, the number of available data for training the predictive models of soil salinity and sodicity decreases and this will impose uncertainty. Also it was not possible to quantify the error that propagates from the first part of the developed two-part predictive models to the second part.

Furthermore, there is an urgent need to develop an independent dataset of measured values of soil salinity and sodicity. This dataset can be used as a benchmark for evaluating the predictive power of the trained models in the current study and also the other available datasets for soil salinity and sodicity. Additionally, it was difficult to compare the total area of salt-affected soils provided here by other references as their definitions of the saline and sodic soils are different.

Finally, socio-economic aspects were not considered in the present study. How soil salinization impacts on the drylands' livelihoods can be monetised? The long-term economic costs of soil salinization can be a topic for future research studies.

References

Abbaspour, M., & Nazaridoust, A. (2007). Determination of environmental water requirements of Lake Urmia, Iran: an ecological approach. *International Journal of Environmental Studies, 64*(2), 161-169.

Agh, N. (2007). *Characterization of Artemia populations from Iran*: Ghent Univ.

Appendix 1

Lake area-level-volume relationships

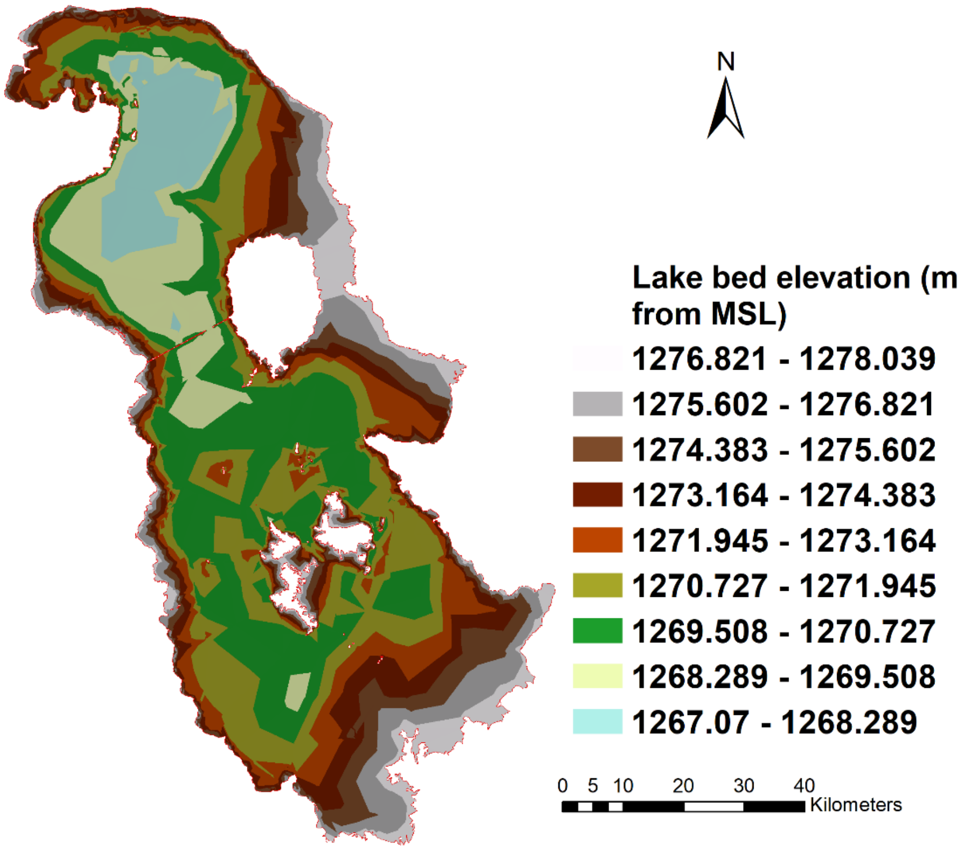


Figure A1-1: The elevation map of the Lake Urmia bed.

The volume of the whole Lake Urmia as a function of its water level (from mean sea level) can be expressed as:

$$Volume = a_1 \exp(-((level - b_1)/c_1)^2) + a_2 \exp(-((level - b_2)/c_2)^2) \qquad \text{(A1.1)}$$

where $a_1 = 4.09 \times 10^{10}$; $b_1 = 1280$; $c_1 = 4.802$; $a_2 = 6.661 \times 10^9$; $b_2 = 1274$; $c_2 = 3.141$

Eq. A1.1 with constant values listed below also can be used for estimation of the lake area as function of its water level.

$a_1 = 2.573 \times 10^9$; $b_1 = 1278$; $c_1 = 2.281$; $a_2 = 4.33 \times 10^9$; $b_2 = 1275$; $c_2 = 5.578$

Moreover, similar relations of volume-level and area-level are derived in the case of dividing the lake into two parts. These relations are obtained for southern part of the lake which will be restored in the lake partial restoration scenario. Using

$a_1 = 3.441 \times 10^{10}$; $b_1 = 1284$; $c_1 = 8.068$; $a_2 = -1.714 \times 10^9$; $b_2 = 1271$; $c_2 = 2.509$

in Eq. A1.1, the relation between the volume and level of the southern part of the lake can be obtained. Eq. A1.2 is also applied for calculation of the area of lake as a function of its water level.

$$\begin{aligned} area = a_{11}exp(-((level-b_{11})/c_{11})^2)+a_{22}exp(-((lev-b_{22})/c_{22})^2)+... \\ a_{33}exp(-((level-b_{33})/c_{33})^2)+a_{44}exp(-((level-b_{44})/c_{44}))^2 \end{aligned} \quad \text{(A1.2)}$$

where:

$a_{11} = 3.606\times10^9$; $b_{11} = 1278$; $c_{11} = 3.373$; $a_{22} = 1.539\times10^9$; $b_{22} = 1274$; $c_{22} = 2.095$; $a_{33} = 1.1\times10^9$; $b_{33} = 1272$; $c_{33} = 1.395$; $a_{44} = 6.513\times10^8$; $b44 = 1271$; $c_{44} = 0.8428$.

Effect of Lake Urmia's desiccation on the nearby environment

A fixed number of 200,000 particles per hour was released from dust sources to compute the concentration and direction of dust storm. These particles originated from Lake Urmia bed, as a result of 5 hours wind blow on 24 March 2018. Emission rate was set to be equal to the vertical dust flux (*F*) as a function of wind speed, calculated from above (vertical dust parametrization). Gridded meteorological data from the output of Global Data Analysis System (0.5-deg, every three hours on the native GFS hybrid sigma coordinate system - data is available at https://www.ready.noaa.gov/archives.php - and the same emission rate was applied to all defined point sources for the duration of the emission. The concentration grid resolution of the model was 0.5 × 0.5 degrees. A mean of 27 ensemble members (offset is one meteorological grid point in the horizontal and, 0.01 sigma units in the vertical direction) is calculated for 18 hours dispersion simulation. The particle speed of deposition was assumed to be 0.001 m s^{-1}.

Threshold friction velocity correction to moisture and roughness elements

Soil moisture increases the threshold friction velocity by strengthening soil cohesion. Y. Shao and E. Jung (personal communication, 2009) parameterised the effect of soil moisture on various soil textures using an experimental study and expressed a relation for the soil moisture correction factor (f_w) as follows

$$\begin{aligned} f_w &= 1 && for \quad w < w_r \\ f_w &= \sqrt{1+a(w-w_r)^b} && for \quad w > w_r \end{aligned} \quad \text{(A1.3)}$$

In above equation, w is the volumetric soil moisture, a and b are constants which depend on the soil type/texture, and w_r is the threshold soil moisture. These parameters can be found in Y. Shao and E. Jung (personal communication, 2009). Non-erodible elements can affect u_{*ft} because they cover the soil surface, protecting it from erosion by reducing the wind momentum energy over the surface (Raupach, 1992). Here, the threshold friction velocity correction factor (f_λ) proposed by Raupach (1992) and Raupach et al. (1993) is used for considering the effect of presence of roughness elements and is given by

$$f_\lambda = \sqrt{1-\sigma m\lambda}\sqrt{1+\beta m\lambda} \quad \text{(A1.4)}$$

where λ is the frontal area index or roughness density, m accounts for the non-uniformity of the surface stress, σ is the ratio of basal to frontal area of the elements, and β is the ratio of the drag coefficient for a single roughness element to that of a surface without roughness elements. Raupach et al. (1993) suggested $\sigma\approx 1$, $\beta\approx 90$, and $m \approx 0.5$. λ is a function of the fraction cover of non-erodible elements as follows

$$\lambda = -c_{\lambda} \ln(1 - f_r) \quad \text{(A1.5)}$$

and c_{λ} is 0.35 (Shao et al., 1996).

Salt mass balance

Table A1-1: The salt mass balance over the period 1996 - 2010.

Year	Average water level of the lake (m)	Lake water density (kg m^{-3})	2D surface (km^2)	Volume (km^3)	Lake modified volume (km^3)	Salt height change on 3D surface (m)
1996	1277.88	1155.97	5721.61	33.71	33.79	
1997	1277.48	1166.13	5717.54	31.41	32.02	-0.007
1998	1277.22	1175.34	5679.48	29.98	29.58	0.013
1999	1275.92	1187.26	5005.78	23.06	22.78	0.113
2000	1274.94	1250.73	4633.58	18.32	18.95	-0.082
2001	1274.44	1309.00	4451.67	16.07	16.34	-0.060
2002	1273.75	1332.60	4185.18	13.11	12.74	0.144
2003	1273.61	1298.59	4126.75	12.51	12.69	0.065
2004	1273.72	1323.41	4172.85	12.98	13.21	-0.087
2005	1273.57	1325.92	4108.60	12.33	12.36	0.041
2006	1273.18	1328.96	3996.50	10.79	11.30	0.057
2007	1273.15	1329.88	3944.54	10.66	11.09	0.007
2008	1272.58	1335.61	3653.32	8.47	87.36	0.134
2009	1272.07	1346.63	3362.97	6.70		0.121
2010	1271.68	1359.19	3098.89	5.43		0.141

References

Raupach, M., Gillette, D., & Leys, J. (1993). The effect of roughness elements on wind erosion threshold. *Journal of Geophysical Research: Atmospheres, 98*(D2), 3023-3029.

Raupach, M. R. (1992). Drag and drag partition on rough surfaces. *Boundary-Layer Meteorology, 60*(4), 375-395.

Shao, Y., Raupach, M. R., & Leys, J. F. (1996). A model for predicting aeolian sand drift and dust entrainment on scales from paddock to region. *Soil Research, 34*(3), 309-342.

Appendix 2

Extended Data

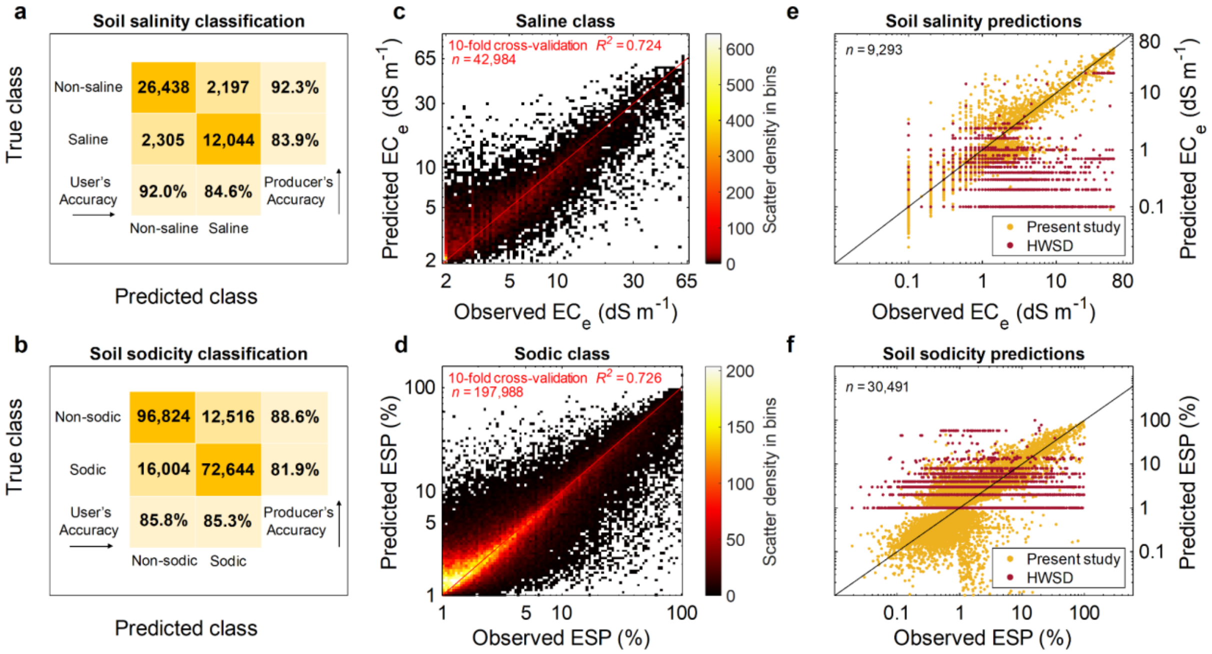


Figure A2-1: Validation of the predictive capability of the developed two-part models for global estimation of soil $\mathrm{EC_e}$ and ESP. a and **b**: The relation between classes of known measurements (True class) and the Predicted class as a result of 10-fold cross validation (10-CV). Producer's Accuracy shows the percentage of correct classifications relative to all classifications made by the classifier. User's Accuracy indicates to what percentage the predictions of the classifier can represent reality. **c** and **d**: Binned scatter plots showing the relation between the measured data and predictions of the regression part for saline and sodic classes as a result of 10-CV. **e** and **f**: Comparison between measured values of $\mathrm{EC_e}$ and ESP at the soil surface (0 - 30 cm) and the predicted values obtained using the model developed in the present study (R^2 $\mathrm{EC_e}$ = 0.83, R^2 ESP = 0.86) and those obtained from the Harmonised World Soil Database (HWSD; R^2 $\mathrm{EC_e}$ = 0.12, R^2 ESP = 0.26). A total of 9,293 and 30,491 measured data points are used in **e** and **f**.

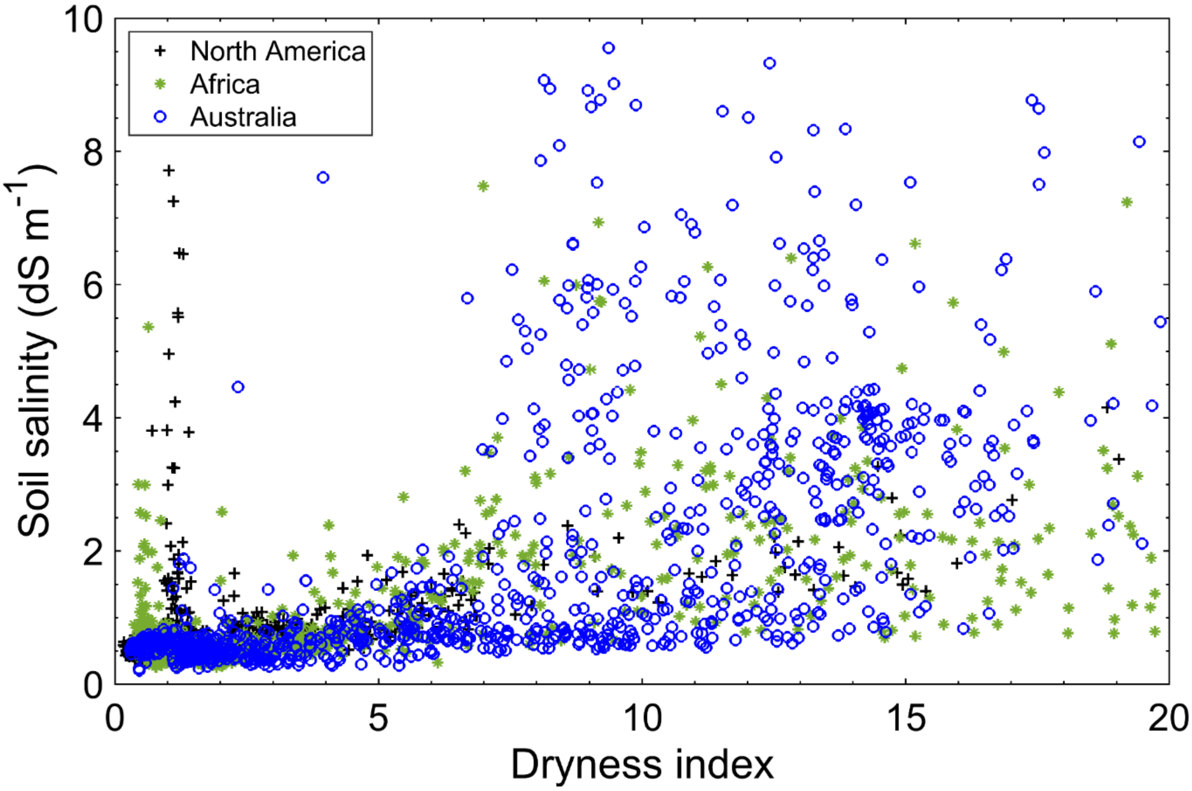


Figure A2-2: Catchment-scale average of the soil salinity predicted by ML-based models developed in the present study versus the dryness index (the ratio of long-term potential evapotranspiration to rainfall) for Australia, Africa, and North America.

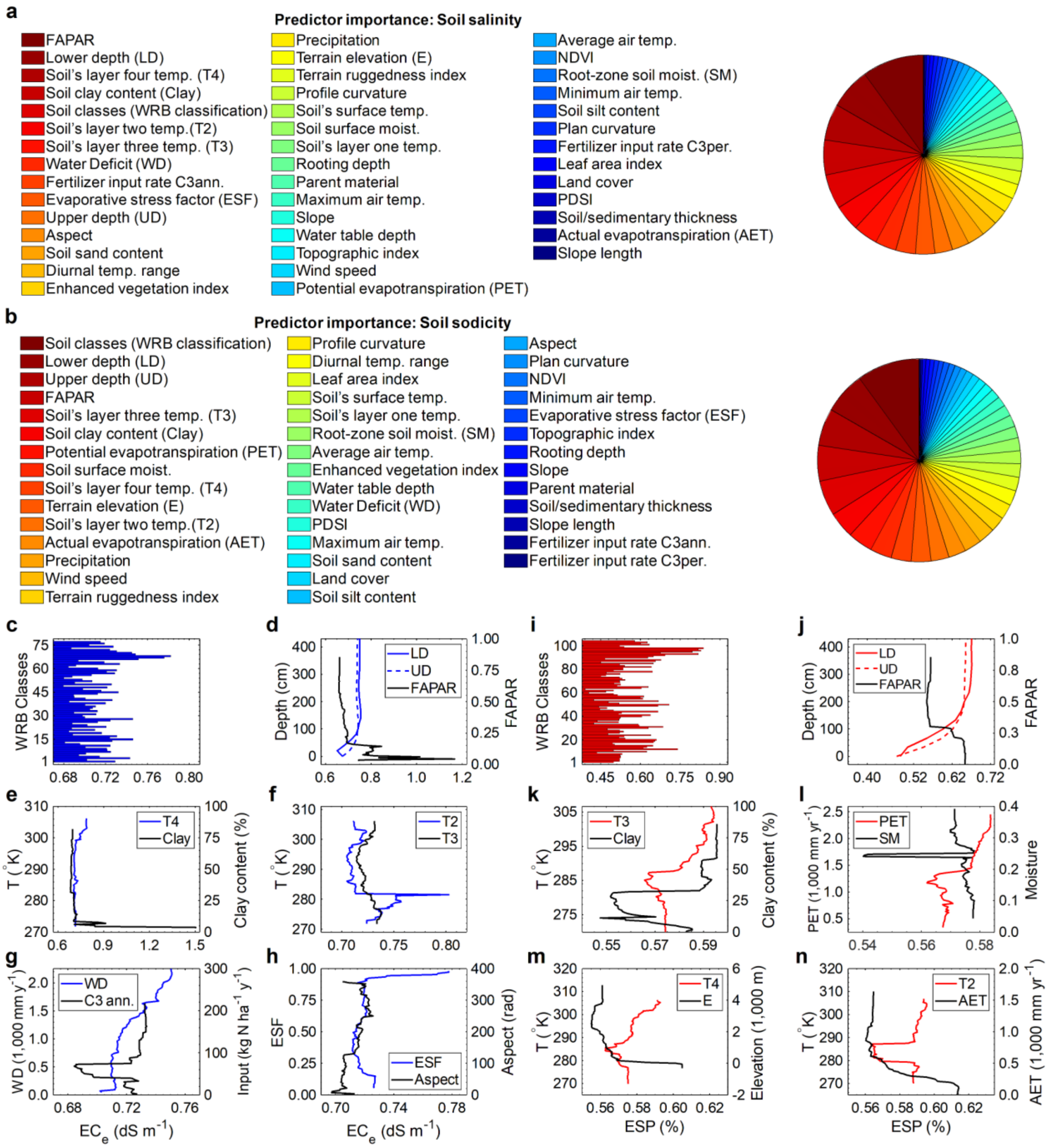
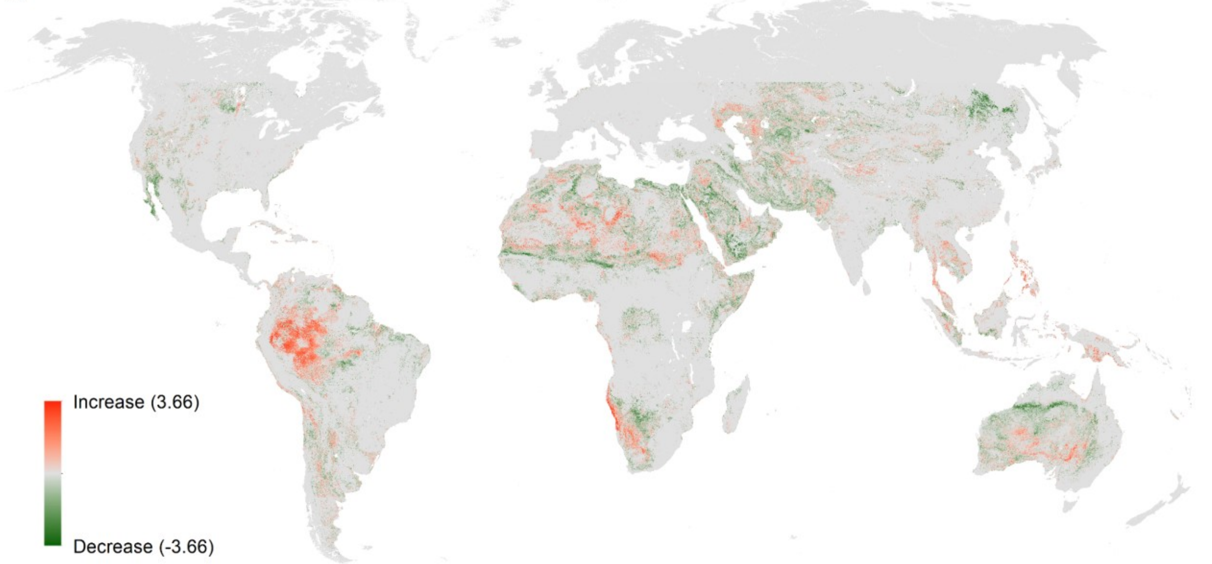


Figure A2-3: Predictor importance and partial dependency plots. a and **b**: The significance of predictors for regression models over the saline and sodic classes. **c** to **h**: The relation of the top 12 important predictors with predicted values of EC_e. **i** to **n**: The relation of the top 12 important predictors with predicted values of ESP. NDVI: Normalised Difference Vegetation Index; PDSI: Palmer Drought Severity Index; FAPAR: Fraction of Absorbed Photo-synthetically Active Radiation; C3ann.: C3-annual crops; C3per.: C3-perennial crops. For the full name and properties of the used predictors, see Table A2-1.

a

Change in likelihood of soils with $EC_e \geq 4$ dS m^{-1}



b

Change in likelihood of soils with ESP ≥ 6%

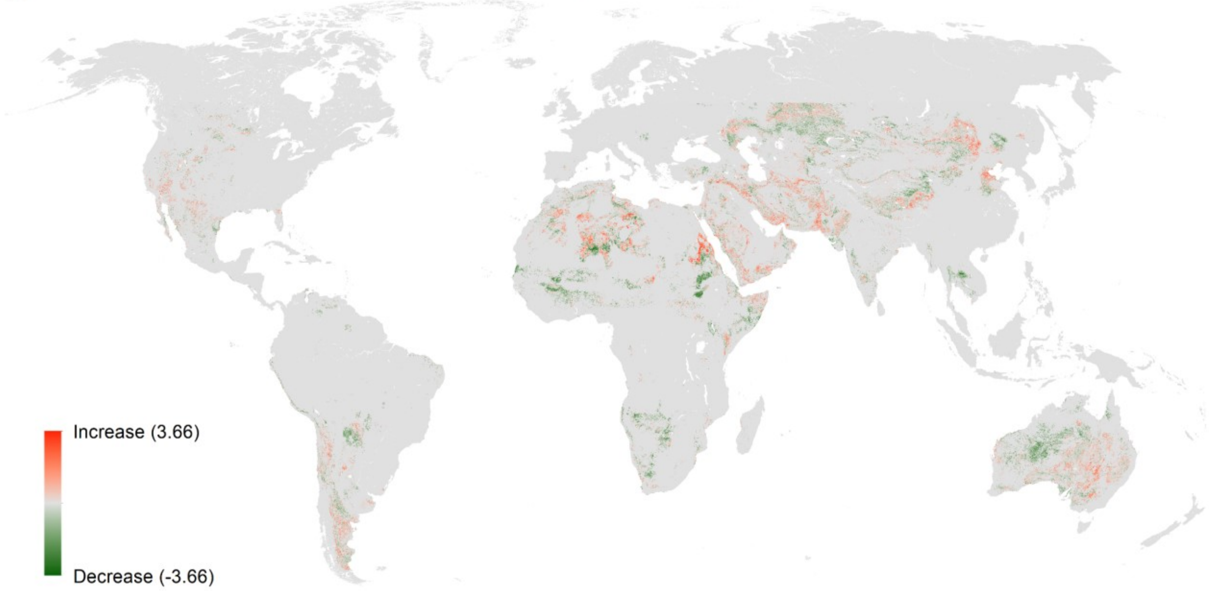


Figure A2-4: Global distribution of the change in likelihood (θ) of surface soils (0 - 30 cm) with an $EC_e \geq 4$ dS m^{-1} and ESP ≥ 6% in the 2000 to 2018 period, relative to the 1981 - 1999 period. A positive θ indicates that the likelihood has increased and a negative value shows that it has decreased. Maps are delimited to -55 and 55 latitudes and higher latitudes are shown only for improving the visualisation of the maps.

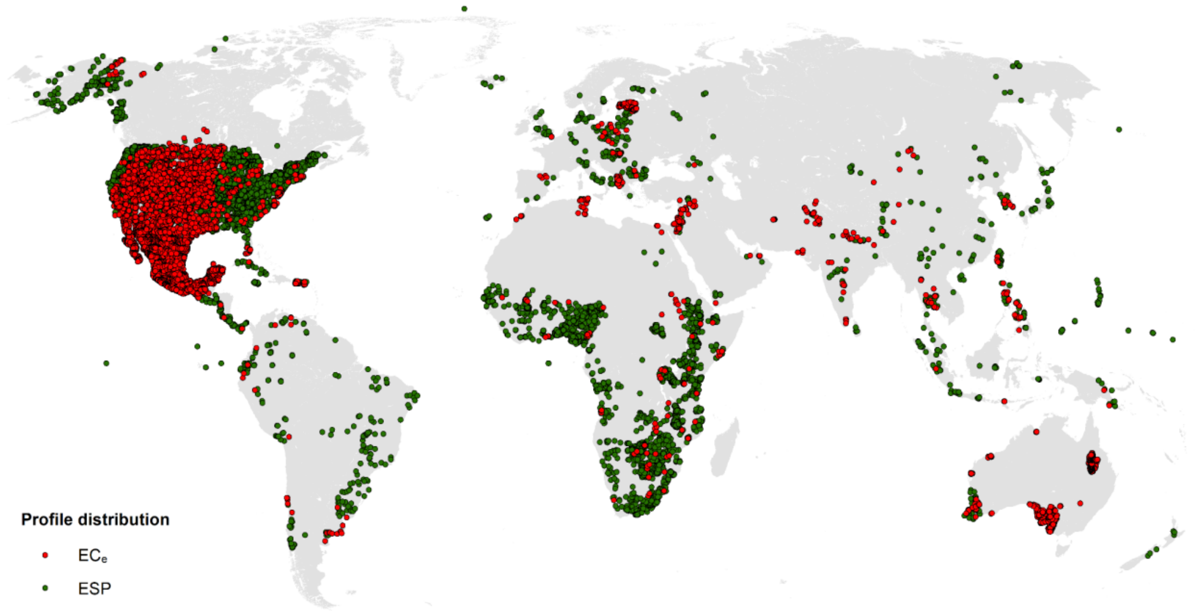


Figure A2-5: Profiles data distribution used as input for training the two-part models.

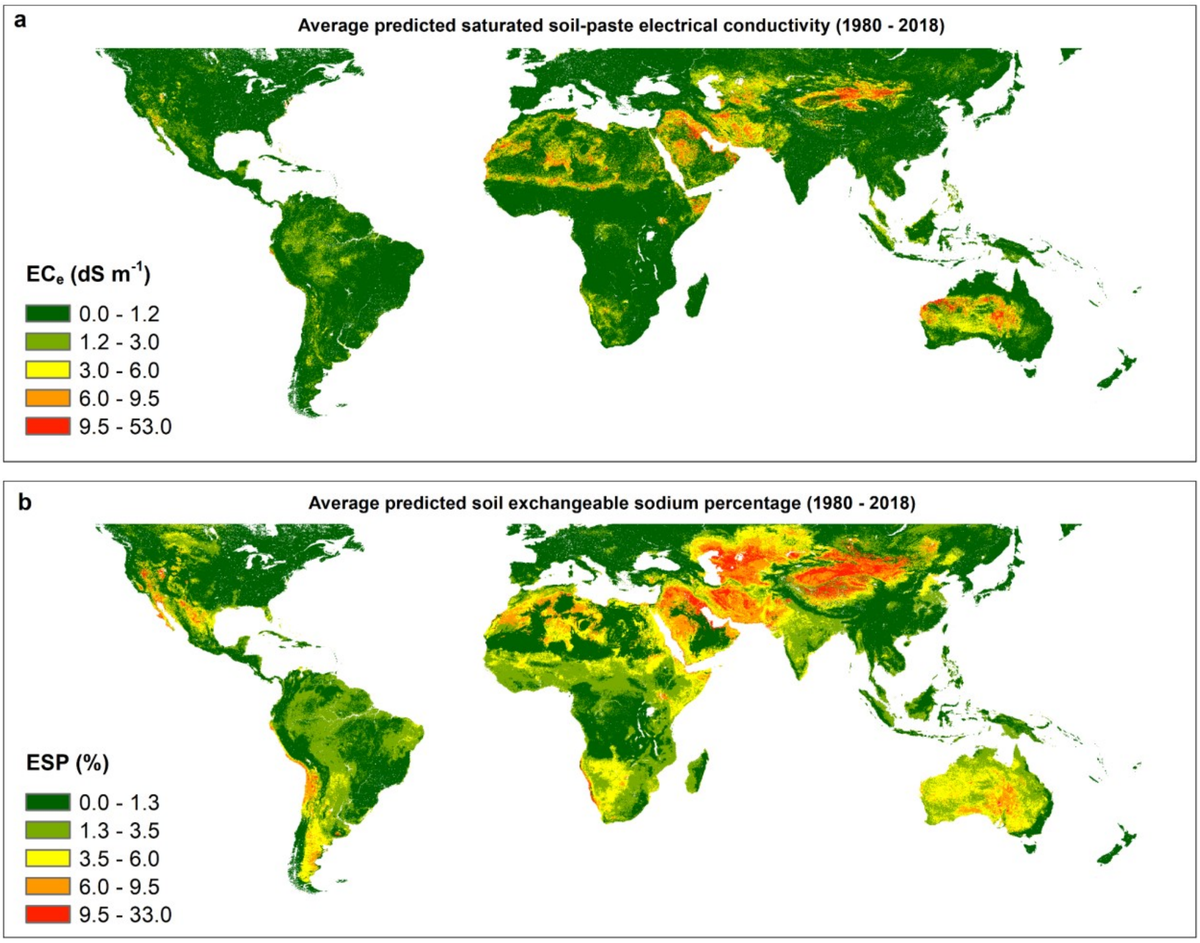


Figure A2-6: Average of annual predictions for surface soil (0 - 30 cm) EC_e and ESP between 1980 and 2018.

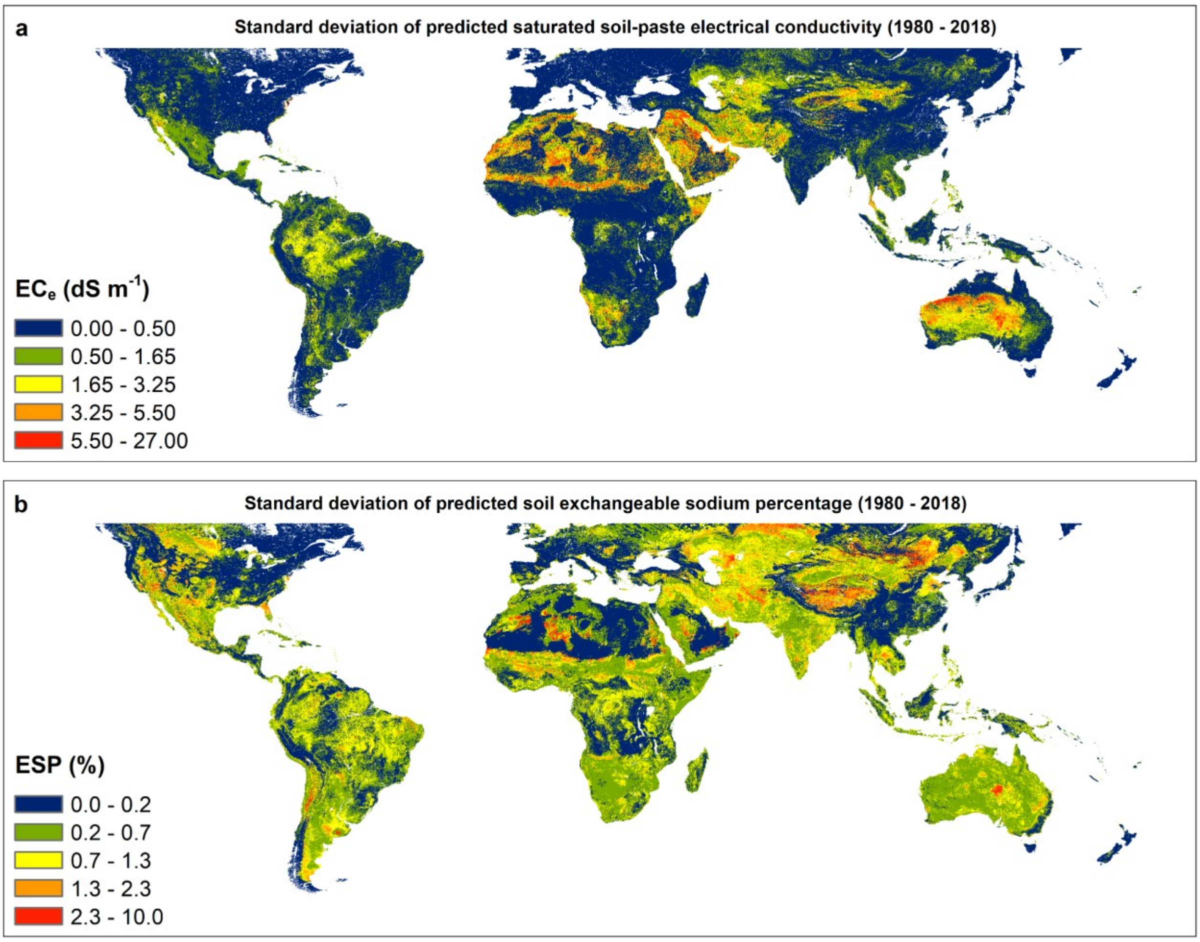


Figure A2-7: Standard deviation of annual predictions for EC_e and ESP between 1980 and 2018.

Table A2-1: Static predictors used for training the two-part models.

Static predictors					
Predictor	**Pre-processing**	**Source**	**Projection**	**Extent**	**Resolution**
Sample's upper depth (cm)	-	Original soil datasets	-	-	-
Sample's lower depth (cm)					
Elevation (m)	Projected the original DEMs to the World Mercator x- y-coordinates system (at 259.511 m resolution) by the cubic convolution method to calculate predictors' values in SAGA GIS (elevation only included re-projection).	SRTM Digital Elevation Database v4.1	GCS WGS 1984	180W-170E, 60S-60N	0.00208°
Plan curvature					
Profile curvature					
Slope (degrees)					
Slope length (m)	Projected the original DEMs to the World Mercator x- y-coordinates system (at 1,000 m resolution) by the cubic convolution method for reducing computational time in SAGA GIS.				
Terrain Ruggedness Index (TRI)					
Aspect (degrees)	-				
Fertiliser input rate for C3-annual and perennial crops (kg nitrogen ha^{-1} y^{-1}). C3 is one of the pathways that plants use to fix carbon during the process of photosynthesis. For C3 plants, the first carbon compound produced during photosynthesis contains three carbon atoms.	The original annual .nc layers were converted to geo-tiff rasters and per-cell average of rasters between 1980 and 2018 was calculated using the ArcGIS "cell statistics" tool.	Land use Harmonization (LUH2 v2h_high)	GCS WGS 1984	180W-180E, 90S-90N	0.25°
World Reference Base soil classes (120 classes)	-	ISRIC-SoilGrids250	GCS WGS 1984	180W-180E, 62S-87.37N	0.00208°
Soil clay content (%)	Per-cell average of five standard soil depths: 0, 15, 30, 60, and 100 cm was calculated using the trapezoidal rule and ArcGIS "cell statistics" tool.				
Soil silt content (%)					
Soil sand content (%)					
Parent material lithological classes (16 classes)	The original shape file was first converted to a geo-tiff format and then re-projected to the GCS WGS 1984 (0.01055° resolution).	GLiM	World Eckert IV	Left: -16,653,453.7 m Right: 16,653,453.7 m Bottom: -8,460,600.9 m Top: 8,376,733.0 m	Polygon
Water table depth (m)	Raster datasets for different continents were merged into a single global one.	Fan et al. (2013)	GCS WGS 1984	180W-180E, 53S-84N	0.00833°
Topographic Index	The original .nc file was converted to a geo-tiff raster.	Marthews et al. (2014)	GCS WGS 1984	180W-180E, 56.35S-86.09N	0.00208°
Average soil and sedimentary-deposit thickness (m)	-	Pelletier et al. (2016)	GCS WGS 1984	180W-180E, 60S-90N	0.00833°
Average plant rooting depth (m)	The original dataset was geo-referenced to the GCS WGS 1984 coordinates system by the nearest neighbour method.	ISLSCP II Ecosystem Rooting Depths (95ecosys_rootdepth)	Undefined	180.25W-179.75E, 90.25S-89.75N	0.5°

Table A2-1 cont.: Dynamic predictors used for training the two-part models.

Dynamic predictors						
Predictor	**Averaging Method**	**Source**	**Projection**	**Extent**	**Spatial resolution**	**Pre-processing**
Precipitation (mm yr^{-1})	Decadal average of yearly accumulations	CRU TS v. 4.03	GCS WGS 1984	180W-180E, 90S-90N	0.5°	Original monthly .nc files were converted to geo-tiff layers and decadal per-cell averages were computed in ArcGIS using the “cell statistics” tool.
Potential evapotranspiration (mm yr^{-1})						
Diurnal temperature range (°C)	Decadal average of monthly means					
Average air temperature (°C)						
Maximum air temperature (°C)						
Minimum air temperature (°C)						
Actual evapotranspiration (mm yr^{-1})	Five-year average of yearly accumulations	TerraClimate	GCS WGS 1984	180W-180E, 90S-90N	0.0416°	Original monthly .nc files were converted to geo-tiff layers and five-year per-cell averages were computed in ArcGIS using the “cell statistics” tool.
Water deficit (mm yr^{-1})						
PDSI	Five-year average of monthly values					
Root-zone soil moisture (mm)						
Soil surface (2-5 cm) moisture (percentage of total saturation), remotely-sensed by satellites	Yearly mean	Soil moisture gridded data (v201812.0.1) Climate Data Store	GCS WGS 1984	180W-180E, 90S-90N	0.25°	Original monthly (combined passive and active sensor type) .nc files were converted to geo-tiff layers and annual per-cell averages were computed in ArcGIS using the “cell statistics” tool.
Evaporative stress factor (S)		GLEAM v3.3 Datasets	GCS WGS 1984	180W-180E, 90S-90N	0.25°	Original daily .nc files were converted to geo-tiff layers and annual per-cell averages were computed in ArcGIS using the “cell statistics” tool.
Two-band Enhanced Vegetation Index (EVI2)		NASA (LP DAAC) Vegetation Index and Phenology Vegetation Indices, VIP30 v. 004 (Note for 2014-2018, we used NDVI and EVI from MOD13C2 Version 6 product from Terra MODIS)	GCS Unknown datum based upon the Clarke 1866 ellipsoid	180W-180E, 90S-90N	0.05°	EVI2 (and/or EVI) and NDVI sub-datasets were extracted from the original monthly .hdf files, re-projected to the GCS WGS 1984 using the bilinear interpolation method, and saved as geo-tiffs. Annual per-cell averages of EVI2 (and/or EVI) were then calculated in ArcGIS using the “cell statistics” tool.
Normalised Difference Vegetation Index (NDVI)						
Fraction of Absorbed Photo-synthetically Active Radiation (FAPAR)		NOAA (National Oceanic and Atmospheric Administration) Climate Data Record (CDR) of Advanced Very High Resolution Radiometer (AVHRR) Surface Reflectance	GCS WGS 1984	180W-180E, 90S-90N	0.05°	Original daily .nc files were converted to geo-tiffs. Annual per-cell averages of FAPAR and LAI were then calculated from the daily raster layers in ArcGIS using the “cell statistics” tool.
Leaf Area Index (LAI)						
Wind speed (m s^{-1})		ERA5 re-analysis monthly averages adopted from Climate Data Store	GCS WGS 1984	0.125W-359.875E, 90.125S-90.125N	0.25°	Original Monthly .nc files were converted to geo-tiffs and re-projected to the GCS WGS 1984 to change the extent of the rasters (180W-180E, 90S-90N was the desirable extent). Annual per-cell averages were then calculated from daily raster layers in ArcGIS using the “cell statistics” tool.
Soil skin temperature (°K)						
Soil layer one (0 -7 cm) temperature (°K)						
Soil layer two (7-28 cm) temperature (°K)						
Soil layer three (28-100 cm) temperature (°K)						
Soil layer four (100-289 cm) temperature (°K)						
Land cover (16 classes)	1980 - 1996, attributed to the 1993 land cover layer.	Global Land Cover Characteristics Data Base Version 2.0	GCS WGS 1984	180W-180E, 90S-90N	0.00833°	The International Geosphere-Biosphere Programme (IGBP) land cover legend was chosen since it was available in both datasets. IGBP sub-datasets were extracted from the original .hdf files and saved as geo-tiffs. Layers with the sinusoidal coordinates system were re-projected to the GCS WGS 1984 using the nearest neighbour method.
	1997 - 2018, attributed to the layer with the nearest year of acquisition.	Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) products for years 2000, 2006, 2014, and 2018	Unknown datum based upon the custom spheroid sinusoidal	Left: -20,015,108.8 m Right: 20,015,107.7 m Bottom: -10,007,554.1 m Top: 10,007,554.1 m	463.31 meter	

Table A2-2: Accuracy metrics and the results of hyperparameter optimisation for different parts of the fitted two-part models.

	Hyperparameters (Breiman 2001)			**Method**	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	**Split Criterion**
EC_e (dS m^{-1})	Classification		LB[a]	-	34.76	0.32	4.97	9,886.65	5.44	-
			UB[b]	-	56.51	0.55	39.58	18,553.18	18.44	-
			Best	AdaBoostM1	50.00	0.47	21.00	2,664.00	-	gdi
	Regression	None (0 - 2)[c]	LB	-	61.37	0.09	5.97	9,716.32	11.34	-
			UB	-	97.83	0.13	12.60	15,242.99	18.04	-
			Best	LSBoost	80.00	0.07	8.00	23,408.00	8.00	-
		Saline (2 - 60)	LB	-	105.78	0.08	4.60	5,037.27	8.80	-
			UB	-	166.89	0.13	11.10	8,632.70	15.96	-
			Best	LSBoost	366.00	0.05	3.00	9,365.00	1.00	-
	Classification accuracy metrics			**Binomial deviance loss**	**Classification error**	**Accuracy (%)**	**Precision**	**Recall**	***MCC*[d]**	***MOF*[e]**
	Classification		LB	0.192	0.120	88.338	0.922	0.897	0.743	0.118
			UB	0.218	0.124	88.876	0.927	0.909	0.753	0.124
			Best	0.187	0.117	89.650	0.921	0.924	0.767	0.109
	Regression accuracy metrics			***RMSE* (log)**	***MAE* (log)**	***NSE*[f] (log)**	***RMSE***	***MAE***	***NSE***	***MOF***
	Regression	None (0 - 2)	LB	0.069	0.047	0.711	0.294	0.189	0.640	0.005
			UB	0.070	0.048	0.718	0.297	0.192	0.648	0.005
			Best	0.068	0.047	0.727	0.289	0.186	0.659	0.005
		Saline (2 - 60)	LB	0.190	0.129	0.730	5.230	2.501	0.703	0.037
			UB	0.193	0.132	0.738	5.318	2.551	0.713	0.038
			Best	0.187	0.127	0.747	5.119	2.451	0.724	0.037
	Hyperparameters (Breiman 2001)			**Method**	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	**Split Criterion**
ESP (%)	Classification		LB	-	75.80	-	1.27	80,234.76	3.42	-
			UB	-	122.04	-	1.74	121,561.84	6.81	-
			Best	Bag	208.00	-	1.00	80,815.00	2.00	deviance
	Regression	None (0 - 1)	LB	-	220.45	0.04	6.80	19,872.75	2.03	-
			UB	-	313.56	0.06	11.33	45,713.13	2.80	-
			Best	LSBoost	378.00	0.03	12.00	95,924.00	2.00	-
		Sodic (1 – 98.59)	LB	-	196.73	0.06	4.80	34,059.28	2.27	-
			UB	-	281.72	0.09	11.75	53,918.61	2.93	-
			Best	LSBoost	295.00	0.03	1.00	36,274.00	2.00	-
	Classification accuracy metrics			**Binomial deviance loss**	**Classification error**	**Accuracy (%)**	**Precision**	**Recall**	***MCC***	***MOF***
	Classification		LB	0.226	0.148	85.053	0.851	0.883	0.697	0.152
			UB	0.229	0.149	85.248	0.854	0.885	0.701	0.154
			Best	0.229	0.144	85.593	0.859	0.885	0.708	0.149
	Regression accuracy metrics			***RMSE* (log)**	***MAE* (log)**	***NSE* (log)**	***RMSE***	***MAE***	***NSE***	***MOF***
	Regression	None (0 - 1)	LB	0.071	0.046	0.556	0.226	0.142	0.530	0.005
			UB	0.071	0.046	0.560	0.227	0.144	0.533	0.005
			Best	0.071	0.046	0.563	0.225	0.142	0.537	0.005
		Sodic (1 – 98.59)	LB	0.231	0.160	0.740	6.924	2.683	0.705	0.056
			UB	0.233	0.162	0.744	7.030	2.726	0.714	0.057
			Best	0.231	0.158	0.744	6.772	2.616	0.726	0.055

[a] Lower band

[b] Upper band

[c] Minimum and maximum of the training set

[d] Mathews Correlation Coefficient

[e] Minimum observed objective function

[f] Nash-Sutcliffe model efficiency coefficient

Table A2-3: Total area of the salt-affected soils at the country level.

Country	Salt-affected area (Mha)	Country	Salt-affected area (Mha)
China	211.748	Tunisia	2.163
Australia	131.407	South Africa	1.995
Kazakhstan	93.312	Colombia	1.657
Iran	88.336	Kuwait	1.573
Saudi Arabia	68.191	Eritrea	1.494
Algeria	63.932	Turkey	1.391
Mongolia	42.981	Malaysia	1.265
Turkmenistan	36.992	Tajikistan	1.228
Pakistan	36.195	Indonesia	1.227
Iraq	30.544	Botswana	1.125
Uzbekistan	27.355	Qatar	1.038
Libya	21.535	Thailand	1.019
Mexico	21.073	Angola	0.79017
United States	20.747	Myanmar	0.75299
Afghanistan	20.379	Israel	0.65719
Niger	17.341	Philippines	0.65335
Argentina	17.244	Nigeria	0.61718
Mauritania	17.024	Azerbaijan	0.59546
Chile	15.758	Djibouti	0.52259
Western Sahara	15.254	Papua New Guinea	0.39629
Chad	14.761	Venezuela	0.36037
Sudan	13.202	Senegal	0.30626
Somalia	13.114	Bahamas	0.27594
Syria	11.647	Kyrgyzstan	0.25453
Oman	10.495	Vietnam	0.21236
Mali	9.863	Burkina Faso	0.19560
Namibia	8.605	Guyana	0.19282
Peru	7.782	Republic of Congo	0.18667
Morocco	7.295	Ecuador	0.17812
India	6.938	North Korea	0.14791
Jordan	6.768	Madagascar	0.12221
Egypt	6.695	Netherlands	0.10879
Yemen	6.545	Tanzania	0.10358
Kenya	5.261	Palestine	0.10204
Brazil	4.998	Gabon	0.10049
Bolivia	4.681	Cuba	0.09567
Ethiopia	3.599	Cambodia	0.09222
United Arab Emirates	2.966	Mozambique	0.08535
Democratic Republic of the Congo	2.230	Japan	0.06442

Table A2-4: Total area of the salt-affected soils at the climate, biome, and land cover levels. For the full name of the climate zones see Figure A2-12.

Climate	Salt-affected area (Mha)	Biome	Salt-affected area (Mha)
BWh	594.398	Deserts and Xeric Shrublands	928.225
Bwk	339.908	Montane Grasslands and Shrublands	86.454
Bsk	103.596	Tropical and Subtropical Grasslands, Savannas and Shrublands	52.459
ET	55.528	Temperate Grasslands, Savannas and Shrublands	38.064
BSh	43.330	Tropical and Subtropical Moist Broadleaf Forests	16.420
Af	9.513	Mediterranean Forests, Woodlands and Scrub	15.019
AW	4.718	Temperate Broadleaf and Mixed Forests	14.220
Csa	4.461	Flooded Grasslands and Savannas	10.305
Am	4.445	Temperate Conifer Forests	5.514
Cfa	3.015	Tropical and Subtropical Dry Broadleaf Forests	2.734
Cwa	1.965	Mangroves	1.538
Dfb	1.940	Tropical and Subtropical Coniferous Forests	0.195
CSb	1.346		
Dwc	1.142	**Land cover**	**Salt-affected area (Mha)**
Cfb	1.112	Barren	536.109
Dfa	0.961	Open Shrublands	144.120
Dfc	0.777	Grasslands	77.372
Dwa	0.506	Croplands	16.490
Dsb	0.480	Evergreen Broadleaf Forests	10.164
Cwb	0.419	Savannas	0.343
Dsa	0.399	Woody Savannas	0.151
As	0.291	Mixed Forests	0.114
Dwb	0.279	Evergreen Needleleaf Forests	0.097
Cwc	0.237	Deciduous Broadleaf Forests	0.020
Dsc	0.090	Closed Shrublands	0.007
EF	0.007		
Csc	0.006		
Cfc	0.004		

Frequency distribution of cell-level likelihoods and trends

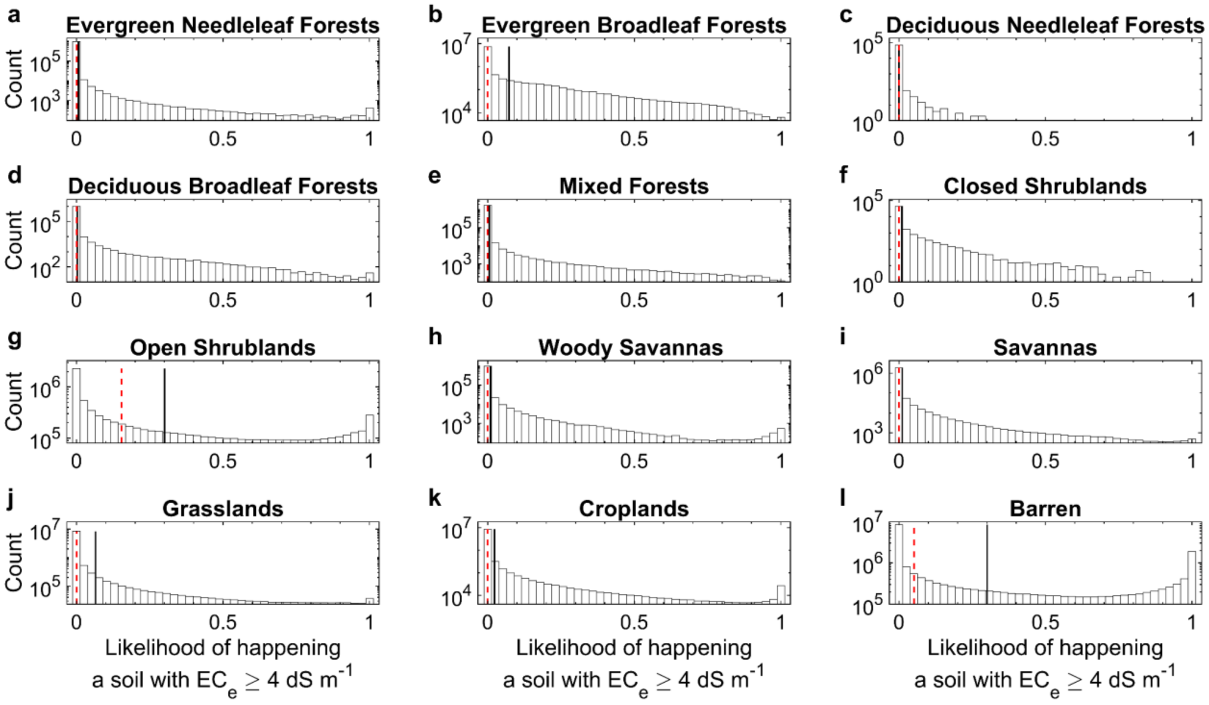


Figure A2-8: Frequency distribution of the cell-level likelihood of soils with an $EC_e \geq 4$ dS m^{-1} between 1980 and 2018 at each land cover type. Black and red dotted lines indicate the mean and median of predictions, respectively.

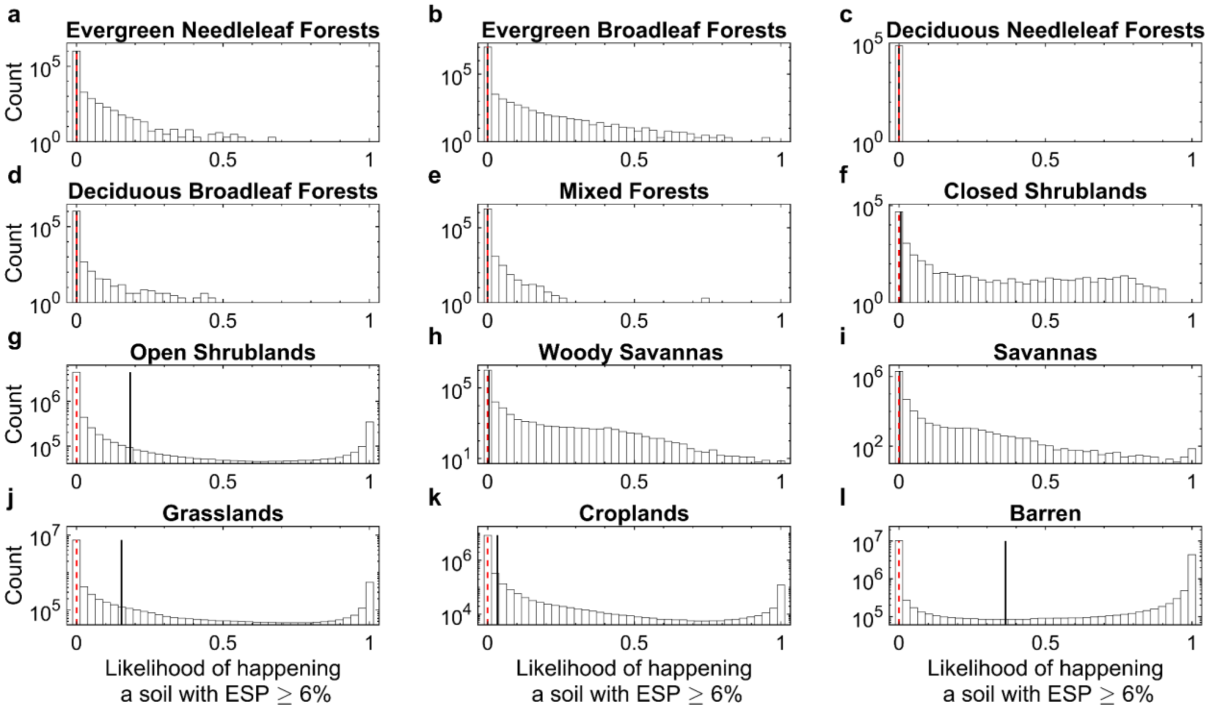


Figure A2-9: Frequency distribution of the cell-level likelihood of soils with an ESP ≥ 6% between 1980 and 2018 at each land cover type. Black and red dotted lines indicate the mean and median of predictions, respectively.

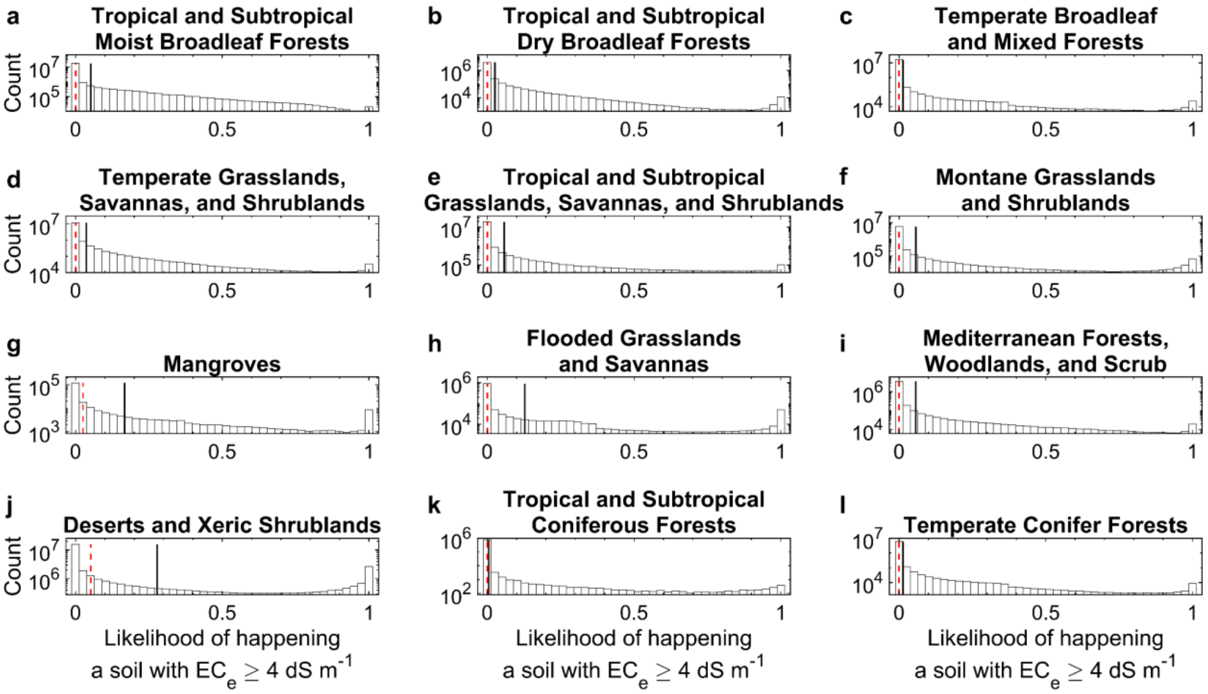


Figure A2-10: Frequency distribution of the cell-level likelihood of soils with an $EC_e \geq 4$ dS m^{-1} between 1980 and 2018 at each biome. Black and red dotted lines indicate the mean and median of predictions, respectively.

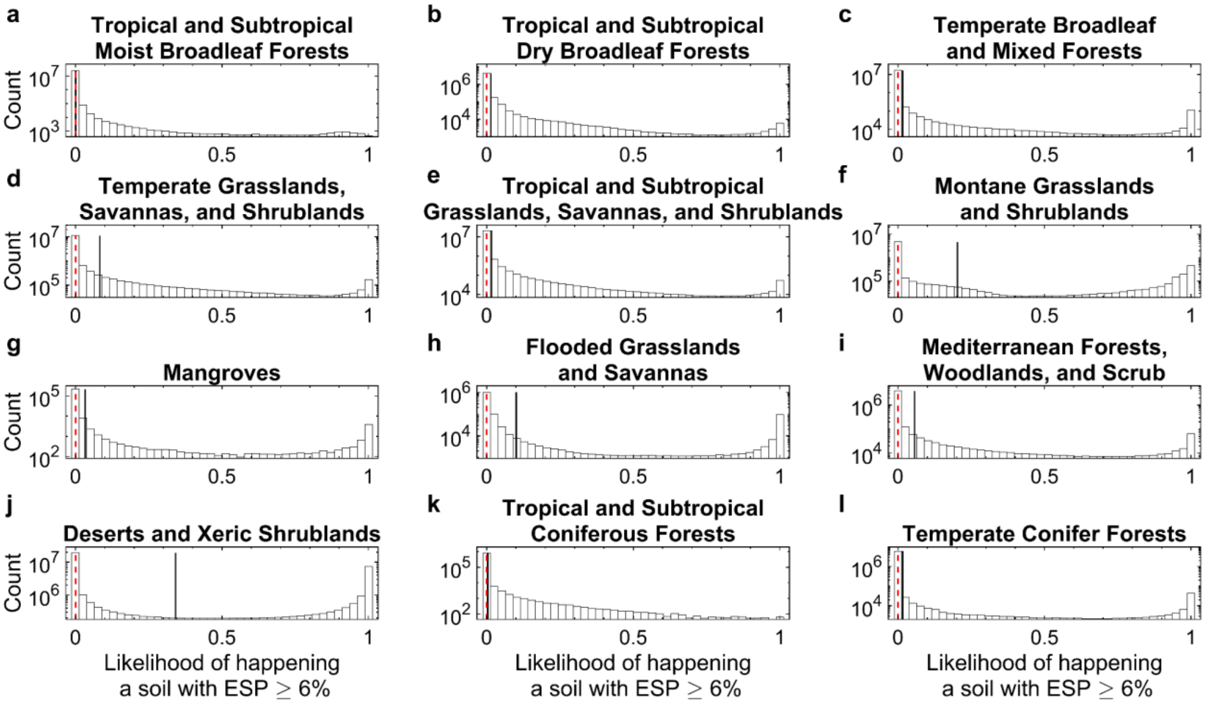


Figure A2-11: Frequency distribution of the cell-level likelihood of soils with an ESP ≥ 6% between 1980 and 2018 at each biome. Black and red dotted lines indicate the mean and median of predictions, respectively.

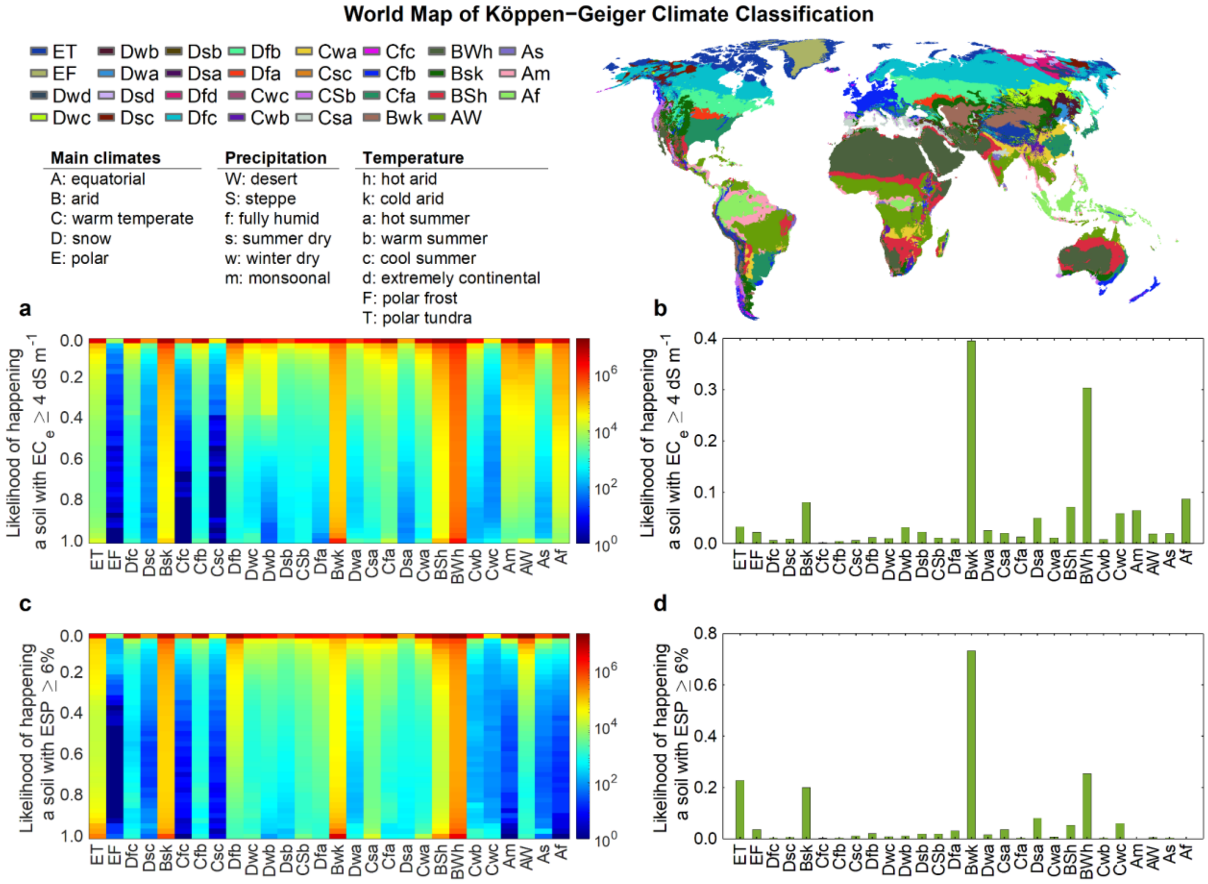


Figure A2-12: Frequency distribution of the cell-level likelihood of salt-affected soils at each climate between 1980 and 2018. a and **b**: Soils with $EC_e \geq 4$ dS m^{-1}. **c** and **d**: Soils with ESP ≥ 6%. Heat map charts show the count of data within each climate zone. Bar charts show the mean of likelihoods within each climate zone.

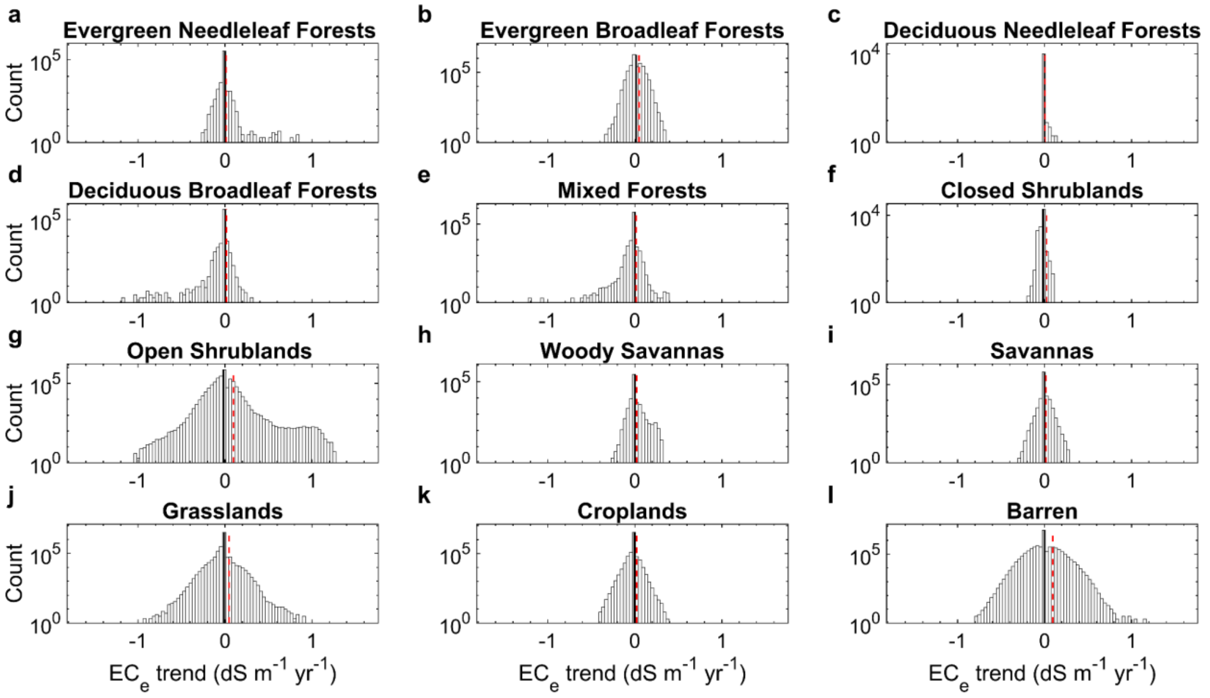


Figure A2-13: Frequency distribution of the cell-level trends in variation of EC$_e$ ($p < 0.05$) between 1980 and 2018 for each land cover type. Black and red dotted lines indicate the mean and standard deviation of the calculated trends.

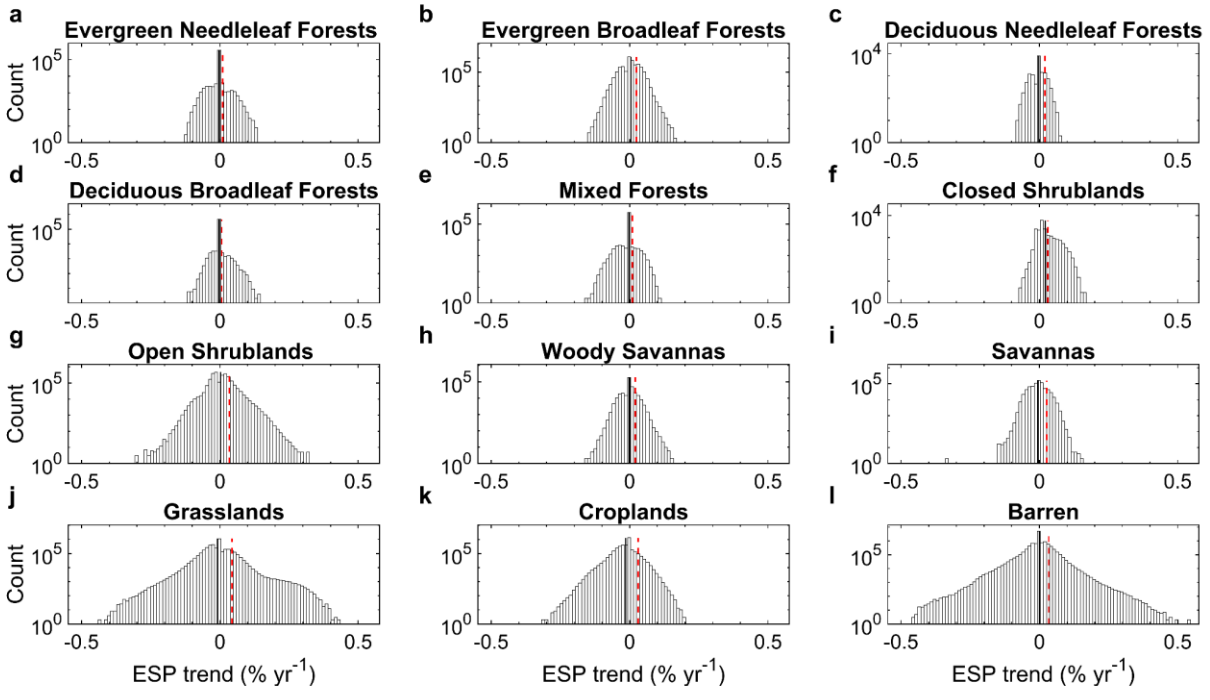


Figure A2-14: Frequency distribution of the cell-level trends in variation of ESP ($p < 0.05$) between 1980 and 2018 for each land cover type. Black and red dotted lines indicate the mean and standard deviation of the calculated trends.

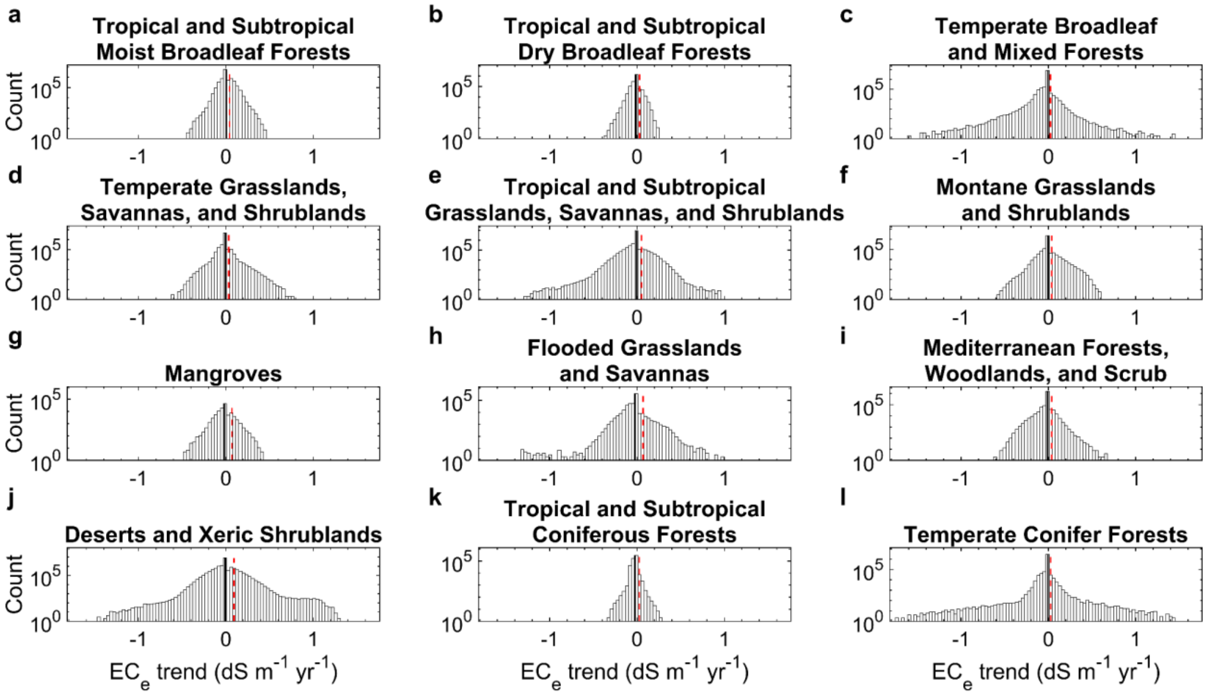


Figure A2-15: Frequency distribution of the cell-level trends in variation of EC_e ($p < 0.05$) between 1980 and 2018 for each biome. Black and red dotted lines indicate the mean and standard deviation of the calculated trends, respectively.

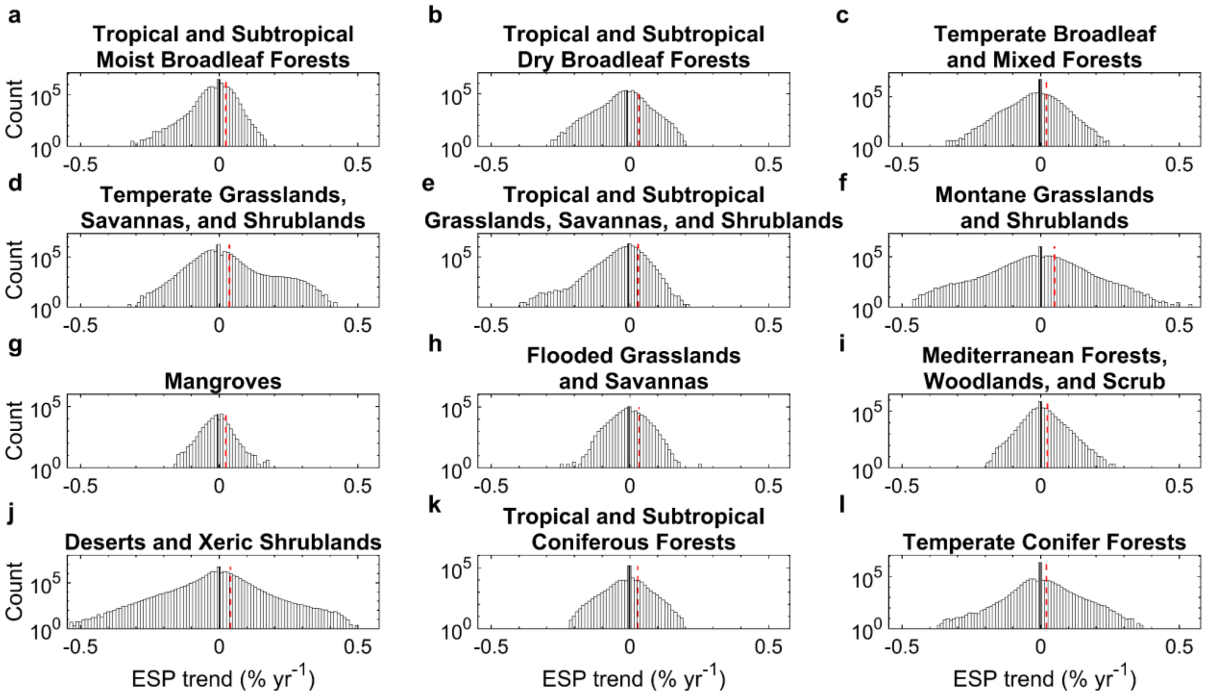


Figure A2-16: Frequency distribution of the cell-level trends in variation of ESP ($p < 0.05$) between 1980 and 2018 for each biome. Black and red dotted lines indicate the mean and standard deviation of the calculated trends, respectively.

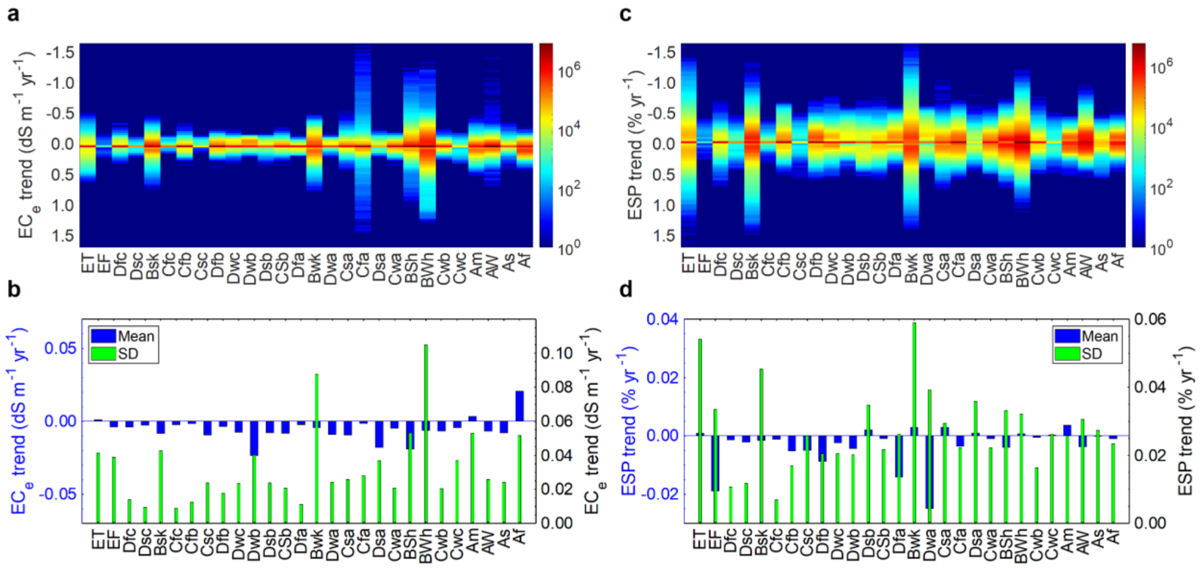


Figure A2-17: Cell-level trends in variation of soil salinity and sodicity for each climate zone between 1980 and 2018 (p < 0.05). a and **b**: Frequency distribution, mean, and standard deviation (SD) of the cell-level trends in vriation of EC_e. **c** and **d**: Frequency distribution, mean, and standard deviation of the cell-level trends in variation of ESP. Heat map charts show the count of data within each climate zone. See Figure A2-12 for the full name of each climate zone.

Model training

Table A2-5: The significance of predictors for classification and regression models over the training sets and saline/sodic classes. Importance values are normalised between 0 and 1 and the higher the value, the more the significance. See Table A2-1 for the full name of the predictors.

Predictor	**Min**	**Max**	**Mean**	**Median**	**Standard deviation**	**Importance for EC_e**		**Importance for ESP**	
						Classification	Regression (saline)	Classification	Regression (sodic)
Sample's upper depth	0.00	3,277.00	60.78	41.00	88.52	0.837	0.287	0.559	0.450
Sample's lower depth	0.00	3,292.00	86.54	66.00	95.66	1.000	0.669	0.641	0.519
Elevation	-315.00	5,158.00	595.06	310.00	643.58	0.216	0.239	0.298	0.157
Plan curvature	-0.34	0.33	0.00	0.00	0.02	0.253	0.071	0.186	0.081
Profile curvature	-0.31	0.39	0.00	0.00	0.02	0.269	0.213	0.285	0.137
Slope	0.00	70.00	6.58	3.00	9.16	0.064	0.154	0.156	0.056
Slope length	0.00	75,083.30	1,007.75	0.00	2,286.11	0.080	0.000	0.047	0.023
TRI	0.00	471.04	23.54	11.95	34.69	0.235	0.219	0.235	0.137
Fertiliser input for C3 annual crops	0.00	329.87	70.06	81.09	25.80	0.000	0.320	0.000	0.012
Fertiliser input for C3 perennial crops	0.00	415.51	36.87	0.00	70.47	0.011	0.071	0.000	0.000
Water table depth	0.00	466.07	23.75	13.43	32.18	0.244	0.140	0.171	0.110
Aspect	0.00	359.00	178.64	180.00	106.01	0.293	0.277	0.186	0.081
Topographic index	-1.85	19.91	6.17	5.95	2.29	0.318	0.127	0.186	0.069
Soil clay content	0.00	90.00	23.11	21.00	11.80	0.347	0.568	0.217	0.168
Soil silt content	0.00	82.00	40.12	40.00	17.92	0.166	0.077	0.204	0.094
Soil sand content	0.00	98.00	36.74	34.00	19.85	0.167	0.274	0.189	0.096
Soil-sedimentary thickness	0.00	50.00	18.36	5.00	20.79	0.066	0.017	0.068	0.030
Average rooting depth	0.40	4.50	1.38	1.20	0.51	0.067	0.186	0.106	0.062
WRB soil classes	-	-	-	-	-	0.996	0.563	1.000	1.000
Parent material classes	-	-	-	-	-	0.146	0.177	0.112	0.044
Diurnal temperature range	4.99	21.00	12.87	12.43	2.17	0.241	0.263	0.304	0.135
Precipitation	0.84	4,475.21	897.53	966.85	423.09	0.201	0.240	0.291	0.145
Average temperature	-5.17	29.76	14.06	13.47	5.84	0.107	0.101	0.247	0.112
Maximum temperature	2.16	37.19	20.52	19.77	5.80	0.123	0.158	0.252	0.100
Minimum temperature	-12.64	24.13	7.64	7.29	6.09	0.117	0.092	0.201	0.077
Root-zone soil moisture	0.00	510.98	68.98	65.17	57.09	0.209	0.094	0.263	0.117
PDSI	-6.37	7.64	0.32	0.26	1.47	0.387	0.024	0.271	0.101
Soil surface moisture (2-5 cm)	0.04	0.39	0.24	0.25	0.06	0.287	0.207	0.300	0.161
Evaporative stress factor	0.00	1.00	0.77	0.86	0.20	0.231	0.314	0.169	0.073
EVI2	-0.0737	0.5721	0.2849	0.3035	0.0962	0.284	0.240	0.230	0.112
NDVI	-0.2149	0.8877	0.4843	0.5167	0.1610	0.187	0.099	0.165	0.078
FAPAR	0.0005	0.8560	0.4059	0.4272	0.1327	0.191	1.000	0.404	0.215
LAI	0.0009	5.6356	1.2700	1.2189	0.7138	0.160	0.065	0.229	0.128
Wind speed	0.84	7.13	3.14	3.08	0.81	0.293	0.118	0.317	0.143
Soil surface (skin) temperature	263.94	306.81	287.46	286.73	6.29	0.080	0.208	0.290	0.127
Soil's layer one temperature	270.27	307.15	288.00	286.61	5.93	0.061	0.187	0.271	0.122
Soil's layer two temperature	270.01	306.84	287.91	286.52	5.88	0.030	0.420	0.359	0.150
Soil's layer three temperature	269.92	306.68	287.90	286.52	5.86	0.005	0.396	0.348	0.169
Soil's layer four temperature	269.79	306.35	287.90	286.57	5.85	0.034	0.593	0.334	0.160
Potential evapotranspiration	330.80	2,450.19	1,205.92	1,148.09	280.03	0.187	0.116	0.338	0.164
Water deficit	0.00	2,435.30	503.90	285.81	442.76	0.201	0.343	0.264	0.110
Actual evapotranspiration	0.00	1,807.37	704.41	750.90	274.72	0.287	0.011	0.279	0.145
Land cover classes	-	-	-	-	-	0.306	0.057	0.178	0.095

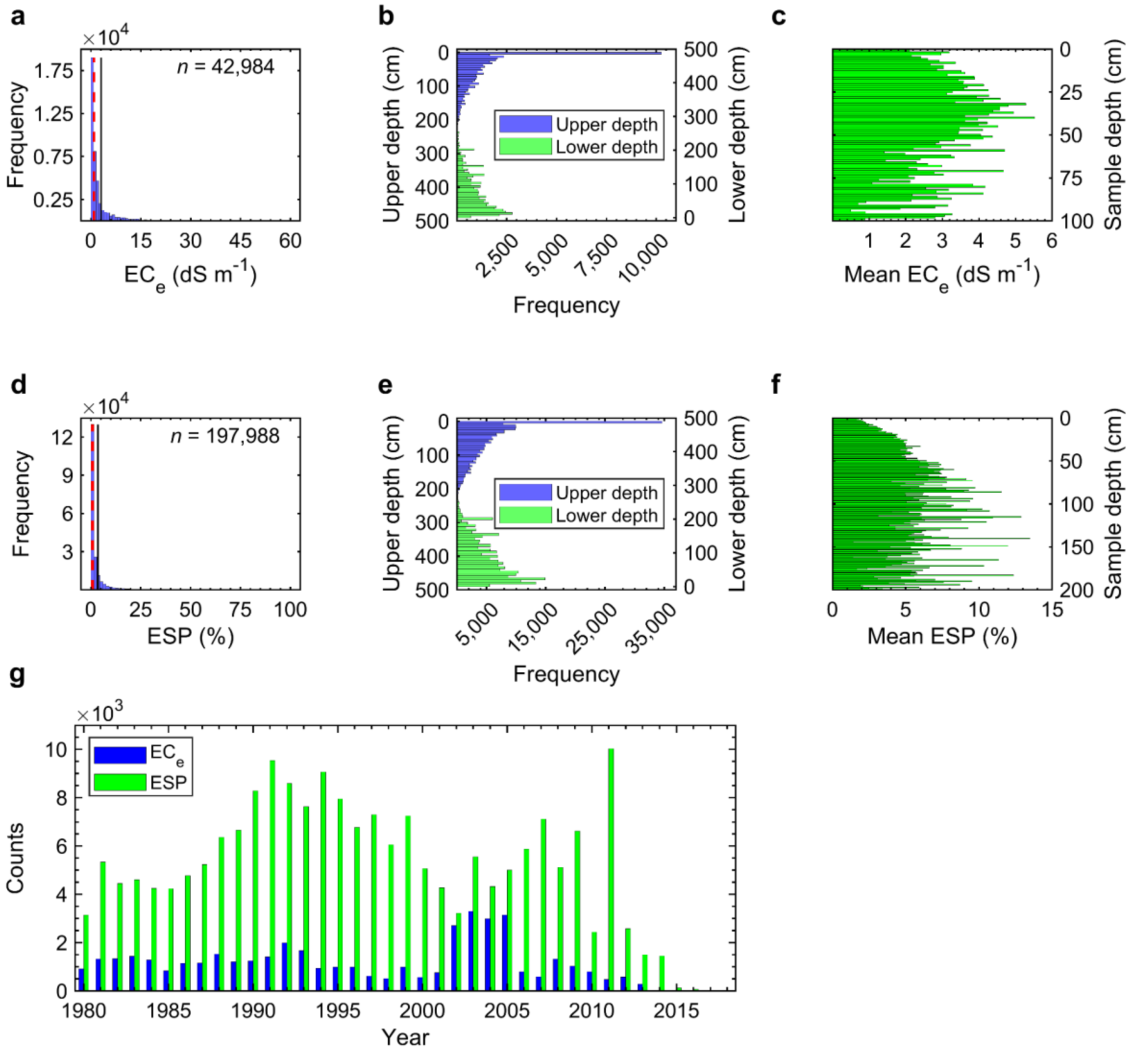


Figure A2-18: Frequency distribution of the input training data. a and **d**: Histograms of the measured values of EC_e and ESP used for training the two-part models, respectively. Black and red dotted lines represent the mean and median of data, respectively. **b**: Histograms of the lower and upper depths for measured EC_e samples used in the training set. **c**, Frequency distribution of the measured EC_e values versus depth. **e**: Histograms of the lower and upper depths for measured ESP samples used in the training set. **f**: Frequency distribution of the measured ESP values versus depth. **g**: Number of measured samples per year used in training process of the two-part models.

Table A2-6: Speed, interpretability, and flexibility of MATLAB built-in classification and regression ML algorithms for training different parts of the two-part predictive models. Information on the interpretability and flexibility are adopted from the MATALB ML toolbox user guide.

Classifier		EC_e classification		ESP classification		Interpretability	Flexibility
		Accuracy (%)	Training time (s)	Accuracy (%)	Training time (s)		
Tree	Fine	74.3	9	68.8	15	Easy	High
	Medium	72.1	8	66.5	13	Easy	Medium
	Coarse	67.7	7	64.0	10	Easy	Low
Logistic Regression		72.1	86	68.2	480	Easy	Low
Naive Bayes	Gaussian Naive	56.6	9	64.3	12	Easy	Low
	Kernel Naive	71.2	487	64.9	8,240	Easy	Medium
Support Vector Machines	Linear	71.2	406	67.8	13,277	Easy	Low
	Quadratic	80.1	838	73.4	58,762	Hard	Medium
	Cubic	85.7	1,897	78.6	77,690	Hard	Medium
	Fine Gaussian	87.0	789	83.8	19,325	Hard	High
	Medium Gaussian	81.2	283	76.2	13,408	Hard	Medium
	Coarse Gaussian	72.4	315	69.2	10,044	Hard	Low
Ensemble	Boosted Trees	75.2	35	69.6	160	Hard	Medium to high
	Bagged Trees	88.6	31	84.2	176	Hard	High
	RUSBoosted Trees	69.6	31	66.4	204	Hard	Medium

Regression Model		EC_e Regression				ESP Regression				Interpretability	Flexibility
		RMSE	*NSE*	*MAE*	Training time (s)	*RMSE*	*NSE*	*MAE*	Training time (s)		
Linear	Linear	0.30	0.33	0.24	11	0.39	0.25	0.31	16	Easy	Very low
	Interaction	53.83	< 0	6.13	445	0.49	< 0	0.27	4,597	Easy	Low
	Robust	0.30	0.32	0.23	24	0.39	0.24	0.31	149	Easy	Very low
	Stepwise	Not completed in 24 hours				Not completed in 24 hours				Easy	Low
Tree	Fine	0.25	0.53	0.17	7	0.34	0.44	0.24	106	Easy	High
	Medium	0.26	0.51	0.18	6	0.34	0.44	0.25	18	Easy	Medium
	Coarse	0.27	0.45	0.20	5	0.35	0.41	0.26	11	Easy	Low
Support Vector Machines	Linear	0.31	0.29	0.24	57	0.40	0.22	0.30	2,627	Easy	Low
	Quadratic	0.26	0.49	0.19	133	0.36	0.37	0.27	7,905	Hard	Medium
	Cubic	0.27	0.45	0.17	415	0.34	0.42	0.23	27,237	Hard	Medium
	Fine Gaussian	0.22	0.65	0.15	62	0.27	0.65	0.19	3,711	Hard	High
	Medium Gaussian	0.25	0.53	0.18	51	0.33	0.47	0.24	2,495	Hard	Medium
	Coarse Gaussian	0.31	0.27	0.24	47	0.38	0.28	0.29	2,182	Hard	Low
Ensemble	Boosted Trees	0.28	0.41	0.22	11	0.37	0.31	0.29	53	Hard	Medium to high
	Bagged Trees	0.22	0.64	0.16	17	0.27	0.63	0.20	104	Hard	High
Gaussian	Squared Exponential	0.23	0.61	0.17	2,155	Not completed in 24 hours				Hard	Automatic
	Matern 5/2	0.24	0.57	0.15	26,648	Ditto				Hard	Automatic
	Exponential	0.20	0.69	0.14	3,393	Ditto				Hard	Automatic
	Rational Quadratic	0.20	0.70	0.14	3,319	Ditto				Hard	Automatic

Table A2-7: Tuned hyperparameters and accuracy metrics for 30 classification models fitted to the EC_e training set. The model with best *MCC* (Mathew's Correlation Coefficient) was chosen for using in the final predictive two-part model of salinity.

EC_e classification												
No.	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	**Binomial deviance loss**	**Classification error**	**Accuracy (%)**	**Precision**	**Recall**	***MCC***	***MOF***
1	23	0.266	3	1,703	-	0.192	0.117	89.298	0.925	0.914	0.761	0.114
2	98	0.385	71	825	-	0.187	0.116	89.545	0.925	0.918	0.766	0.114
3	107	-	1	25,989	10	0.226	0.117	88.814	0.931	0.899	0.754	0.117
4	11	0.532	4	1,256	-	0.203	0.128	87.954	0.921	0.896	0.733	0.120
5	10	0.008	7	3,541	-	0.306	0.138	85.881	0.931	0.852	0.701	0.152
6	21	0.963	1	1,020	-	0.191	0.125	88.949	0.916	0.918	0.751	0.119
7	18	-	2	26,939	5	0.227	0.122	88.177	0.929	0.890	0.741	0.122
8	21	0.216	1	16,529	-	0.190	0.122	89.352	0.917	0.924	0.760	0.117
9	96	0.477	7	18,762	-	0.135	0.121	89.082	0.920	0.915	0.755	0.115
10	99	-	1	2,338	10	0.227	0.116	88.763	0.934	0.895	0.754	0.117
11	44	0.077	4	32,789	-	0.193	0.120	89.287	0.920	0.919	0.759	0.114
12	44	0.607	6	22,169	-	0.190	0.121	89.471	0.916	0.926	0.762	0.110
13	50	0.474	21	2,664	-	0.187	0.117	89.650	0.921	0.924	0.767	0.109
14	15	-	3	27,232	5	0.227	0.123	88.086	0.930	0.888	0.740	0.131
15	74	0.028	6	24,562	-	0.196	0.121	88.424	0.929	0.895	0.746	0.120
16	67	-	1	38,908	3	0.227	0.118	88.579	0.933	0.893	0.750	0.124
17	19	0.703	11	24,248	-	0.192	0.126	88.842	0.915	0.918	0.749	0.120
18	34	-	1	17,420	3	0.226	0.118	88.582	0.932	0.894	0.750	0.118
19	12	-	2	27,493	4	0.227	0.128	87.621	0.926	0.885	0.729	0.130
20	39	0.444	5	684	-	0.196	0.118	88.945	0.928	0.904	0.755	0.115
21	50	0.191	16	11,886	-	0.189	0.119	89.352	0.922	0.918	0.761	0.113
22	64	0.281	180	1,308	-	0.284	0.117	88.128	0.940	0.877	0.745	0.130
23	63	-	1	2,615	30	0.226	0.117	88.849	0.931	0.899	0.755	0.117
24	44	-	2	8,278	11	0.226	0.119	88.438	0.932	0.892	0.747	0.123
25	89	0.904	2	154	-	0.199	0.121	88.400	0.930	0.893	0.746	0.114
26	12	0.640	13	8,186	-	0.195	0.129	88.293	0.916	0.907	0.738	0.121
27	10	0.422	1	22,855	-	0.193	0.126	88.854	0.916	0.917	0.749	0.120
28	12	0.764	1	36,342	-	0.194	0.128	88.877	0.912	0.922	0.749	0.121
29	45	0.481	4	2,039	-	0.136	0.124	88.812	0.919	0.913	0.749	0.116
30	41	0.142	1	2,261	-	0.145	0.119	89.026	0.925	0.909	0.756	0.115

Table A2-8: Tuned hyperparameters and accuracy metrics for 30 regression models fitted to the non-saline class (EC$_e$ < 2 dS m^{-1}). The model with best *NSE* (Nash-Sutcliffe model efficiency coefficient) was chosen for using in the final predictive two-part model of salinity.

EC$_e$ regression non-saline												
No.	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	***RMSE* (log)**	***MAE* (log)**	***NSE* (log)**	***RMSE***	***MAE***	***NSE***	***MOF*** × 10^4
1	50	0.094	19	13,701	9	0.069	0.049	0.718	0.294	0.193	0.647	50.85
2	110	0.034	1	7,603	10	0.070	0.047	0.714	0.296	0.188	0.644	50.35
3	206	0.053	9	18,769	7	0.069	0.048	0.720	0.292	0.191	0.652	48.93
4	29	0.224	10	20,978	16	0.070	0.048	0.710	0.297	0.191	0.642	50.97
5	92	0.077	2	862	39	0.068	0.047	0.725	0.290	0.187	0.658	49.39
6	60	0.206	3	701	42	0.068	0.047	0.727	0.289	0.188	0.659	50.13
7	155	0.035	24	6,852	19	0.068	0.048	0.727	0.290	0.188	0.657	48.24
8	66	0.067	1	18,230	9	0.069	0.047	0.718	0.295	0.186	0.646	49.72
9	28	0.134	6	9,724	8	0.069	0.048	0.720	0.294	0.189	0.648	50.03
10	59	0.068	4	11,243	7	0.069	0.046	0.723	0.292	0.185	0.653	48.35
11	21	0.196	1	18,079	5	0.074	0.050	0.680	0.310	0.198	0.609	51.51
12	127	0.035	27	2,023	18	0.069	0.049	0.718	0.295	0.193	0.646	50.44
13	56	0.066	3	23,331	8	0.069	0.046	0.723	0.292	0.184	0.652	49.12
14	28	0.133	16	23,402	17	0.070	0.048	0.711	0.296	0.190	0.642	49.37
15	71	0.063	6	17,945	10	0.068	0.046	0.725	0.291	0.185	0.654	49.60
16	58	0.114	1	517	16	0.071	0.050	0.705	0.298	0.196	0.637	49.54
17	33	0.151	2	18,453	8	0.070	0.050	0.712	0.297	0.198	0.640	48.50
18	45	0.077	5	1,993	26	0.069	0.047	0.719	0.294	0.187	0.647	48.86
19	180	0.070	31	21,983	11	0.069	0.048	0.723	0.292	0.190	0.652	48.16
20	80	0.072	8	23,408	8	0.068	0.047	0.727	0.289	0.186	0.659	47.63
21	50	0.088	2	21,229	16	0.070	0.047	0.709	0.298	0.188	0.637	51.21
22	44	0.242	14	242	20	0.069	0.047	0.724	0.291	0.188	0.656	50.46
23	80	0.037	6	10,144	15	0.069	0.047	0.718	0.297	0.187	0.640	50.11
24	110	0.154	1	18,459	4	0.071	0.047	0.707	0.299	0.189	0.636	51.77
25	109	0.101	10	7,701	14	0.069	0.047	0.722	0.291	0.187	0.654	48.91
26	22	0.201	7	6,285	31	0.071	0.048	0.706	0.298	0.192	0.637	51.01
27	176	0.091	1	19,052	6	0.070	0.046	0.715	0.294	0.185	0.647	51.41
28	84	0.047	13	4,221	5	0.071	0.051	0.703	0.300	0.201	0.632	50.12
29	90	0.118	31	553	12	0.071	0.048	0.703	0.304	0.193	0.624	50.92
30	18	0.209	4	25,273	6	0.070	0.048	0.713	0.297	0.191	0.640	51.00

Table A2-9: Tuned hyperparameters and accuracy metrics for 30 regression models fitted to the saline class ($EC_e \geq 2$ dS m^{-1}). The model with best *NSE* was chosen for using in the final predictive two-part model of salinity.

EC_e Regression saline												
No.	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	***RMSE* (log)**	***MAE* (log)**	***NSE* (log)**	***RMSE***	***MAE***	***NSE***	***MOF* × 10^4**
1	144	0.235	8	7,380	1	0.193	0.134	0.728	5.223	2.560	0.713	391.75
2	366	0.046	3	9,365	1	0.187	0.127	0.747	5.119	2.451	0.725	367.26
3	59	0.075	6	8,900	9	0.191	0.131	0.735	5.423	2.551	0.691	384.33
4	140	0.043	7	14,179	3	0.188	0.129	0.742	5.299	2.526	0.705	364.09
5	229	0.042	1	1,812	1	0.189	0.128	0.741	5.168	2.472	0.719	360.28
6	107	0.051	5	903	26	0.189	0.129	0.741	5.214	2.507	0.714	369.81
7	52	0.153	11	12,602	3	0.192	0.133	0.733	5.372	2.586	0.697	387.43
8	152	0.145	13	219	18	0.189	0.132	0.741	5.172	2.522	0.719	370.23
9	103	0.068	7	6,004	2	0.189	0.129	0.741	5.231	2.508	0.713	373.13
10	128	0.040	1	6,038	12	0.189	0.126	0.741	5.131	2.428	0.724	379.53
11	45	0.099	7	13,238	9	0.192	0.132	0.731	5.436	2.579	0.690	385.47
12	143	0.187	38	9,795	14	0.190	0.132	0.736	5.245	2.549	0.711	379.96
13	114	0.045	1	816	27	0.188	0.130	0.743	5.195	2.514	0.717	365.84
14	110	0.092	2	367	32	0.190	0.131	0.738	5.247	2.529	0.711	378.18
15	113	0.035	4	4,955	20	0.190	0.128	0.738	5.246	2.491	0.711	373.13
16	83	0.061	4	10,608	22	0.188	0.126	0.743	5.177	2.453	0.718	365.07
17	84	0.052	1	9,323	13	0.190	0.129	0.738	5.263	2.486	0.709	365.50
18	33	0.157	1	11,757	4	0.192	0.129	0.733	5.242	2.502	0.711	377.85
19	86	0.105	9	11,060	12	0.195	0.131	0.722	5.265	2.521	0.709	368.75
20	44	0.137	5	11,099	2	0.191	0.131	0.735	5.266	2.534	0.709	380.64
21	359	0.196	1	1,111	3	0.191	0.128	0.735	5.204	2.466	0.716	375.64
22	168	0.027	5	2,205	11	0.186	0.126	0.748	5.136	2.445	0.723	365.27
23	132	0.124	2	478	32	0.189	0.130	0.741	5.182	2.494	0.718	370.60
24	257	0.034	17	2,040	10	0.189	0.130	0.740	5.326	2.541	0.702	359.86
25	90	0.123	1	13,140	5	0.191	0.127	0.736	5.178	2.439	0.718	384.88
26	261	0.025	10	13,856	4	0.189	0.130	0.741	5.323	2.536	0.702	362.90
27	35	0.213	1	202	25	0.205	0.145	0.695	5.631	2.771	0.667	405.56
28	55	0.088	2	10,499	11	0.190	0.127	0.737	5.136	2.440	0.723	377.69
29	248	0.140	13	78	16	0.198	0.132	0.714	5.539	2.597	0.678	378.02
30	37	0.166	6	13,604	5	0.195	0.133	0.723	5.498	2.602	0.682	372.97

Table A2-10: Tuned hyperparameters and accuracy metrics for 30 classification models fitted to the ESP training set. The model with best *MCC* (Mathew's Correlation Coefficient) was chosen for using in the final predictive two-part model of sodicity.

ESP classification												
No.	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	**Binomial deviance loss**	**Classification error**	**Accuracy (%)**	**Precision**	**Recall**	***MCC***	***MOF***
1	28	0.231	1	9,436	-	0.209	0.149	85.065	0.853	0.882	0.697	0.154
2	94	-	1	40,194	11	0.229	0.146	85.390	0.855	0.886	0.704	0.155
3	65	-	1	186,555	5	0.229	0.150	85.007	0.850	0.885	0.696	0.154
4	175	-	4	53,692	2	0.230	0.148	85.249	0.853	0.886	0.701	0.152
5	50	-	1	98,133	3	0.229	0.147	85.274	0.855	0.884	0.702	0.152
6	112	-	2	60,487	2	0.230	0.148	85.240	0.854	0.883	0.701	0.153
7	74	-	1	190,915	2	0.229	0.146	85.411	0.857	0.884	0.704	0.151
8	50	-	1	197,849	2	0.229	0.146	85.379	0.856	0.883	0.704	0.150
9	16	-	1	194,982	4	0.229	0.153	84.736	0.850	0.879	0.691	0.159
10	208	-	1	80,815	2	0.229	0.144	85.593	0.859	0.885	0.708	0.149
11	77	-	1	113,132	2	0.229	0.150	85.034	0.852	0.883	0.697	0.151
12	29	-	1	119,019	4	0.229	0.149	85.077	0.851	0.884	0.698	0.154
13	52	-	2	185,249	2	0.230	0.148	85.233	0.855	0.883	0.701	0.155
14	82	-	1	123,708	2	0.229	0.147	85.331	0.857	0.882	0.703	0.155
15	26	-	2	71,302	2	0.230	0.151	84.938	0.852	0.880	0.695	0.156
16	225	-	2	102,714	2	0.229	0.149	85.114	0.853	0.882	0.698	0.151
17	48	-	1	118,810	12	0.229	0.148	85.156	0.853	0.884	0.699	0.157
18	155	-	1	51,096	14	0.229	0.145	85.509	0.856	0.887	0.706	0.151
19	80	-	1	64,011	3	0.229	0.147	85.329	0.854	0.886	0.703	0.150
20	155	-	2	14,430	4	0.230	0.149	85.119	0.850	0.887	0.698	0.154
21	18	-	2	169,555	3	0.229	0.151	84.948	0.851	0.882	0.695	0.156
22	72	-	2	149,721	2	0.229	0.146	85.363	0.855	0.884	0.703	0.150
23	111	-	2	51,520	2	0.229	0.146	85.394	0.855	0.885	0.704	0.153
24	25	-	1	29,331	11	0.229	0.149	85.110	0.853	0.883	0.698	0.157
25	164	-	2	145,917	4	0.229	0.147	85.343	0.853	0.887	0.703	0.154
26	209	-	1	30,842	4	0.229	0.148	85.163	0.852	0.886	0.699	0.151
27	180	-	1	34,111	4	0.229	0.146	85.391	0.855	0.886	0.704	0.150
28	48	-	1	67,923	20	0.231	0.159	84.147	0.837	0.885	0.679	0.158
29	53	-	1	92,899	4	0.229	0.151	84.859	0.853	0.878	0.693	0.155
30	211	-	2	138,786	2	0.229	0.148	85.177	0.852	0.885	0.700	0.151

Table A2-11: Tuned hyperparameters and accuracy metrics for 30 regression models fitted to the non-sodic class (ESP < 1%). The model with best *NSE* (Nash–Sutcliffe model efficiency coefficient) was chosen for using in the final predictive two-part model of sodicity.

ESP regression non-sodic												
No.	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	***RMSE* (log)**	***MAE* (log)**	***NSE* (log)**	***RMSE***	***MAE***	***NSE***	***MOF*** $\times 10^4$
1	196	0.058	27	7,078	3	0.071	0.045	0.561	0.226	0.141	0.536	51.95
2	210	0.052	6	34,502	1	0.071	0.046	0.558	0.226	0.143	0.532	51.22
3	69	0.120	12	23,081	2	0.072	0.047	0.549	0.229	0.145	0.523	52.44
4	265	0.031	6	27,557	1	0.071	0.046	0.563	0.225	0.143	0.536	51.67
5	155	0.072	4	3,473	2	0.071	0.046	0.563	0.225	0.144	0.536	51.05
6	238	0.045	17	72,003	2	0.071	0.047	0.563	0.225	0.145	0.537	51.89
7	166	0.045	16	96,243	3	0.072	0.045	0.548	0.229	0.141	0.523	51.71
8	121	0.155	10	3,881	3	0.071	0.047	0.560	0.226	0.145	0.534	51.99
9	478	0.009	1	6,981	4	0.071	0.046	0.563	0.226	0.143	0.534	51.64
10	479	0.020	18	1,906	3	0.071	0.048	0.560	0.227	0.150	0.532	51.76
11	170	0.058	5	13,936	1	0.072	0.046	0.549	0.228	0.142	0.524	51.99
12	489	0.022	13	96,009	4	0.071	0.046	0.560	0.226	0.143	0.535	51.49
13	117	0.092	4	5,232	2	0.071	0.047	0.555	0.227	0.145	0.529	52.49
14	318	0.016	5	15,194	2	0.071	0.045	0.562	0.226	0.141	0.535	52.34
15	272	0.039	16	2,251	2	0.071	0.048	0.562	0.226	0.148	0.534	52.18
16	113	0.064	20	5,851	5	0.071	0.046	0.559	0.226	0.143	0.533	52.21
17	318	0.017	6	78,130	3	0.071	0.045	0.562	0.226	0.141	0.535	51.44
18	378	0.030	12	95,924	2	0.071	0.046	0.563	0.225	0.142	0.537	51.14
19	192	0.030	9	64,674	2	0.071	0.046	0.562	0.226	0.144	0.535	51.44
20	75	0.087	5	6,384	2	0.071	0.046	0.558	0.227	0.144	0.532	52.13
21	222	0.037	1	3,599	4	0.071	0.046	0.561	0.226	0.144	0.533	52.03
22	368	0.027	1	3,022	1	0.072	0.045	0.553	0.228	0.140	0.527	52.15
23	365	0.052	5	1,245	3	0.071	0.046	0.558	0.227	0.143	0.531	51.40
24	421	0.014	4	55,809	2	0.071	0.044	0.553	0.228	0.139	0.528	51.71
25	353	0.043	5	87,892	1	0.071	0.045	0.560	0.226	0.140	0.534	52.80
26	194	0.058	9	12,642	3	0.071	0.046	0.555	0.227	0.142	0.530	52.22
27	395	0.027	1	18,392	1	0.071	0.045	0.556	0.227	0.141	0.530	52.11
28	87	0.048	6	83,758	4	0.071	0.046	0.558	0.227	0.143	0.530	51.72
29	217	0.037	9	3,935	2	0.071	0.044	0.555	0.227	0.139	0.528	51.89
30	489	0.021	8	1,968	2	0.071	0.048	0.559	0.227	0.150	0.531	52.12

Table A2-12: Tuned hyperparameters and accuracy metrics for 30 regression models fitted to the sodic class (ESP ≥ 1%). The model with best *NSE* was chosen for using in the final predictive two-part model of sodicity.

ESP regression sodic												
No.	**Number of learning cycles**	**Learn rate**	**Minimum leaf size**	**Maximum number of splits**	**Number of variables to sample**	***RMSE* (log)**	***MAE* (log)**	***NSE* (log)**	***RMSE***	***MAE***	***NSE***	***MOF***
1	121	0.055	1	58,832	2	0.232	0.160	0.743	6.832	2.668	0.721	0.057
2	172	0.102	6	12,029	2	0.230	0.159	0.745	6.784	2.628	0.725	0.056
3	394	0.053	18	77,202	4	0.229	0.159	0.749	6.979	2.674	0.709	0.055
4	144	0.117	12	86,769	4	0.236	0.165	0.733	7.093	2.798	0.700	0.057
5	295	0.031	1	36,274	2	0.231	0.158	0.744	6.772	2.616	0.726	0.055
6	465	0.012	1	82,208	2	0.231	0.160	0.744	6.960	2.707	0.711	0.055
7	262	0.023	7	20,550	4	0.233	0.163	0.740	7.312	2.780	0.681	0.057
8	218	0.042	7	11,874	3	0.231	0.160	0.745	7.091	2.702	0.700	0.056
9	302	0.028	1	23,465	2	0.231	0.160	0.745	6.866	2.689	0.719	0.055
10	333	0.019	1	21,829	3	0.231	0.160	0.743	6.947	2.664	0.712	0.056
11	116	0.066	1	8,039	3	0.232	0.161	0.742	6.995	2.695	0.708	0.056
12	347	0.035	8	58,001	2	0.227	0.157	0.753	6.795	2.634	0.725	0.054
13	75	0.058	1	17,824	2	0.236	0.165	0.732	7.109	2.755	0.698	0.058
14	96	0.097	7	6,191	2	0.232	0.162	0.741	7.130	2.728	0.697	0.056
15	33	0.222	2	15,408	2	0.234	0.159	0.738	6.780	2.649	0.726	0.059
16	444	0.086	24	73,908	2	0.229	0.159	0.749	6.852	2.667	0.720	0.055
17	117	0.033	1	72,240	2	0.234	0.162	0.738	7.121	2.713	0.697	0.057
18	481	0.019	5	70,110	2	0.227	0.158	0.752	6.817	2.634	0.723	0.055
19	218	0.062	18	86,189	2	0.231	0.162	0.744	7.125	2.780	0.697	0.055
20	125	0.121	10	15,807	1	0.235	0.165	0.736	7.182	2.829	0.692	0.057
21	317	0.041	2	6,448	3	0.230	0.158	0.747	6.929	2.644	0.714	0.056
22	144	0.029	1	62,848	4	0.233	0.163	0.739	7.172	2.776	0.693	0.058
23	309	0.108	41	72,322	4	0.234	0.165	0.738	7.239	2.843	0.687	0.056
24	66	0.115	4	53,906	1	0.233	0.162	0.739	6.945	2.714	0.712	0.058
25	182	0.065	11	81,746	3	0.233	0.163	0.739	6.996	2.756	0.708	0.056
26	295	0.171	12	1,713	4	0.234	0.162	0.736	6.981	2.688	0.709	0.058
27	374	0.015	1	58,737	4	0.235	0.162	0.736	7.009	2.687	0.707	0.057
28	188	0.131	1	2,564	3	0.232	0.160	0.742	6.892	2.688	0.717	0.056
29	228	0.104	6	62,585	1	0.231	0.160	0.744	6.794	2.666	0.725	0.057
30	210	0.057	3	53,132	2	0.230	0.158	0.746	6.794	2.619	0.725	0.055

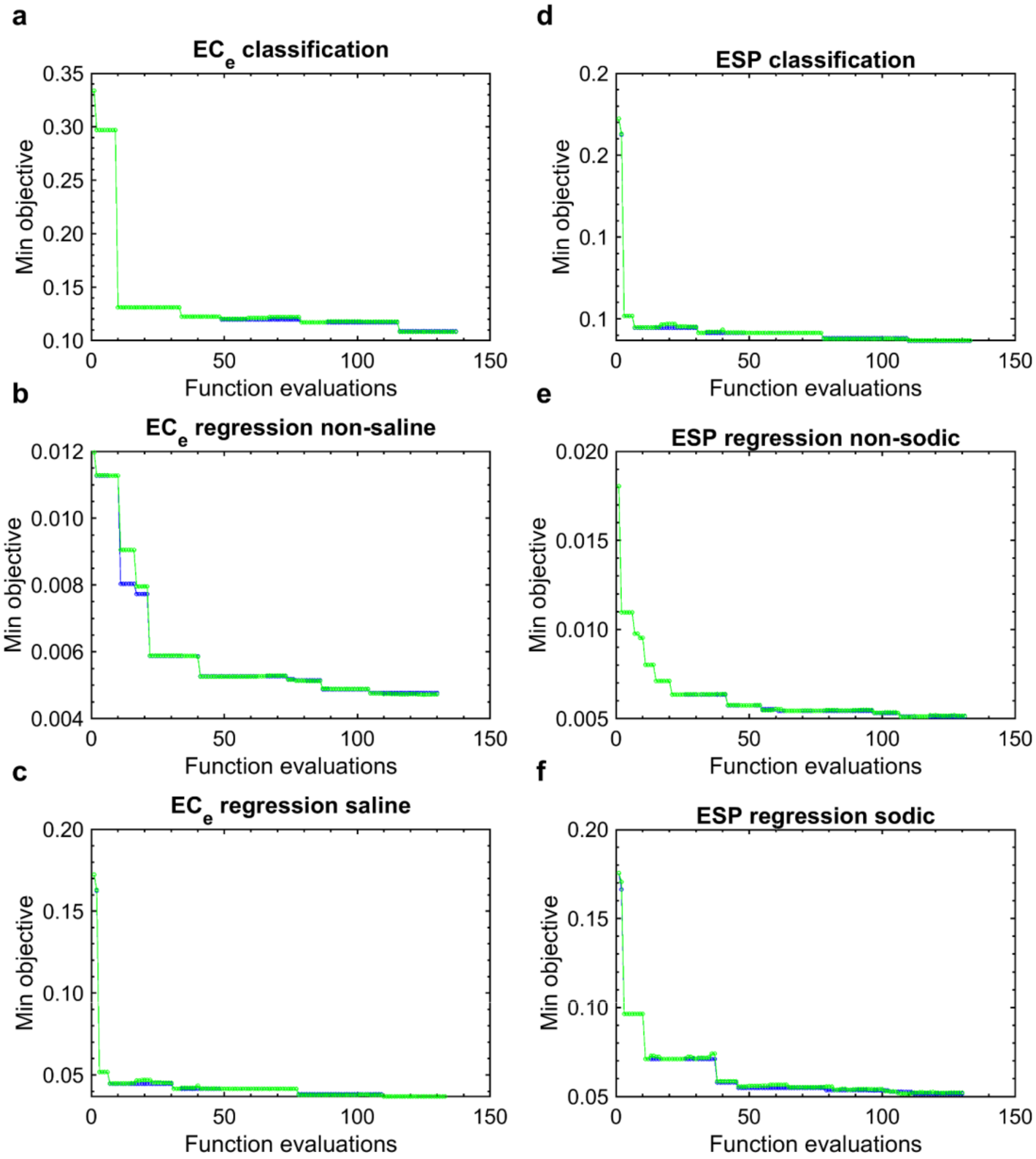


Figure A2-19: Objective function value against the iteration number during model hyperparameter tuning of the best-fitted models. The maximum number of objective function evaluations was 130. Green and blue lines show the observed and estimated objective function values, respectively.

Limitations

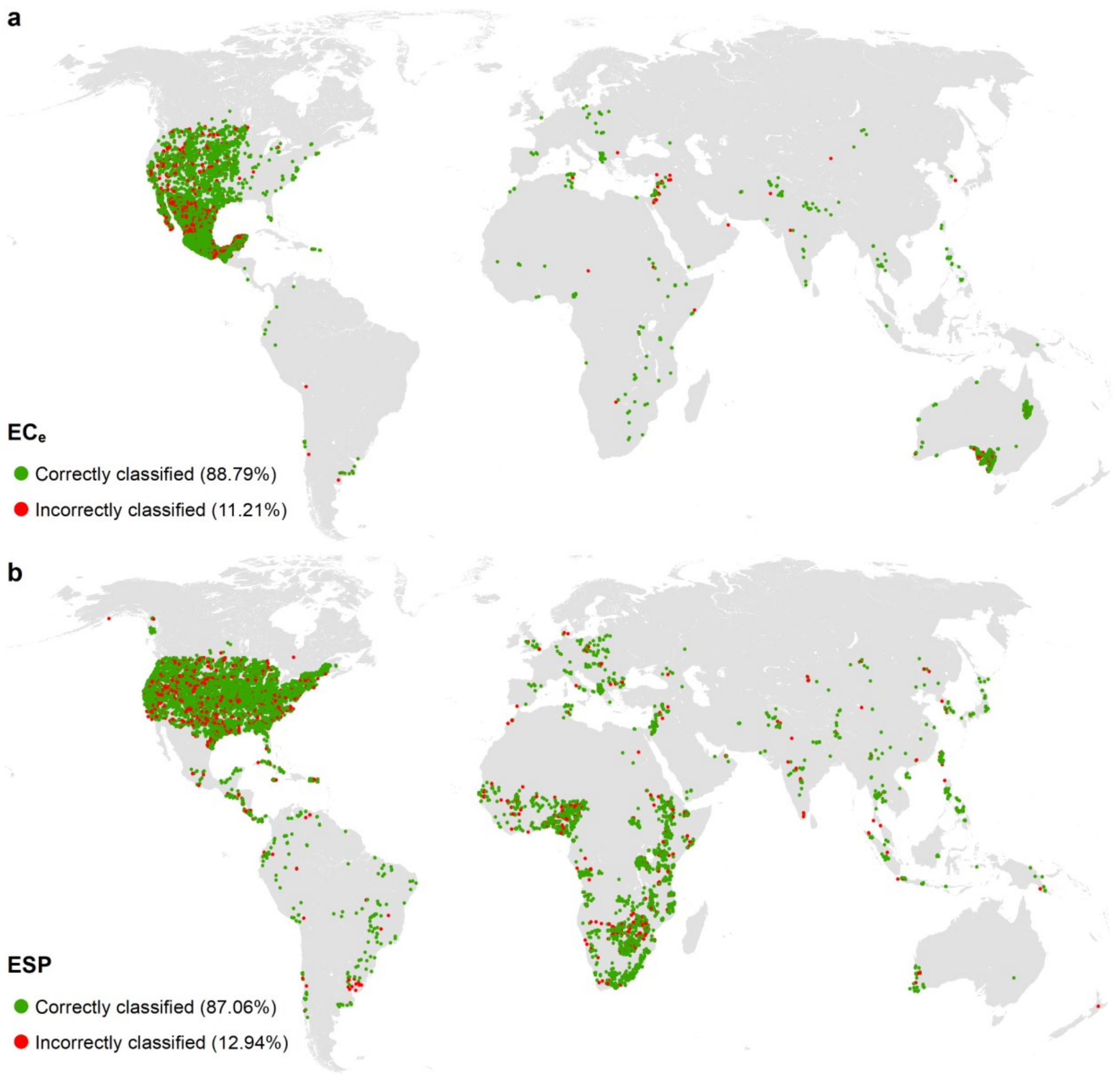


Figure A2-20: Spatial distribution of the known surface measurements classified by the two-part models. **a**: EC$_e$. **b**: ESP. Any available EC$_e$ or ESP measurement from 1980 with zero upper sample's depth and a maximum lower sample's depth equal to 30 cm was used as a known surface measurement. We categorised the known surface measurements of EC$_e$ into five classes of non-saline (0 - 2 dS m^{-1}), slightly saline (2 - 4 dS m^{-1}), moderately saline (4 - 8 dS m^{-1}), highly saline (8 - 16 dS m^{-1}), and extremely saline (> 16 dS m^{-1}) and similarly, the known surface measurements of ESP into five classes of non-sodic (0 - 1%), slightly sodic (1 - 6%), moderately sodic (6 -15%), highly sodic (16 - 30%), and extremely sodic (> 30%). The above maps are generated by comparison between the measured data classes and final predictions of the two-part models falling into each class.

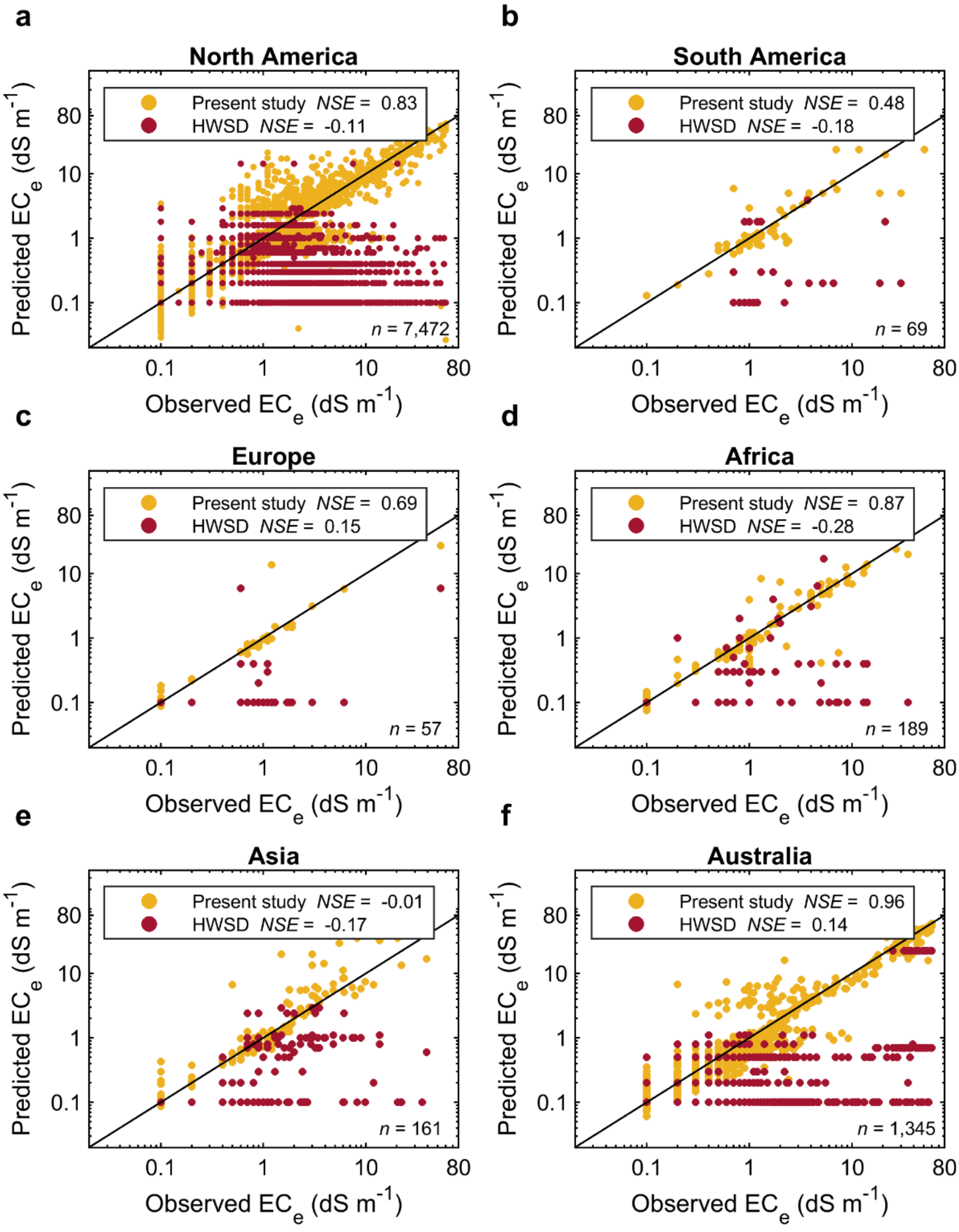


Figure A2-21: Comparison between the measured soil surface electrical conductivity of saturated-paste extract (EC_e) and the values predicted by the two-part models developed in this study as well as the values presented by Harmonised World Soil Database (HWSD) at the continent level. *NSE* = Nash-Sutcliffe model efficiency coefficient (Nash et al. 1970) ranging from -∞ to 1; *NSE* = 1 shows a perfect match.

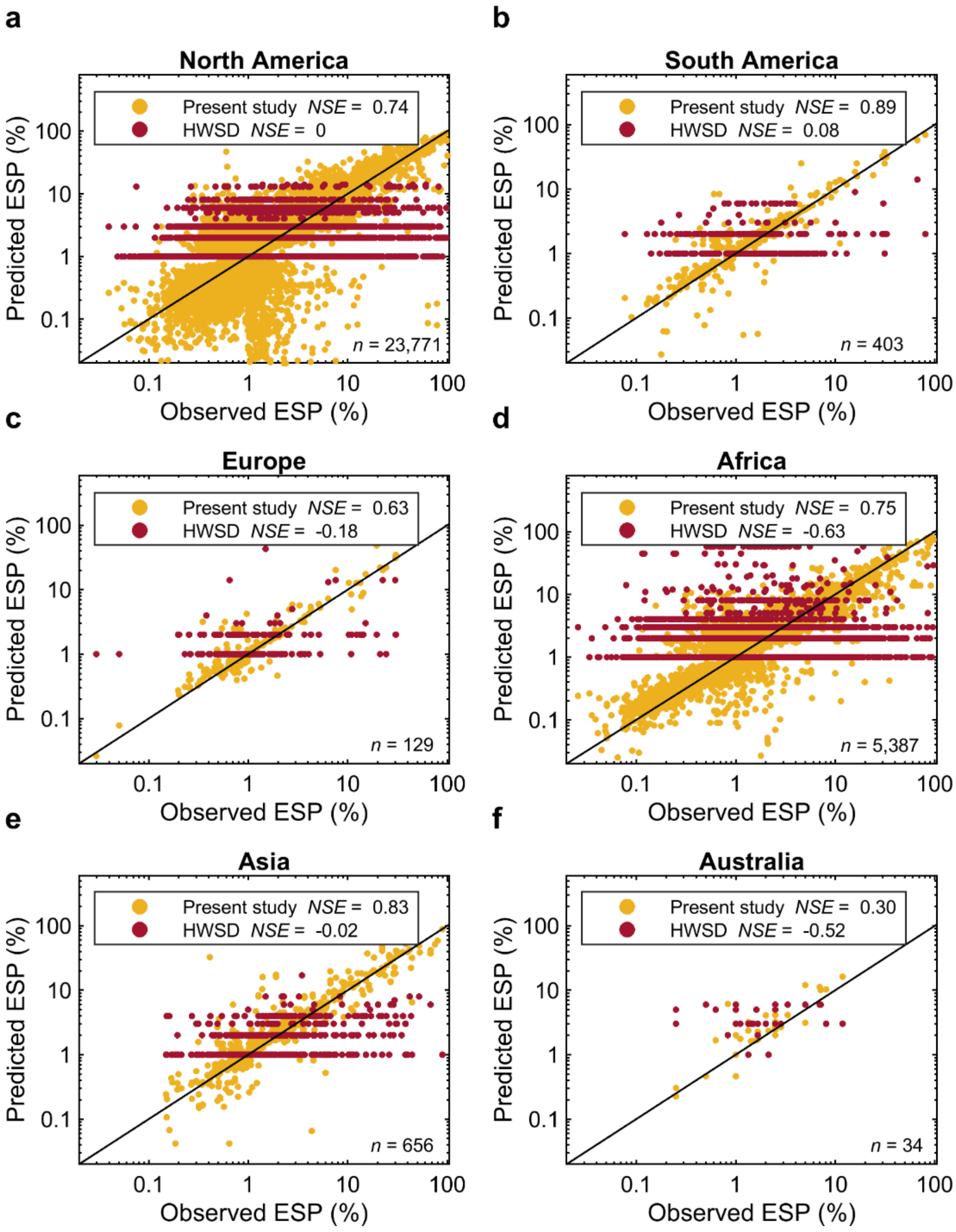


Figure A2-22: Comparison between the measured surface soil exchangeable sodium percentage (ESP) and the values predicted by the two-part models developed in this study as well as the values presented by Harmonised World Soil Database (HWSD) at the continent level. Nash-Sutcliffe model efficiency coefficient **(5)** (*NSE*, ranging from -∞ to 1; *NSE* = 1 shows a perfect match) for South America with ~1% contribution to ESP training dataset was 0.89 whereas *NSE* of North America with > 80% involvement in the training dataset was 0.74. In spite of this, predictions of surface EC_e and ESP for Asia (*NSE* = -0.01*,* number of observations = 161) and Australia (*NSE* = 0.30, umber of observations = 34) were still of the lowest certainty, respectively. The low *NSE* values might be due to the insufficient number of validating surface measurements. Overall, the performance of the models in estimation of individual continents' soil surface salinity/sodicity with an *NSE* ranging from -0.01 to 0.96 (mean for all continents = 0.63) was better than the mean *NSE* of -0.14 for HWSD predictions.

Statistics on salt-affected regions

We would assume soils of a particular location as salt-affected if the annual predicted EC_e of that location were ≥ 4 dS m^{-1} and/or its predicted ESP were ≥ 6% in at least 75% of the years between 1980 and 2018 period.

Note: All statistical analysis presented here were calculated for the regions delimited to -55° and 55° latitudes, i.e. tropics, subtropics, and temperate zones. Statistics of the countries located above 55° latitude (e.g. Canada, Russia, and United Kingdom) are not reported here.

Table A2-13: Mean cell-level likelihood of the soils with an $EC_e \geq 4$ dS m^{-1} or ESP ≥ 6% for different biomes and land cover types between 1980 and 2018. Also this Table shows statistics on the soil cell-level trends in EC_e and ESP ($p < 0.05$) for each land cover and biome in the 1980 - 2018 period.

	EC_e		**ESP**		**EC_e**	**ESP**
	Trend (× 10^5 dS m^{-1} yr^{-1})		**Trend (× 10^5 % yr^{-1})**		**Likelihood (× 10^5)**	
Land cover	**MEAN**	**STD**	**MEAN**	**STD**	**MEAN**	
Evergreen Needleleaf Forests	-193.94	1,330.98	-218.25	1,013.15	66.02	1.96
Evergreen Broadleaf Forests	1,373.16	4,812.37	440.51	2,459.06	727.24	0.52
Deciduous Needleleaf Forests	-110.53	425.51	-221.27	2,001.71	1.63	0.00
Deciduous Broadleaf Forests	-60.81	1,496.96	-213.23	665.46	32.26	0.44
Mixed Forests	-138.18	1,381.59	-204.27	982.49	51.78	0.39
Closed Shrublands	-1,366.40	2,460.76	2,151.29	3,087.65	90.90	61.96
Open Shrublands	-1,758.10	9,716.26	275.03	3,420.53	3,002.27	1,832.48
Woody Savannas	-147.79	2,023.91	-94.37	2,018.11	88.04	42.36
Savannas	87.18	1,579.73	-481.02	2,665.50	108.94	26.87
Grasslands	-1,045.83	4,600.50	-799.55	4,401.40	651.07	1,531.19
Croplands	-414.23	2,067.68	-1,439.65	3,094.71	237.19	337.82
Barren	-8.47	9,614.04	78.34	3,400.27	3,009.40	3,633.16
	EC_e		**ESP**		**EC_e**	**ESP**
	Trend (× 10^5 dS m^{-1} yr^{-1})		**Trend (× 10^5 % yr^{-1})**		**Likelihood (× 10^5)**	
Biome	**MEAN**	**STD**	**MEAN**	**STD**	**MEAN**	
Tropical and Subtropical Moist Broadleaf Forests	483.8978	4326.879	-18.63	2,438.43	518.48	7.99
Tropical and Subtropical Dry Broadleaf Forests	-1244.43	3000.977	-1,005.10	3,214.77	256.79	139.22
Temperate Broadleaf and Mixed Forests	-499.945	2495.365	-347.28	2,061.64	133.55	153.88
Tropical and Subtropical Grasslands, Savannas and Shrublands	-756.812	5115.055	-589.24	3,000.58	580.47	166.23
Temperate Grasslands, Savannas and Shrublands	-569.106	3165.172	-818.54	3,683.85	366.81	830.73
Montane Grasslands and Shrublands	-329.883	4007.414	105.22	4,965.57	574.33	2,032.18
Mangroves	-1590.22	7012.804	-577.33	2,401.83	1,672.36	332.69
Flooded Grasslands and Savannas	-2653.01	6889.474	-314.04	3,296.76	1,282.01	1,005.42
Mediterranean Forests, Woodlands and Scrub	-836.737	3866.973	122.03	2,462.69	568.10	565.74
Deserts and Xeric Shrublands	-706.209	9238.99	159.96	4,029.28	2,777.99	3,416.36
Tropical and Subtropical Coniferous Forests	-2439.09	2505.256	-191.37	2,820.89	51.70	38.13
Temperate Conifer Forests	-467.236	2588.667	-169.80	2,061.95	139.45	162.14

Table A2-14: Mean cell-level likelihood of the soils with an $EC_e \geq 4$ dS m^{-1} or ESP ≥ 6% for different climate zones between 1980 and 2018. Also this Table shows statistics on the soil cell-level trends in EC_e and ESP ($p < 0.05$) for each climate zone in the 1980 - 2018 period. For the full name of the climate zones see Figure A2-12.

Climate	EC_e		ESP		EC_e	ESP
	Trend (× 10^5 dS m^{-1} yr^{-1})		Trend (× 10^5 % yr^{-1})		Likelihood (× 10^5)	
	MEAN	STD	MEAN	STD	MEAN	
ET	111.18	4,136.40	82.27	5,413.40	324.89	2,276.64
EF	-383.11	3,905.90	-1,882.60	3,343.22	221.74	360.87
Dfc	-391.58	1,395.77	-134.66	1,078.30	61.88	28.50
Dsc	-246.93	960.50	-204.56	1,181.56	84.01	56.12
Bsk	-835.32	4,268.78	-154.16	4,532.88	801.45	1,990.51
Cfc	-228.74	890.41	-118.11	688.70	16.75	16.82
Cfb	-152.44	1,237.17	-519.27	1,683.15	45.48	29.10
Csc	-925.37	2,375.44	-496.08	2,575.24	62.62	109.31
Dfb	-350.37	1,766.04	-870.50	2,054.51	116.81	200.76
Dwc	-743.63	2,342.93	-238.56	2,051.83	96.02	85.25
Dwb	-2,356.96	4,014.65	-421.00	2,014.50	314.16	110.69
Dsb	-765.30	2,372.16	208.22	3,471.50	223.41	193.45
CSb	-816.80	2,084.19	-87.14	2,172.00	108.94	196.48
Dfa	-237.36	1,126.71	-1,415.54	2,617.47	95.77	312.66
Bwk	-439.18	8,768.37	290.37	5,891.73	3,948.60	7,330.95
Dwa	-897.44	2,407.54	-2,483.69	3,915.74	256.64	156.36
Csa	-934.47	2,594.95	288.00	2,942.60	199.09	357.49
Cfa	-126.89	2,797.89	-356.57	2,212.29	128.52	36.10
Dsa	-1,788.15	3,690.43	94.90	3,583.49	497.52	808.61
Cwa	-475.22	2,062.28	-92.62	2,217.25	109.35	74.41
BSh	-1,896.97	5,317.69	-381.05	3,308.79	713.64	531.97
BWh	-611.38	10,495.24	78.17	3,205.38	3,038.27	2,538.39
Cwb	-652.95	2,053.46	-43.37	1,633.36	79.63	31.00
Cwc	-426.70	3,692.22	26.40	2,616.76	587.85	600.27
Am	321.16	5,290.85	359.73	2,572.52	649.19	2.31
AW	-669.06	2,592.30	-367.78	3,051.78	191.40	40.76
As	-791.21	2,408.20	-9.37	2,740.20	193.97	31.36
Af	2,054.75	5,162.73	-100.90	2,341.68	867.14	1.25

Table A2-15: Statistics on the total area of soils with an EC_e or ESP between certain thresholds in the 1980 - 2018 period at the land cover level. This table also contains information about the land cover-level trends in the total area of soils with an $EC_e \geq 4$ dS m^{-1} or ESP $\geq 6\%$ since 1980 and their statistical significance (each class includes its left class edge).

EC_e

Land cover	Mean, EC_e 4 - 8 dS m^{-1} (km^2)	SD[a] EC_e 4 - 8 dS m^{-1} (km^2)	Mean, EC_e 8 - 16 dS m^{-1} (km^2)	SD, EC_e 8 - 16 dS m^{-1} (km^2)	Mean, $EC_e \geq 16$ dS m^{-1} (km^2)	SD, $EC_e \geq 16$ dS m^{-1} (km^2)	Mean of saline area, $EC_e \geq 4$ dS m^{-1} (km^2)	1980 - 2018 trend (km^2 yr^{-1})	*p*-value (1980 - 2018)
Evergreen Needleleaf Forests	4,835.6	1,149.4	80.5	22.0	28.1	17.6	4,944.2	-49.8	0.002
Evergreen Broadleaf Forests	654,492.4	229,964.5	3,858.4	5,396.0	8.6	33.5	658,359.4	10,349.0	0.001
Deciduous Needleleaf Forests	28.9	6.6	0.0	0.0	0.0	0.0	28.9	0.2	0.031
Deciduous Broadleaf Forests	2,156.9	915.2	132.3	50.8	64.2	32.4	2,353.4	-47.1	0.000
Mixed Forests	21,494.0	5,005.5	200.1	75.8	110.3	50.5	21,804.4	-184.6	0.008
Closed Shrublands	365.2	279.7	8.8	9.9	0.0	0.2	374.1	-3.5	0.388
Open Shrublands	1,244,384.5	185,061.8	490,577.4	156,118.0	87,384.3	32,279.9	1,822,346.1	-5,010.4	0.115
Woody Savannas	6,760.7	1,062.9	234.9	188.6	3.2	3.4	6,998.8	0.5	0.975
Savannas	17,339.4	4,783.1	836.0	277.1	3.4	12.9	18,178.9	73.7	0.300
Grasslands	390,667.1	62,346.5	88,496.8	16,908.6	20,122.9	7,515.4	499,286.7	-4,836.6	0.000
Croplands	148,125.4	24,137.5	14,562.3	3,294.7	350.4	367.3	163,038.0	-1,723.6	0.000
Barren	3,013,899.4	221,158.4	1,089,902.7	115,492.3	226,593.9	76,531.5	4,330,396.0	-961.0	0.796

ESP

Land cover	Mean, ESP 6 - 15% (km^2)	SD, ESP 6 - 15% (km^2)	Mean, ESP 15 - 30% (km^2)	SD, ESP 15 - 30% (km^2)	Mean, ESP $\geq$ 30% (km^2)	SD, ESP $\geq$ 30% (km^2)	Mean of sodic area, ESP $\geq 6\%$ (km^2)	1980 - 2018 trend (km^2 yr^{-1})	*p*-value (1980 - 2018)
Evergreen Needleleaf Forests	111.4	115.0	0.0	0.0	0.0	0.0	111.4	-4.6	0.004
Evergreen Broadleaf Forests	465.6	596.3	0.0	0.0	0.0	0.0	465.6	5.3	0.541
Deciduous Broadleaf Forests	35.4	41.5	0.0	0.0	0.0	0.0	35.4	-0.3	0.678
Mixed Forests	83.5	48.3	0.0	0.1	0.0	0.0	83.5	-2.3	0.000
Closed Shrublands	231.2	128.0	0.0	0.0	0.0	0.0	231.2	5.1	0.003
Open Shrublands	1,073,504.8	110,501.5	1,017.7	300.8	0.2	0.7	1,074,522.6	1,268.4	0.428
Woody Savannas	3,473.7	1,471.0	0.0	0.3	0.0	0.0	3,473.8	-52.2	0.011
Savannas	4,381.2	3,867.6	2.0	2.1	0.0	0.0	4,383.2	-103.1	0.060
Grasslands	1,032,743.3	99,680.9	5,836.9	1,797.5	1.2	2.2	1,038,581.4	-3,518.9	0.012
Croplands	228,444.5	29,256.5	1,527.1	1,140.7	0.4	1.4	229,972.0	-1,463.9	0.000
Barren	5,000,492.0	158,500.1	115,200.6	9,591.5	17.1	16.8	5,115,709.7	2,519.1	0.279

[a] Standard deviation

Table A2-16: Statistics on the total area of soils with an EC_e between certain thresholds in the 1980 - 2018 period at the biome level. This table also shows information about the biome-level trends in the total area of soils with an $EC_e \geq 4$ dS m^{-1} since 1980 and their statistical significance (each class includes its left class edge).

EC_e									
Biome	**Mean, EC_e 4 - 8 dS m^{-1} (km^2)**	**SD, EC_e 4 - 8 dS m^{-1} (km^2)**	**Mean, EC_e 8 - 16 dS m^{-1} (km^2)**	**SD, EC_e 8 - 16 dS m^{-1} (km^2)**	**Mean, $EC_e \geq 16$ dS m^{-1} (km^2)**	**SD, $EC_e \geq$ 16 dS m^{-1} (km^2)**	**Mean of saline area, $EC_e \geq 4$ dS m^{-1} (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
Deserts and Xeric Shrublands	5,402,957.2	434,427.8	1,854,798.2	284,550.9	356,676.0	116,349.0	7,614,431.4	-13,300.2	0.085
Flooded Grasslands and Savannas	85,444.4	24,048.8	34,070.8	4,060.5	4,737.1	1,330.5	124,252.3	-1,941.6	0.000
Mangroves	27,773.9	5,585.1	5,402.7	1,003.0	47.7	59.7	33,224.3	-254.6	0.002
Mediterranean Forests, Woodlands and Scrub	152,534.2	23,283.3	26,783.9	7,987.4	6,883.4	3,243.9	186,201.5	-1,315.4	0.001
Montane Grasslands and Shrublands	243,040.0	18,332.3	42,733.1	12,717.5	5,910.1	2,591.7	291,683.1	-835.2	0.011
Temperate Broadleaf and Mixed Forests	144,539.4	50,965.8	15,098.6	4,907.8	6,233.1	2,518.8	165,871.1	-4,044.7	0.000
Temperate Conifer Forests	52,836.1	11,162.5	3,965.3	1,381.7	945.9	384.3	57,747.3	-911.5	0.000
Temperate Grasslands, Savannas and Shrublands	320,498.4	50,867.7	33,101.5	9,013.1	7,243.3	2,747.7	360,843.2	-2,771.7	0.000
Tropical and Subtropical Coniferous Forests	2,867.4	789.4	408.1	235.7	9.5	12.1	3,285.0	-39.1	0.000
Tropical and Subtropical Dry Broadleaf Forests	88,770.6	23,136.6	5,546.4	1,687.3	141.8	123.8	94,458.7	-664.2	0.056
Tropical and Subtropical Grasslands, Savannas and Shrublands	696,230.8	92,651.8	368,470.1	41,665.8	48,707.3	15,603.3	1,113,408.2	-7,071.2	0.000
Tropical and Subtropical Moist Broadleaf Forests	992,815.8	254,469.2	24,965.9	14,011.8	85.2	310.3	1,017,866.8	11,638.1	0.001

Table A2-17: Statistics on the total area of soils with an ESP between certain thresholds in the 1980 - 2018 period at the biome level. This table also contains information about the biome-level trends in the total area of soils with an ESP ≥ 6% since 1980 and their statistical significance (each class includes its left class edge).

ESP									
Biome	**Mean, ESP 6 - 15% (km^2)**	**SD, ESP 6 - 15% (km^2)**	**Mean, ESP 15 - 30% (km^2)**	**SD, ESP 15 - 30% (km^2)**	**Mean, ESP ≥ 30% (km^2)**	**SD, ESP ≥ 30% (km^2)**	**Mean of sodic area, ESP ≥ 6% (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
Deserts and Xeric Shrublands	8,865,009.9	240,647.5	121,507.3	13,231.2	12.7	21.0	8,986,529.8	5,458.4	0.120
Flooded Grasslands and Savannas	97,621.6	8,122.4	1,536.0	574.5	0.0	0.0	99,157.7	-296.9	0.009
Mangroves	6,269.6	632.1	60.0	49.8	0.0	0.0	6,329.6	-8.6	0.339
Mediterranean Forests, Woodlands and Scrub	185,835.7	18,231.2	681.0	441.2	0.2	0.6	186,516.9	-115.6	0.664
Montane Grasslands and Shrublands	903,607.4	46,474.4	120,237.8	9,995.9	18.9	18.4	1,023,864.0	1,271.3	0.082
Temperate Broadleaf and Mixed Forests	183,566.4	21,192.3	3,317.3	1,608.2	0.0	0.1	186,883.7	1,037.7	0.000
Temperate Conifer Forests	63,595.0	4,002.7	1,436.9	436.8	0.2	0.6	65,032.0	191.1	0.000
Temperate Grasslands, Savannas and Shrublands	759,323.5	114,238.6	690.8	391.3	0.6	1.5	760,015.0	1,986.6	0.227
Tropical and Subtropical Coniferous Forests	2,342.5	769.5	0.0	0.0	0.0	0.0	2,342.5	-33.5	0.001
Tropical and Subtropical Dry Broadleaf Forests	50,892.0	20,114.1	1.5	4.8	0.0	0.0	50,893.6	-686.0	0.014
Tropical and Subtropical Grasslands, Savannas and Shrublands	317,844.9	87,560.4	402.5	299.4	0.1	0.2	318,247.5	-4,180.4	0.000
Tropical and Subtropical Moist Broadleaf Forests	15,013.0	7,338.5	36.5	66.0	0.1	0.3	15,049.6	-98.3	0.354

Table A2-18: Statistics on the total area of soils with an EC_e between certain thresholds in the 1980 - 2018 period at the climate level. This table also shows information about the climate-level trends in the total area of soils with an $EC_e \geq 4$ dS m^{-1} since 1980 and their statistical significance (each class includes its left class edge). For the full name of the climate zones see Figure A2-12.

EC_e									
Climate	**Mean, EC_e 4 - 8 dS m^{-1} (km^2)**	**SD, EC_e 4 - 8 dS m^{-1} (km^2)**	**Mean, EC_e 8 - 16 dS m^{-1} (km^2)**	**SD, EC_e 8 - 16 dS m^{-1} (km^2)**	**Mean, $EC_e \geq 16$ dS m^{-1} (km^2)**	**SD, $EC_e \geq 16$ dS m^{-1} (km^2)**	**Mean of saline area, $EC_e \geq$ 4 dS m^{-1} (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
AW	297,081.99	74,633.23	19,561.43	5,771.02	117.71	115.47	316,761.13	-2,145.8	0.046
Af	541,186.77	196,155.54	2,035.31	3,250.37	18.66	70.64	543,240.74	9,737.9	0.000
Am	280,332.75	59,046.66	16,546.19	9,773.84	62.93	197.40	296,941.87	2,383.3	0.006
As	11,671.52	5,684.02	1,090.04	497.14	3.17	5.62	12,764.74	-254.6	0.002
BSh	409,557.19	84,311.35	216,775.80	98,817.24	30,056.33	17,035.57	656,389.32	-9,150.2	0.000
BWh	4,245,174.05	429,867.83	1,677,044.14	197,420.42	326,347.89	107,275.99	6,248,566.08	-10,717.6	0.081
Bsk	571,005.65	55,111.88	35,192.30	17,223.06	5,502.52	3,966.61	611,700.47	-2,760.9	0.002
Bwk	1,418,580.40	61,572.82	417,484.26	56,922.10	69,824.86	16,669.50	1,905,889.51	-1,276.2	0.358
CSb	14,854.89	2,464.72	494.31	870.16	54.98	256.01	15,404.18	-149.5	0.000
Cfa	75,351.50	16,775.68	8,282.03	1,729.48	2,514.85	721.54	86,148.39	-312.4	0.225
Cfb	19,047.57	3,591.70	153.90	98.44	17.14	21.44	19,218.61	-99.3	0.055
Cfc	318.11	402.30	1.03	3.02	0.04	0.16	319.18	-20.7	0.000
Csa	38,271.52	12,398.11	3,368.09	1,580.96	625.74	523.26	42,265.35	-750.3	0.000
Csc	205.03	228.71	2.12	7.83	0.44	1.58	207.60	-17.1	0.000
Cwa	42,075.57	12,256.45	110.50	99.56	0.75	4.15	42,186.82	-527.5	0.002
Cwb	12,599.48	4,065.05	214.13	356.74	38.96	110.96	12,852.56	-252.2	0.000
Cwc	2,370.27	426.20	232.13	192.56	26.29	46.03	2,628.69	-10.1	0.092
Dfa	11,833.93	3,300.65	4.43	5.75	0.21	1.30	11,838.57	41.3	0.386
Dfb	92,024.22	32,921.59	74.00	34.19	14.27	10.94	92,112.48	-2,197.0	0.000
Dfc	65,091.16	16,722.42	427.59	290.20	108.94	56.98	65,627.69	-1,246.9	0.000
Dsa	5,892.86	2,987.85	97.79	173.30	11.77	28.15	6,002.42	-171.5	0.000
Dsb	11,391.52	4,213.15	103.00	198.65	31.41	62.26	11,525.92	-210.1	0.000
Dsc	1,235.83	412.55	19.67	55.44	2.93	6.88	1,258.43	1.8	0.775
Dwa	13,587.65	9,593.01	10.32	17.22	0.00	0.00	13,597.97	-498.3	0.000
Dwb	38,645.93	41,051.29	0.79	2.53	0.02	0.12	38,646.75	-2,874.8	0.000
Dwc	18,647.40	7,578.91	885.67	187.83	22.08	33.71	19,555.15	-522.2	0.000
EF	109.29	54.77	13.92	14.72	1.37	2.51	124.58	-0.6	0.499
ET	72,479.87	9,945.22	21,437.79	5,826.41	2,996.06	1,048.90	96,913.73	253.8	0.117

Table A2-19: Statistics on the total area of soils with an ESP between certain thresholds in the 1980 - 2018 period at the climate level. This table also contains information about the climate-level trends in the total area of soils with an ESP ≥ 6% since 1980 and their statistical significance (each class includes its left class edge). For the full name of the climate zones see Figure A2-12.

ESP									
Climate	**Mean, ESP 6 - 15% (km^2)**	**SD, ESP 6 - 15% (km^2)**	**Mean, ESP 15 - 30% (km^2)**	**SD, ESP 15 - 30% (km^2)**	**Mean, ESP ≥ 30% (km^2)**	**SD, ESP ≥ 30% (km^2)**	**Mean of sodic area, ESP ≥ 6% (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
AW	66,723.39	52,393.80	34.39	75.12	0.04	0.19	66,757.81	-1,874.5	0.010
Af	783.47	544.49	0.00	0.00	0.00	0.00	783.47	10.4	0.181
Am	1,035.27	1,495.96	0.00	0.00	0.00	0.00	1,035.27	-6.6	0.763
As	2,022.62	728.97	0.00	0.00	0.00	0.00	2,022.62	3.2	0.764
BSh	477,114.91	48,531.17	2,153.58	925.53	0.29	0.82	479,268.78	-945.0	0.178
BWh	5,103,247.28	230,356.67	24,666.78	5,859.54	0.44	1.35	5,127,914.50	5,231.9	0.117
Bsk	1,446,213.25	104,661.58	10,955.36	2,601.10	9.63	18.39	1,457,178.24	785.3	0.607
Bwk	3,442,886.47	39,033.27	92,799.41	12,339.29	1.91	2.93	3,535,687.80	678.2	0.235
CSb	27,319.54	5,758.71	3.51	3.27	0.00	0.00	27,323.05	170.3	0.036
Cfa	24,097.24	4,831.52	41.32	48.22	0.03	0.15	24,138.60	61.9	0.376
Cfb	11,965.59	2,085.80	270.49	284.62	0.17	0.80	12,236.24	87.9	0.002
Cfc	314.84	138.80	0.03	0.12	0.00	0.00	314.87	3.9	0.045
Csa	75,978.28	13,442.30	146.73	75.03	0.45	2.13	76,125.46	714.9	0.000
Csc	366.19	146.14	0.00	0.00	0.00	0.00	366.19	5.5	0.006
Cwa	27,282.22	4,382.41	222.44	291.63	0.00	0.00	27,504.65	-74.4	0.243
Cwb	5,191.37	768.95	0.39	0.64	0.00	0.00	5,191.77	-14.8	0.179
Cwc	2,703.67	422.31	1.83	7.81	0.00	0.00	2,705.49	28.4	0.000
Dfa	37,722.40	11,566.32	0.51	1.74	0.00	0.00	37,722.91	-31.5	0.851
Dfb	118,230.61	42,891.06	35.32	21.50	0.01	0.09	118,265.94	-804.6	0.191
Dfc	8,092.44	2,012.67	21.38	16.93	0.00	0.00	8,113.82	-29.2	0.316
Dsa	9,946.86	2,348.92	0.05	0.34	0.00	0.00	9,946.91	15.8	0.643
Dsb	9,917.17	3,429.96	0.07	0.21	0.00	0.00	9,917.24	-190.0	0.000
Dsc	832.18	190.24	0.15	0.50	0.00	0.00	832.33	-7.1	0.007
Dwa	8,514.13	3,174.45	0.84	5.26	0.00	0.00	8,514.97	-79.8	0.077
Dwb	13,590.04	10,218.55	0.00	0.00	0.00	0.00	13,590.04	259.2	0.074
Dwc	17,752.58	3,039.78	34.44	12.33	0.00	0.00	17,787.02	-27.8	0.528
EF	195.73	193.41	0.02	0.11	0.00	0.00	195.74	-2.9	0.291
ET	544,214.49	33,185.41	119,381.63	9,766.84	19.73	19.00	663,615.85	550.3	0.329

Table A2-20: Statistics on the total area of soils with an EC_e between certain thresholds in the 1980 - 2018 period at the continent level. This table also shows information about the continent-level trends in the total area of soils with an $EC_e \geq 4$ dS m^{-1} since 1980 and 1999 and the corresponding statistical significance (each class includes its left class edge).

Continent	Africa	Asia	Australia	Europe	North America	South America
Mean , EC_e 4 - 8 dS m^{-1} (km^2)	2,128,611.23	3,694,834.77	1,209,129.26	57,763.38	243,792.83	945,111.33
SD, EC_e 4 - 8 dS m^{-1} (km^2)	244,655.94	311,957.67	217,864.27	10,122.29	50,681.79	249,556.22
Mean, EC_e 8 - 16 dS m^{-1} (km^2)	785,682.43	1,038,846.99	532,286.04	101.79	31,727.17	25,199.79
SD, EC_e 8 - 16 dS m^{-1} (km^2)	79,069.54	121,000.08	193,677.91	62.10	7,402.80	13,599.98
Mean, $EC_e \geq 16$ dS m^{-1} (km^2)	159,847.99	181,227.37	90,395.53	13.89	4,282.98	1,938.54
SD, $EC_e \geq 16$ dS m^{-1} (km^2)	46,494.72	64,174.65	35,886.55	14.01	1,750.71	3,242.45
Mean of saline area, $EC_e \geq 4$ dS m^{-1} (km^2)	3,074,141.66	4,914,909.12	1,831,810.83	57,879.06	279,802.98	972,249.67
1980 - 2018 trend (km^2 yr^{-1})	-2,724.26	-22,663.81	-4,573.79	139.97	-3,092.30	9,466.74
***p*-value (1980 - 2018)**	0.49	< 0.01	0.20	0.34	< 0.01	< 0.01
1999 - 2018 trend (km^2 yr^{-1})	23,166.34	19,933.41	18,532.56	-165.41	3,077.91	1,372.62
***p*-value (1999 - 2018)**	< 0.05	0.06	0.06	0.74	< 0.01	0.88

Table A2-21: Statistics on the total area of soils with an ESP between certain thresholds in the 1980 - 2018 period at the continent level. This table also contains information about the continent-level trends in the total area of soils with an ESP ≥ 6% since 1980 and 1999 and the corresponding statistical significance (each class includes its left class edge).

Continent	Africa	Asia	Australia	Europe	North America	South America
Mean, ESP 6 - 15% (km^2)	1,875,394.56	7,585,807.83	838,087.86	55,355.12	499,594.85	612,090.06
SD, ESP 6 - 15% (km^2)	183,184.69	170,525.82	125,561.82	18,357.54	31,844.39	56,079.12
Mean, ESP 15 - 30% (km^2)	1,715.51	231,208.24	3,211.83	0.95	13,876.67	504.01
SD, ESP 15 - 30% (km^2)	361.10	18,398.63	1,291.58	2.38	3,116.81	397.25
Mean, ESP ≥ 30% (km^2)	0.15	21.29	0.00	0.02	11.14	0.12
SD, ESP ≥ 30% (km^2)	0.43	19.64	0.00	0.10	20.96	0.76
Mean of sodic area, ESP ≥ 6% (km^2)	1,877,110.22	7,817,037.36	841,299.68	55,356.09	513,482.66	612,594.18
1980 - 2018 trend (km^2 yr^{-1})	-3,860.03	5,616.02	-485.28	-215.59	1,652.23	1,813.94
***p*-value (1980 - 2018)**	0.14	< 0.05	0.79	0.42	< 0.01	< 0.05
1999 - 2018 trend (km^2 yr^{-1})	12,990.22	15,112.59	1,978.87	1,488.74	1,135.92	6,094.53
***p*-value (1999 - 2018)**	< 0.01	< 0.05	0.71	< 0.05	0.23	< 0.01

Table A2-22: Statistics on the total area of soils with an EC_e between certain thresholds in the 1980 - 2018 period at the country level. This table also shows information about the country-level trends in the total area of soils with an $EC_e \geq 4$ dS m^{-1} since 1980 and the corresponding statistical significance (each class includes its left class edge).

Country	**Mean, EC_e 4 - 8 dS m^{-1} (km^2)**	**SD, EC_e 4 - 8 dS m^{-1} (km^2)**	**Mean, EC_e 8 - 16 dS m^{-1} (km^2)**	**SD, EC_e 8 - 16 dS m^{-1} (km^2)**	**Mean, $EC_e \geq 16$ dS m^{-1} (km^2)**	**SD, $EC_e \geq$ 16 dS m^{-1} (km^2)**	**Mean of saline area, $EC_e \geq 4$ dS m^{-1} (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
Afghanistan	154,952.93	15,910.35	26,452.15	7,514.74	4,375.77	6,893.46	185,780.85	263.2	0.296
Akrotiri and Dhekelia	3.66	1.82	0.00	0.00	0.00	0.00	3.66	0.0	0.253
Albania	135.44	59.13	25.88	10.53	0.00	0.00	161.32	-0.3	0.717
Algeria	397,251.03	60,983.19	0.00	0.00	0.00	0.00	397,251.03	984.4	0.262
Angola	7,073.69	2,709.56	3,352.82	776.92	108.71	140.61	10,535.21	137.6	0.002
Anguilla	2.94	3.45	0.29	0.48	0.00	0.00	3.23	-0.1	0.062
Antigua and Barbuda	30.53	23.57	1.67	4.17	0.08	0.52	32.28	-0.3	0.414
Argentina	157,583.30	23,789.40	4,826.04	6,284.36	592.67	2,006.17	163,002.01	-942.4	0.011
Armenia	142.59	42.68	0.22	0.55	0.00	0.00	142.81	0.1	0.813
Aruba	1.97	1.89	0.00	0.00	0.00	0.00	1.97	0.0	0.070
Australia	1,210,798.30	217,933.19	532,833.12	193,709.07	90,402.19	35,891.29	1,834,033.61	-4,613.3	0.193
Austria	63.46	55.14	0.04	0.28	0.00	0.00	63.51	-0.2	0.830
Azerbaijan	5,919.39	2,139.13	6.37	6.28	0.17	0.86	5,925.94	-93.3	0.001
Bahamas	3,528.15	695.46	739.42	425.27	22.32	26.87	4,289.89	-25.2	0.001
Bahrain	252.76	69.14	164.68	58.28	19.98	24.82	437.42	-0.5	0.178
Bangladesh	2,116.78	1,632.50	22.10	45.19	0.00	0.00	2,138.88	-119.1	0.000
Barbados	5.36	4.89	0.00	0.00	0.00	0.00	5.36	-0.2	0.002
Belarus	39.23	24.61	1.27	3.00	0.07	0.21	40.56	-1.0	0.005
Belgium	106.55	85.05	3.22	4.66	1.29	1.67	111.06	1.9	0.141
Belize	256.93	162.94	0.48	0.55	0.00	0.00	257.41	-4.9	0.033
Benin	534.80	701.34	131.37	71.01	0.07	0.41	666.24	-14.7	0.154
Bhutan	422.53	287.16	6.04	12.47	0.47	2.34	429.03	-19.2	0.000
Bolivia	68,937.29	27,095.86	1,861.11	1,063.18	300.04	216.41	71,098.43	580.0	0.143
Bonaire, Sint Eustatius and Saba	18.93	9.26	1.03	1.13	0.00	0.00	19.96	-0.5	0.000
Bosnia and Herzegovina	12.69	8.59	0.08	0.21	0.00	0.00	12.77	-0.1	0.353
Botswana	38,058.35	18,237.21	9,590.68	4,728.57	2,723.35	1,617.03	50,372.38	-980.9	0.002
Brazil	413,316.80	149,060.67	7,736.28	7,544.48	3.96	7.79	421,057.04	5,637.2	0.007
British Virgin Islands	12.70	4.70	0.04	0.18	0.00	0.00	12.74	-0.1	0.239
Brunei	1,287.17	454.55	0.07	0.23	0.00	0.00	1,287.24	3.0	0.654
Bulgaria	43.54	44.33	0.16	1.02	0.00	0.00	43.70	1.0	0.104
Burkina Faso	3,176.36	2,389.85	2,457.07	1,134.33	178.67	222.90	5,812.10	-208.6	0.000
Burundi	59.19	33.42	0.00	0.00	0.00	0.00	59.19	-2.0	0.000
Cambodia	13,036.29	8,492.24	212.87	136.97	0.00	0.00	13,249.16	-85.4	0.489
Cameroon	2,111.47	935.22	454.34	194.10	0.37	0.88	2,566.18	-58.9	0.000
Caspian Sea	540.62	139.81	16.15	4.75	2.60	2.07	559.37	0.5	0.801

Country	Mean, EC$_e$ 4 - 8 dS m^{-1} (km^2)	SD, EC$_e$ 4 - 8 dS m^{-1} (km^2)	Mean, EC$_e$ 8 - 16 dS m^{-1} (km^2)	SD, EC$_e$ 8 - 16 dS m^{-1} (km^2)	Mean, EC$_e$ $\geq$ 16 dS m^{-1} (km^2)	SD, EC$_e$ $\geq$ 16 dS m^{-1} (km^2)	Mean of saline area, EC$_e$ $\geq$ 4 dS m^{-1} (km^2)	1980 - 2018 trend (km^2 yr^{-1})	*p*-value (1980 - 2018)
Cayman Islands	30.11	16.16	0.00	0.00	0.00	0.00	30.11	-0.8	0.000
Central African Republic	222.74	209.64	1.03	3.44	0.00	0.00	223.76	-5.4	0.070
Chad	165,824.58	39,522.81	68,212.55	9,494.78	16,069.39	4,803.66	250,106.51	-906.2	0.159
Chile	28,621.56	9,396.46	1,935.81	2,765.11	231.79	716.90	30,789.16	-50.2	0.712
China	800,013.32	41,439.25	278,443.13	20,103.35	57,912.77	11,581.37	1,136,369.22	-2,354.9	0.000
Colombia	90,275.31	46,654.57	632.72	343.86	14.80	60.91	90,922.83	2,007.6	0.002
Comoros	0.97	1.19	0.00	0.00	0.00	0.00	0.97	-0.1	0.001
Costa Rica	295.66	177.75	0.13	0.81	0.00	0.00	295.79	-3.2	0.217
Croatia	90.36	49.44	0.10	0.27	0.00	0.00	90.46	0.8	0.263
Cuba	4,014.30	1,868.15	43.21	30.03	0.00	0.00	4,057.51	-34.3	0.204
Curacao	2.47	3.55	0.00	0.00	0.00	0.00	2.47	-0.2	0.000
Cyprus	0.92	1.26	0.00	0.00	0.00	0.00	0.92	0.0	0.009
Czech Republic	1.53	1.83	0.00	0.00	0.00	0.00	1.53	-0.1	0.000
Côte d'Ivoire	8,407.83	4,827.58	120.23	136.34	0.48	2.86	8,528.54	-170.5	0.012
Democratic Republic of the Congo	72,557.16	30,734.60	41.73	35.72	0.02	0.14	72,598.91	-1,356.5	0.001
Denmark	1,486.04	431.33	0.69	2.32	0.09	0.27	1,486.82	13.0	0.032
Djibouti	2,485.01	573.64	1,635.33	616.20	2.60	8.90	4,122.93	32.1	0.038
Dominica	2.88	4.90	0.00	0.00	0.00	0.00	2.88	0.1	0.069
Dominican Republic	299.18	164.98	0.23	0.45	0.00	0.00	299.41	-0.1	0.969
Ecuador	9,841.35	5,164.36	19.07	68.77	3.75	13.25	9,864.16	241.3	0.000
Egypt	73,666.16	25,364.14	26,059.94	5,561.81	4,922.35	1,673.96	104,648.45	-1,596.8	0.000
El Salvador	73.93	25.24	0.00	0.00	0.00	0.00	73.93	-1.5	0.000
Equatorial Guinea	224.86	104.54	15.44	17.37	0.02	0.14	240.33	-2.5	0.120
Eritrea	10,902.19	1,986.97	3,691.25	637.32	123.61	149.96	14,717.05	-48.1	0.141
Estonia	649.13	502.84	0.61	2.47	0.00	0.00	649.74	-28.9	0.000
Ethiopia	29,209.64	7,558.67	28,789.31	5,303.59	1,030.77	645.51	59,029.71	-160.2	0.182
Falkland Islands	0.66	1.49	0.00	0.00	0.00	0.00	0.66	-0.1	0.000
Fiji	634.93	724.46	0.84	4.57	0.02	0.13	635.78	14.2	0.173
Finland	0.02	0.10	0.00	0.00	0.00	0.00	0.02	0.0	0.074
France	789.29	376.72	7.65	7.08	0.00	0.00	796.94	-20.9	0.000
French Guiana	680.96	498.86	0.13	0.37	0.00	0.00	681.09	-21.3	0.002
French Southern Territories	0.80	1.12	0.00	0.00	0.00	0.00	0.80	-0.1	0.000
Gabon	5,241.24	2,177.61	266.08	240.99	0.00	0.00	5,507.32	36.1	0.276
Gambia	587.30	313.13	155.00	106.71	0.11	0.47	742.40	-20.4	0.000
Georgia	30.79	9.18	0.11	0.25	0.00	0.00	30.90	-0.3	0.055
Germany	890.81	315.98	11.77	11.21	3.42	4.89	906.00	-11.9	0.008
Ghana	6,059.31	2,845.48	396.23	216.14	0.00	0.00	6,455.54	-108.3	0.009
Greece	430.03	334.76	1.09	4.80	0.00	0.00	431.11	-11.6	0.013
Grenada	13.57	15.78	0.00	0.00	0.00	0.00	13.57	-0.2	0.407
Guadeloupe	19.76	23.73	0.00	0.00	0.00	0.00	19.76	0.6	0.053

Country	Mean, EC_e 4 - 8 dS m^{-1} (km^2)	SD, EC_e 4 - 8 dS m^{-1} (km^2)	Mean, EC_e 8 - 16 dS m^{-1} (km^2)	SD, EC_e 8 - 16 dS m^{-1} (km^2)	Mean, $EC_e \geq 16$ dS m^{-1} (km^2)	SD, $EC_e \geq 16$ dS m^{-1} (km^2)	Mean of saline area, $EC_e \geq 4$ dS m^{-1} (km^2)	1980 - 2018 trend (km^2 yr^{-1})	*p*-value (1980 - 2018)
Guatemala	112.41	74.59	0.00	0.00	0.00	0.00	112.41	0.4	0.685
Guernsey	0.01	0.09	0.00	0.00	0.00	0.00	0.01	0.0	0.663
Guinea	380.50	612.68	33.03	30.89	0.04	0.27	413.57	-23.7	0.007
Guinea-Bissau	215.10	244.88	147.49	93.06	0.32	1.41	362.91	-15.2	0.001
Guyana	17,005.91	9,664.04	135.84	123.17	0.00	0.00	17,141.75	-254.2	0.066
Haiti	350.79	342.50	0.14	0.49	0.00	0.00	350.94	1.5	0.760
Honduras	542.92	424.89	1.33	2.62	0.00	0.00	544.25	-21.6	0.000
Hong Kong	95.34	36.37	0.00	0.00	0.00	0.00	95.34	-1.7	0.000
Hungary	666.81	365.46	10.57	28.52	0.33	0.85	677.71	4.8	0.372
India	72,207.91	19,327.21	13,864.71	4,961.64	1,421.86	2,452.09	87,494.48	-223.9	0.523
Indonesia	78,193.87	24,840.74	104.67	290.06	6.45	23.76	78,304.99	-122.6	0.734
Iran	558,214.60	65,832.39	137,230.30	23,629.17	20,921.61	13,633.69	716,366.51	-5,737.4	0.000
Iraq	128,021.00	20,191.40	91,339.31	13,348.61	19,408.21	8,040.06	238,768.53	-1,251.4	0.000
Ireland	86.00	36.68	1.63	2.26	0.00	0.00	87.63	0.9	0.117
Isle of Man	0.75	1.20	0.00	0.00	0.00	0.00	0.75	0.0	0.016
Israel	2,369.85	449.58	1,120.20	462.79	1,143.75	233.29	4,633.80	-47.6	0.000
Italy	821.51	568.46	3.82	3.56	0.00	0.00	825.33	-19.4	0.015
Jamaica	195.48	163.23	0.04	0.18	0.00	0.00	195.52	5.3	0.020
Japan	3,580.48	438.55	24.68	14.80	0.02	0.12	3,605.18	-6.6	0.300
Jordan	29,663.91	3,676.89	10,346.28	1,883.93	2,191.31	1,140.57	42,201.50	-169.5	0.022
Kazakhstan	349,786.28	50,861.68	26,930.67	8,931.20	823.56	694.18	377,540.51	-1,839.6	0.006
Kenya	30,922.65	6,254.87	25,159.21	4,200.75	30.43	130.54	56,112.28	-480.6	0.000
Kosovo	0.36	0.71	0.00	0.00	0.00	0.00	0.36	0.0	0.016
Kuwait	8,501.79	2,302.16	4,793.23	1,727.05	1,273.16	734.82	14,568.18	-36.4	0.003
Kyrgyzstan	1,773.83	418.22	17.35	22.82	1.68	3.62	1,792.86	-18.5	0.001
Laos	3,243.74	1,990.73	95.94	113.16	0.00	0.00	3,339.68	-10.2	0.735
Latvia	460.55	135.03	0.47	1.62	0.02	0.11	461.05	-7.5	0.000
Lebanon	36.89	40.37	1.20	2.96	0.00	0.00	38.09	-2.3	0.000
Lesotho	5.68	7.86	0.00	0.00	0.00	0.00	5.68	-0.2	0.076
Liberia	1,633.51	2,087.26	391.28	219.63	0.15	0.51	2,024.95	-122.4	0.000
Libya	158,962.14	39,878.72	58,823.00	14,616.99	13,503.88	2,064.65	231,289.02	-715.1	0.240
Liechtenstein	0.32	0.32	0.00	0.00	0.00	0.00	0.32	0.0	0.000
Lithuania	56.15	52.85	0.20	0.43	0.01	0.08	56.36	-3.0	0.000
Macao	2.01	1.44	0.00	0.00	0.00	0.00	2.01	-0.1	0.000
Macedonia	9.35	10.31	0.00	0.00	0.00	0.00	9.35	0.2	0.262
Madagascar	6,005.63	3,270.22	55.71	72.74	1.75	10.55	6,063.10	-213.1	0.000
Malawi	1,029.32	1,545.31	0.02	0.13	0.00	0.00	1,029.34	-28.8	0.195
Malaysia	37,842.00	11,262.18	23.94	85.35	0.00	0.00	37,865.95	439.7	0.005
Mali	112,299.91	22,409.65	51,336.56	8,803.16	8,481.53	4,653.01	172,118.00	-79.3	0.844
Malta	0.07	0.27	0.00	0.00	0.00	0.00	0.07	0.0	0.112

Country	Mean, EC_e 4 - 8 dS m^{-1} (km^2)	SD, EC_e 4 - 8 dS m^{-1} (km^2)	Mean, EC_e 8 - 16 dS m^{-1} (km^2)	SD, EC_e 8 - 16 dS m^{-1} (km^2)	Mean, $EC_e \geq 16$ dS m^{-1} (km^2)	SD, $EC_e \geq$ 16 dS m^{-1} (km^2)	Mean of saline area, $EC_e \geq 4$ dS m^{-1} (km^2)	1980 - 2018 trend (km^2 yr^{-1})	*p*-value (1980 - 2018)
Martinique	5.90	6.41	0.00	0.00	0.00	0.00	5.90	0.0	0.706
Mauritania	163,944.59	28,457.22	75,209.51	14,302.27	16,467.33	6,110.87	255,621.43	-48.8	0.926
Mexico	43,247.71	19,897.31	7,243.72	2,246.88	643.36	718.73	51,134.79	-1,074.5	0.000
Moldova	3.93	6.04	0.00	0.00	0.00	0.00	3.93	-0.3	0.002
Monaco	0.05	0.17	0.00	0.00	0.00	0.00	0.05	0.0	0.029
Mongolia	202,353.51	14,219.52	38,371.14	8,013.43	5,914.46	1,499.60	246,639.11	-578.6	0.000
Montenegro	7.22	5.58	0.00	0.00	0.00	0.00	7.22	-0.2	0.006
Montserrat	0.84	1.09	0.06	0.39	0.00	0.00	0.90	0.0	0.412
Morocco	71,746.82	27,576.81	12,301.31	4,820.45	1,160.62	1,377.16	85,208.75	432.8	0.335
Mozambique	3,525.23	1,773.31	12.23	11.46	0.00	0.00	3,537.46	-121.0	0.000
Myanmar	31,952.33	11,258.24	2,548.16	974.57	8.53	30.34	34,509.02	43.7	0.798
Namibia	86,060.52	28,098.19	23,187.76	7,270.07	4,044.06	2,215.62	113,292.33	1,483.3	0.002
Nepal	354.40	181.89	20.08	16.05	2.00	2.83	376.49	-4.3	0.101
Netherlands	2,356.79	818.58	18.23	30.63	4.55	12.25	2,379.57	11.7	0.327
New Caledonia	981.37	581.43	0.61	3.55	0.10	0.63	982.08	15.1	0.068
New Zealand	93.79	138.33	1.05	3.39	0.00	0.00	94.83	-4.7	0.018
Nicaragua	1,154.48	912.61	2.15	4.67	0.00	0.00	1,156.63	-29.3	0.022
Niger	195,201.72	32,047.93	78,157.78	14,247.23	26,030.46	5,701.82	299,389.96	760.3	0.126
Nigeria	28,880.82	10,999.63	12,696.09	4,619.99	696.20	636.48	42,273.10	-777.5	0.000
North Korea	2,152.15	618.72	1.39	1.70	0.00	0.00	2,153.54	-47.8	0.000
Northern Cyprus	7.63	6.97	0.00	0.00	0.00	0.00	7.63	-0.2	0.060
Norway	2.92	5.43	0.66	2.11	0.04	0.12	3.61	-0.4	0.000
Oman	51,408.83	12,282.33	34,412.65	5,980.26	5,907.46	4,882.27	91,728.94	-567.0	0.014
Pakistan	191,486.76	28,171.70	48,332.79	8,503.59	4,425.30	4,761.69	244,244.85	-1,584.6	0.000
Palestine	765.01	163.43	267.32	142.09	53.27	64.44	1,085.61	-6.7	0.000
Panama	1,513.66	1,042.17	7.63	27.12	0.07	0.41	1,521.36	34.3	0.019
Papua New Guinea	22,279.96	13,841.84	738.59	779.22	1.59	6.37	23,020.14	867.8	0.000
Paraguay	9,012.33	7,101.04	0.06	0.38	0.00	0.00	9,012.39	91.5	0.372
Peru	119,258.89	70,082.12	8,003.55	1,448.13	795.30	342.29	128,057.74	2,308.3	0.019
Philippines	33,395.51	17,119.77	788.86	629.98	0.11	0.68	34,184.47	1,240.9	0.000
Poland	185.60	65.94	5.19	3.62	1.60	1.61	192.38	1.0	0.315
Portugal	231.80	164.74	7.51	5.05	2.66	1.03	241.98	5.5	0.017
Puerto Rico	153.97	144.64	0.46	1.16	0.00	0.00	154.43	7.6	0.000
Qatar	4,242.01	2,183.97	4,552.07	1,650.50	1,183.34	721.18	9,977.42	17.7	0.042
Republic of Congo	18,533.10	9,311.56	0.68	2.49	0.02	0.14	18,533.80	-142.5	0.288
Romania	295.05	279.33	0.05	0.16	0.00	0.00	295.10	-4.5	0.268
Rwanda	19.84	16.50	0.00	0.00	0.04	0.27	19.88	-0.2	0.499
Saint Kitts and Nevis	5.20	5.80	0.00	0.00	0.00	0.00	5.20	-0.1	0.506
Saint Lucia	6.62	7.86	0.00	0.00	0.00	0.00	6.62	0.1	0.208
Saint Pierre and Miquelon	0.80	1.14	0.00	0.00	0.00	0.00	0.80	-0.1	0.000

Country	**Mean, EC$_e$ 4 - 8 dS m^{-1} (km^2)**	**SD, EC$_e$ 4 - 8 dS m^{-1} (km^2)**	**Mean, EC$_e$ 8 - 16 dS m^{-1} (km^2)**	**SD, EC$_e$ 8 - 16 dS m^{-1} (km^2)**	**Mean, EC$_e$ ≥ 16 dS m^{-1} (km^2)**	**SD, EC$_e$ ≥ 16 dS m^{-1} (km^2)**	**Mean of saline area, EC$_e$ ≥ 4 dS m^{-1} (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
Saint Vincent and the Grenadines	8.13	7.55	0.00	0.00	0.00	0.00	8.13	-0.3	0.017
Saint-Barthélemy	0.02	0.13	0.00	0.00	0.00	0.00	0.02	0.0	0.253
Saint-Martin	0.44	1.09	0.00	0.00	0.00	0.00	0.44	0.0	0.754
Saudi Arabia	380,561.24	61,027.91	165,425.49	26,109.18	39,163.39	13,433.20	585,150.12	-4,399.9	0.000
Senegal	6,224.53	3,488.72	6,415.38	2,941.95	409.72	294.48	13,049.63	-469.5	0.000
Serbia	78.21	71.30	0.00	0.00	0.00	0.00	78.21	-0.2	0.865
Sierra Leone	2,535.22	4,025.45	462.75	456.58	1.15	2.79	2,999.13	-176.6	0.002
Singapore	33.26	10.74	0.42	1.78	0.00	0.00	33.68	-0.3	0.124
Sint Maarten	0.08	0.41	0.00	0.00	0.00	0.00	0.08	0.0	0.357
Slovakia	8.08	8.85	0.04	0.15	0.00	0.00	8.13	-0.2	0.172
Slovenia	35.91	13.06	0.12	0.67	0.00	0.00	36.03	-0.1	0.463
Solomon Islands	835.19	286.87	0.02	0.14	0.00	0.00	835.21	13.0	0.001
Somalia	86,377.38	12,393.26	75,367.30	9,215.17	3,182.40	2,002.27	164,927.08	-699.3	0.010
South Africa	32,483.59	5,964.61	4,372.70	3,372.51	1,095.75	1,912.33	37,952.03	10.5	0.934
South Korea	923.40	309.35	25.90	13.13	0.00	0.00	949.30	-14.6	0.001
South Sudan	617.63	529.26	18.45	20.09	0.00	0.00	636.09	-20.7	0.005
Spain	446.57	254.09	0.60	1.68	0.00	0.00	447.17	-11.7	0.001
Sri Lanka	360.67	134.20	3.36	3.00	0.00	0.00	364.03	-0.9	0.638
Sudan	170,056.91	35,272.55	55,918.49	12,620.13	10,552.34	5,361.61	236,527.74	2,294.6	0.000
Suriname	3,271.49	2,497.50	0.04	0.27	0.00	0.00	3,271.53	-30.7	0.395
Swaziland	25.71	79.05	0.00	0.00	0.00	0.00	25.71	-0.4	0.725
Sweden	129.77	99.05	0.08	0.46	0.00	0.00	129.85	-6.3	0.000
Switzerland	30.68	20.57	0.02	0.09	0.00	0.00	30.69	-1.4	0.000
Syria	51,002.45	7,452.21	22,859.12	6,685.03	7,326.31	3,134.39	81,187.89	-41.0	0.650
São Tomé and Príncipe	0.39	0.43	0.00	0.00	0.00	0.00	0.39	0.0	0.001
Taiwan	533.85	104.96	5.75	31.57	0.10	0.63	539.70	-4.3	0.017
Tajikistan	3,528.58	585.37	582.93	378.42	102.44	72.82	4,213.95	-10.5	0.307
Tanzania	3,855.78	2,684.81	20.11	25.25	0.00	0.00	3,875.89	-185.2	0.000
Thailand	34,780.67	15,173.31	13,378.82	7,584.43	19.34	29.44	48,178.82	740.3	0.011
Timor-Leste	41.90	38.71	0.00	0.00	0.00	0.00	41.90	-1.3	0.020
Togo	256.81	279.55	3.58	3.57	0.00	0.00	260.39	-13.8	0.000
Trinidad and Tobago	292.52	227.76	0.00	0.00	0.00	0.00	292.52	1.9	0.563
Tunisia	19,615.14	6,112.50	6,301.58	1,868.58	2,356.56	662.74	28,273.27	-274.8	0.015
Turkey	8,835.55	3,206.00	1,223.61	721.65	16.34	19.84	10,075.50	-278.4	0.000
Turkmenistan	144,587.05	20,064.07	50,617.06	14,590.47	3,111.81	2,263.98	198,315.93	-444.9	0.259
Turks and Caicos Islands	217.51	84.17	64.86	75.75	2.77	5.34	285.14	-2.6	0.016
Uganda	138.05	184.25	0.00	0.00	0.00	0.00	138.05	-8.8	0.000
Ukraine	427.25	137.60	0.21	0.36	0.00	0.00	427.46	-0.8	0.676
United Arab Emirates	16,964.19	4,013.65	6,615.55	2,957.21	1,708.59	1,222.72	25,288.32	-115.7	0.042
United States	160,671.09	23,163.32	24,495.28	5,692.75	3,700.50	1,180.61	188,866.87	-1,225.3	0.001

Country	**Mean, EC_e 4 - 8 $dS\ m^{-1}$ (km^2)**	**SD, EC_e 4 - 8 $dS\ m^{-1}$ (km^2)**	**Mean, EC_e 8 - 16 $dS\ m^{-1}$ (km^2)**	**SD, EC_e 8 - 16 $dS\ m^{-1}$ (km^2)**	**Mean, $EC_e \geq 16$ $dS\ m^{-1}$ (km^2)**	**SD, $EC_e \geq$ 16 $dS\ m^{-1}$ (km^2)**	**Mean of saline area, $EC_e \geq 4$ $dS\ m^{-1}$ (km^2)**	**1980 - 2018 trend ($km^2\ yr^{-1}$)**	***p*-value (1980 - 2018)**
United States Minor Outlying Islands	1.10	1.51	0.00	0.00	0.00	0.00	1.10	-0.1	0.000
Uruguay	1,402.25	511.01	40.78	38.85	0.00	0.00	1,443.03	-28.5	0.000
Uzbekistan	105,515.20	23,702.97	27,064.20	11,276.84	1,389.98	1,094.17	133,969.37	-684.4	0.136
Vanuatu	271.67	217.21	0.00	0.00	0.00	0.00	271.67	-2.9	0.352
Venezuela	28,784.86	16,664.07	255.02	149.36	2.75	14.50	29,042.64	-96.0	0.692
Vietnam	10,792.66	3,732.20	126.89	123.82	0.00	0.00	10,919.54	-11.7	0.832
Virgin Islands, U.S.	6.67	9.84	0.00	0.00	0.00	0.00	6.67	0.2	0.127
Western Sahara	102,941.68	18,916.15	41,349.66	11,398.44	11,078.28	4,953.38	155,369.61	649.8	0.022
Yemen	52,529.11	17,254.51	32,682.20	11,926.33	1,681.85	2,772.69	86,893.16	-1,878.2	0.000
Zambia	1,347.27	1,686.24	16.63	22.64	0.00	0.00	1,363.90	-58.0	0.014
Zimbabwe	2,490.70	2,749.74	4.24	6.18	0.12	0.76	2,495.06	-107.5	0.004
Åland	0.01	0.07	0.00	0.00	0.00	0.00	0.01	0.0	0.092

Table A2-23: Statistics on the total area of soils with an ESP between certain thresholds in the 1980 - 2018 period at the country level. This table also contains information about the country-level trends in the total area of soils with an ESP ≥ 6% since 1980 and the corresponding statistical significance (each class includes its left class edge).

Country	**Mean, ESP 6 - 15% (km^2)**	**SD, ESP 6 - 15% (km^2)**	**Mean, ESP 15 - 30% (km^2)**	**SD, ESP 15 - 30% (km^2)**	**Mean, ESP ≥ 30% (km^2)**	**SD, ESP ≥ 30% (km^2)**	**Mean of sodic area, ESP ≥ 6% (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
Afghanistan	216,750.94	20,860.38	3,014.32	1,254.46	0.02	0.12	219,765.28	1,483.7	0.000
Akrotiri and Dhekelia	4.50	6.24	0.00	0.00	0.00	0.00	4.50	0.1	0.519
Albania	36.54	28.84	0.12	0.63	0.00	0.00	36.65	-0.4	0.355
Algeria	626,841.62	72,092.34	0.00	0.00	0.00	0.00	626,841.62	510.1	0.625
Angola	13,647.56	5,882.03	0.02	0.13	0.00	0.00	13,647.58	-276.4	0.000
Anguilla	0.67	2.00	0.00	0.00	0.00	0.00	0.67	0.0	0.449
Antigua and Barbuda	2.78	4.39	0.00	0.00	0.00	0.00	2.78	0.0	0.454
Argentina	288,013.11	43,405.75	359.46	326.93	0.12	0.76	288,372.69	2,012.4	0.001
Armenia	50.06	30.47	0.00	0.00	0.00	0.00	50.06	-1.6	0.000
Aruba	0.21	0.60	0.00	0.00	0.00	0.00	0.21	0.0	0.785
Australia	838,355.96	125,602.81	3,211.83	1,291.58	0.00	0.00	841,567.78	-486.2	0.791
Austria	1.85	1.80	0.00	0.00	0.00	0.00	1.85	-0.1	0.000
Azerbaijan	7,027.13	588.47	0.05	0.32	0.00	0.00	7,027.18	12.0	0.156
Bahamas	373.98	129.71	0.00	0.00	0.00	0.00	373.98	-0.4	0.843
Bahrain	445.01	4.06	1.68	4.09	0.00	0.00	446.69	0.0	0.002
Bangladesh	41.33	38.04	0.00	0.00	0.00	0.00	41.33	-1.2	0.026
Barbados	0.30	1.87	0.00	0.00	0.00	0.00	0.30	0.0	0.335
Belize	0.56	1.12	0.00	0.00	0.00	0.00	0.56	0.0	0.364
Benin	1,400.40	2,554.13	0.00	0.00	0.00	0.00	1,400.40	-72.7	0.044
Bhutan	0.08	0.48	0.00	0.00	0.00	0.00	0.08	0.0	0.158
Bolivia	45,532.04	5,034.51	99.96	121.90	0.00	0.00	45,631.99	196.3	0.005
Bonaire, Sint Eustatius and Saba	14.47	15.29	0.00	0.00	0.00	0.00	14.47	-0.9	0.000
Botswana	27,785.34	13,305.02	1.17	4.28	0.02	0.13	27,786.53	-110.8	0.565
Brazil	5,724.73	5,339.75	0.00	0.00	0.00	0.00	5,724.73	-80.7	0.294
British Virgin Islands	0.54	3.00	0.00	0.00	0.00	0.00	0.54	0.0	0.249
Brunei	0.46	0.64	0.00	0.00	0.00	0.00	0.46	0.0	0.049
Bulgaria	4.41	10.36	0.28	1.73	0.02	0.10	4.70	0.0	0.932
Burkina Faso	9,474.89	9,683.42	0.02	0.13	0.00	0.00	9,474.91	-514.0	0.000
Burundi	11.18	1.99	3.26	1.85	0.07	0.23	14.51	0.0	0.981
Cambodia	1,542.09	2,256.18	0.00	0.00	0.00	0.00	1,542.09	-29.6	0.364
Cameroon	576.39	752.65	22.67	112.58	0.02	0.13	599.08	-18.1	0.111
Caspian Sea	302.26	82.10	0.47	0.39	0.00	0.00	302.73	-3.2	0.005
Cayman Islands	14.89	23.84	0.00	0.00	0.00	0.00	14.89	0.2	0.593
Central African Republic	49.56	143.54	0.00	0.00	0.00	0.00	49.56	-0.2	0.938
Chad	50,295.87	8,051.47	0.70	0.36	0.00	0.00	50,296.56	-74.0	0.525
Chile	180,306.30	10,376.65	41.13	32.93	0.00	0.00	180,347.44	492.8	0.000

Country	Mean, ESP 6 - 15% (km^2)	SD, ESP 6 - 15% (km^2)	Mean, ESP 15 - 30% (km^2)	SD, ESP 15 - 30% (km^2)	Mean, ESP ≥ 30% (km^2)	SD, ESP ≥ 30% (km^2)	Mean of sodic area, ESP ≥ 6% (km^2)	1980 - 2018 trend (km^2 yr^{-1})	*p*-value (1980 - 2018)
China	2,136,070.51	47,271.28	161,035.31	12,608.93	20.50	19.23	2,297,126.32	-715.8	0.328
Colombia	704.33	323.74	0.00	0.00	0.00	0.00	704.33	-17.2	0.000
Costa Rica	2.52	14.80	0.00	0.00	0.00	0.00	2.52	-0.1	0.668
Cuba	147.89	85.80	0.06	0.22	0.00	0.00	147.95	-1.0	0.404
Curaçao	0.96	1.82	0.00	0.00	0.00	0.00	0.96	-0.1	0.028
Cyprus	27.43	58.34	0.00	0.00	0.00	0.00	27.43	0.3	0.713
Czech Republic	1.02	4.53	0.00	0.00	0.00	0.00	1.02	0.0	0.446
Côte d'Ivoire	302.48	1,376.11	0.00	0.00	0.00	0.00	302.48	-10.4	0.601
Democratic Republic of the Congo	100.24	55.10	1.93	2.36	0.02	0.14	102.19	0.9	0.251
Djibouti	6,726.63	779.05	0.04	0.19	0.00	0.00	6,726.67	-45.5	0.000
Dominica	0.02	0.13	0.00	0.00	0.00	0.00	0.02	0.0	0.158
Dominican Republic	6.50	8.18	0.00	0.00	0.00	0.00	6.50	-0.1	0.588
Ecuador	55.23	37.43	0.00	0.00	0.00	0.00	55.23	-1.0	0.063
Egypt	52,104.65	5,142.35	86.07	15.44	0.00	0.00	52,190.72	288.2	0.000
El Salvador	7.61	18.94	0.00	0.00	0.00	0.00	7.61	0.0	0.944
Eritrea	19,264.50	2,842.37	1.36	2.02	0.00	0.00	19,265.86	-14.3	0.729
Estonia	77.88	209.58	0.00	0.00	0.00	0.00	77.88	-1.4	0.635
Ethiopia	24,980.82	5,996.57	0.17	0.58	0.00	0.00	24,980.99	-179.9	0.033
France	37.89	57.41	0.00	0.00	0.00	0.00	37.89	3.1	0.000
French Southern Territories	0.19	0.58	0.00	0.00	0.00	0.00	0.19	0.0	0.155
Gabon	128.38	213.91	0.00	0.00	0.00	0.00	128.38	1.0	0.750
Gambia	79.86	26.17	0.00	0.00	0.00	0.00	79.86	-0.5	0.197
Georgia	13.02	48.81	0.00	0.00	0.00	0.00	13.02	-0.4	0.560
Germany	0.77	3.69	0.00	0.00	0.00	0.00	0.77	0.0	0.546
Ghana	3,161.34	3,545.31	0.00	0.00	0.00	0.00	3,161.34	-113.6	0.022
Greece	29.18	26.27	0.00	0.00	0.00	0.00	29.18	0.9	0.013
Guadeloupe	0.97	4.62	0.00	0.00	0.00	0.00	0.97	0.0	0.779
Guatemala	35.08	216.47	0.00	0.00	0.00	0.00	35.08	-1.4	0.666
Guinea	411.49	534.09	0.00	0.00	0.00	0.00	411.49	-1.0	0.895
Guinea-Bissau	113.08	363.57	0.00	0.00	0.00	0.00	113.08	-2.4	0.647
Guyana	120.84	142.83	0.00	0.00	0.00	0.00	120.84	-4.5	0.024
Haiti	16.22	35.49	0.00	0.00	0.00	0.00	16.22	-0.2	0.676
Honduras	8.65	26.80	0.00	0.00	0.00	0.00	8.65	-0.4	0.321
Hungary	385.75	759.49	0.14	0.75	0.00	0.00	385.89	-32.9	0.001
India	103,978.73	26,904.56	259.27	308.06	0.00	0.00	104,238.00	-558.7	0.150
Indonesia	245.14	311.54	0.00	0.00	0.00	0.00	245.14	4.1	0.362
Iran	923,430.25	48,133.47	12,786.84	5,039.74	0.00	0.00	936,217.09	3,499.1	0.000
Iraq	311,083.31	9,109.68	2,938.91	1,043.44	0.00	0.00	314,022.22	258.8	0.044
Ireland	0.47	1.44	0.00	0.00	0.00	0.00	0.47	0.0	0.355
Israel	6,720.86	336.71	0.70	0.91	0.00	0.00	6,721.56	16.5	0.000
Italy	571.39	305.60	0.00	0.00	0.00	0.00	571.39	10.0	0.019

Country	**Mean, ESP 6 - 15% (km^2)**	**SD, ESP 6 - 15% (km^2)**	**Mean, ESP 15 - 30% (km^2)**	**SD, ESP 15 - 30% (km^2)**	**Mean, ESP ≥ 30% (km^2)**	**SD, ESP ≥ 30% (km^2)**	**Mean of sodic area, ESP ≥ 6% (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
Jamaica	2.30	2.19	0.00	0.00	0.00	0.00	2.30	-0.1	0.009
Japan	0.03	0.15	0.00	0.00	0.00	0.00	0.03	0.0	0.073
Jordan	67,658.45	1,984.86	551.32	150.22	0.00	0.00	68,209.77	61.8	0.033
Kazakhstan	1,152,995.42	94,477.26	24,518.30	5,620.83	0.00	0.00	1,177,513.72	-3,516.1	0.009
Kenya	50,245.27	9,298.22	512.15	268.42	0.00	0.00	50,757.42	-124.2	0.358
Kuwait	15,613.91	124.84	72.18	73.43	0.00	0.00	15,686.09	2.8	0.007
Kyrgyzstan	3,933.39	779.11	18.52	15.88	0.00	0.00	3,951.91	-48.6	0.000
Laos	7.39	33.74	0.00	0.00	0.00	0.00	7.39	-0.1	0.874
Latvia	3.47	15.52	0.00	0.00	0.00	0.00	3.47	-0.1	0.790
Lebanon	67.54	35.43	0.00	0.00	0.00	0.00	67.54	-0.8	0.108
Lesotho	17.77	17.24	0.00	0.00	0.00	0.00	17.77	0.2	0.435
Liberia	0.09	0.26	0.00	0.00	0.00	0.00	0.09	0.0	0.650
Libya	239,674.27	23,913.78	82.31	11.14	0.00	0.00	239,756.59	100.6	0.772
Lithuania	3.43	14.56	0.00	0.00	0.00	0.00	3.43	-0.3	0.097
Macedonia	0.05	0.31	0.00	0.00	0.00	0.00	0.05	0.0	0.431
Madagascar	397.87	205.02	0.00	0.00	0.00	0.00	397.87	-1.6	0.596
Malawi	108.54	98.18	0.00	0.00	0.00	0.00	108.54	1.2	0.379
Malaysia	203.87	123.33	0.00	0.00	0.00	0.00	203.87	1.8	0.313
Mali	75,362.91	16,841.52	10.00	26.70	0.02	0.13	75,372.93	-763.0	0.001
Malta	0.21	0.46	0.00	0.00	0.00	0.00	0.21	0.0	0.950
Mauritania	87,249.26	9,205.02	176.88	36.66	0.00	0.00	87,426.14	-483.9	0.000
Mexico	253,139.40	18,967.02	1,567.56	461.44	0.00	0.00	254,706.96	383.0	0.166
Moldova	0.69	4.02	0.00	0.00	0.00	0.00	0.69	0.0	0.576
Mongolia	490,605.97	21,187.50	11,782.06	1,966.31	0.44	1.29	502,388.47	166.0	0.589
Morocco	73,776.38	9,306.70	177.78	17.01	0.00	0.00	73,954.17	-289.3	0.027
Mozambique	3,171.49	1,805.45	0.12	0.75	0.00	0.00	3,171.60	-71.1	0.004
Myanmar	759.43	1,198.80	0.00	0.00	0.00	0.00	759.43	-19.0	0.271
Namibia	91,363.26	6,806.23	7.11	8.64	0.00	0.00	91,370.37	-200.3	0.037
Nepal	1,318.28	1,002.96	0.00	0.00	0.00	0.00	1,318.28	-3.8	0.797
Netherlands	0.03	0.17	0.00	0.00	0.00	0.00	0.03	0.0	0.485
New Zealand	0.03	0.14	0.00	0.00	0.00	0.00	0.03	0.0	0.020
Nicaragua	23.28	125.62	0.00	0.00	0.00	0.00	23.28	-0.8	0.680
Niger	65,697.80	11,906.57	0.00	0.00	0.00	0.00	65,697.80	199.3	0.245
Nigeria	4,498.27	5,989.30	3.36	14.11	0.00	0.00	4,501.63	-73.5	0.396
North Korea	11.43	33.07	0.00	0.00	0.00	0.00	11.43	1.1	0.014
Northern Cyprus	54.39	56.78	0.00	0.00	0.00	0.00	54.39	0.1	0.872
Oman	105,289.20	5,868.45	1.13	1.42	0.00	0.00	105,290.33	75.6	0.372
Pakistan	398,545.71	19,378.30	1,772.13	1,204.69	0.32	1.58	400,318.16	958.7	0.000
Palestina	797.15	149.99	1.31	2.83	0.00	0.00	798.45	6.0	0.004
Panama	2.30	14.23	0.00	0.00	0.00	0.00	2.30	-0.1	0.671
Papua New Guinea	19.50	53.35	0.00	0.00	0.00	0.00	19.50	0.0	0.993
Paraguay	12,485.37	12,676.35	0.00	0.00	0.00	0.00	12,485.37	-492.2	0.005

Country	**Mean, ESP 6 - 15% (km^2)**	**SD, ESP 6 - 15% (km^2)**	**Mean, ESP 15 - 30% (km^2)**	**SD, ESP 15 - 30% (km^2)**	**Mean, ESP ≥ 30% (km^2)**	**SD, ESP ≥ 30% (km^2)**	**Mean of sodic area, ESP ≥ 6% (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
Peru	78,620.16	4,684.19	3.46	4.23	0.00	0.00	78,623.62	-248.6	0.000
Philippines	79.56	134.01	0.00	0.00	0.00	0.00	79.56	2.7	0.157
Poland	0.12	0.38	0.00	0.00	0.00	0.00	0.12	0.0	0.158
Portugal	1.93	3.94	0.00	0.00	0.00	0.00	1.93	-0.1	0.078
Puerto Rico	0.69	2.31	0.00	0.00	0.00	0.00	0.69	0.0	0.454
Qatar	10,239.71	65.80	48.01	64.82	0.00	0.00	10,287.72	0.4	0.036
Republic of Congo	129.20	122.21	0.00	0.00	0.00	0.00	129.20	4.0	0.019
Romania	136.44	111.90	0.00	0.00	0.00	0.00	136.44	0.7	0.669
Rwanda	1.18	2.11	0.00	0.00	0.00	0.00	1.18	-0.1	0.001
Saint-Martin	0.08	0.31	0.00	0.00	0.00	0.00	0.08	0.0	0.401
Saudi Arabia	677,693.88	41,498.95	87.12	66.44	0.00	0.00	677,781.00	2,256.2	0.000
Senegal	2,044.49	1,720.61	3.31	5.70	0.00	0.00	2,047.80	-114.8	0.000
Serbia	2.11	2.80	0.00	0.00	0.00	0.00	2.11	0.0	0.276
Sierra Leone	34.69	55.96	0.00	0.00	0.00	0.00	34.69	-0.3	0.670
Singapore	0.18	0.84	0.00	0.00	0.00	0.00	0.18	0.0	0.068
Slovakia	9.87	25.83	0.00	0.00	0.00	0.00	9.87	-1.0	0.005
Solomon Islands	0.89	3.20	0.00	0.00	0.00	0.00	0.89	-0.1	0.004
Somalia	85,739.52	12,169.87	0.60	0.98	0.00	0.00	85,740.12	398.3	0.019
South Africa	30,585.66	7,201.22	0.10	0.39	0.00	0.00	30,585.76	-335.9	0.000
South Korea	0.85	1.37	0.00	0.00	0.00	0.00	0.85	0.0	0.038
South Sudan	4,256.83	4,513.56	0.00	0.00	0.00	0.00	4,256.83	-162.4	0.009
Spain	560.95	174.23	0.00	0.00	0.00	0.00	560.95	-0.4	0.873
Sri Lanka	20.35	17.33	0.00	0.00	0.00	0.00	20.35	0.1	0.705
Sudan	120,928.87	24,802.58	208.71	180.55	0.00	0.00	121,137.58	-1,190.5	0.000
Suriname	0.17	1.09	0.00	0.00	0.00	0.00	0.17	0.0	0.335
Swaziland	1.49	2.46	0.00	0.00	0.00	0.00	1.49	0.0	0.588
Sweden	0.66	1.91	0.00	0.00	0.00	0.00	0.66	0.0	0.686
Switzerland	9.03	7.20	0.00	0.00	0.00	0.00	9.03	-0.5	0.000
Syria	123,697.38	7,403.43	479.93	413.83	0.00	0.00	124,177.30	316.7	0.002
Taiwan	0.66	0.29	0.00	0.00	0.00	0.00	0.66	0.0	0.002
Tajikistan	13,168.44	982.59	373.58	198.73	0.00	0.00	13,542.02	-13.2	0.372
Tanzania	1,607.10	549.61	0.02	0.14	0.00	0.00	1,607.12	18.3	0.017
Thailand	8,772.53	10,338.59	31.13	66.11	0.02	0.13	8,803.68	-306.7	0.035
Timor-Leste	1.02	1.96	0.00	0.00	0.00	0.00	1.02	0.1	0.015
Togo	552.78	605.87	0.00	0.00	0.00	0.00	552.78	-22.4	0.007
Tunisia	22,799.58	2,648.03	156.86	190.38	0.00	0.00	22,956.45	-60.3	0.111
Turkey	25,912.43	5,540.79	1.09	1.46	0.00	0.00	25,913.52	-340.0	0.000
Turkmenistan	378,681.94	14,504.83	1,987.60	1,227.82	0.00	0.00	380,669.54	912.5	0.000
Turks and Caicos Islands	14.16	13.05	0.00	0.00	0.00	0.00	14.16	0.2	0.217
Uganda	70.77	165.29	0.00	0.00	0.00	0.00	70.77	1.5	0.529
Ukraine	257.02	169.78	0.00	0.00	0.00	0.00	257.02	2.2	0.368
United Arab Emirates	29,690.05	1,112.37	157.18	204.86	0.00	0.00	29,847.23	83.1	0.000

Country	**Mean, ESP 6 - 15% (km^2)**	**SD, ESP 6 - 15% (km^2)**	**Mean, ESP 15 - 30% (km^2)**	**SD, ESP 15 - 30% (km^2)**	**Mean, ESP ≥ 30% (km^2)**	**SD, ESP ≥ 30% (km^2)**	**Mean of sodic area, ESP ≥ 6% (km^2)**	**1980 - 2018 trend (km^2 yr^{-1})**	***p*-value (1980 - 2018)**
United States	244,008.42	20,967.09	12,421.72	3,133.88	11.14	20.96	256,441.28	1,316.6	0.000
Uruguay	315.90	402.25	0.00	0.00	0.00	0.00	315.90	3.2	0.588
Uzbekistan	274,290.01	8,495.61	9,344.50	4,175.01	0.00	0.00	283,634.51	-11.7	0.935
Vanuatu	0.38	1.39	0.00	0.00	0.00	0.00	0.38	-0.1	0.004
Venezuela	1,559.29	2,819.07	0.00	0.00	0.00	0.00	1,559.29	-47.4	0.243
Vietnam	161.87	459.80	0.00	0.00	0.00	0.00	161.87	-2.9	0.666
Virgin Islands, U.S.	0.06	0.29	0.00	0.00	0.00	0.00	0.06	0.0	0.528
Western Sahara	79,663.44	5,560.93	257.01	14.40	0.00	0.00	79,920.45	-6.1	0.940
Yemen	75,857.80	17,624.44	7.31	8.47	0.00	0.00	75,865.11	1,054.6	0.000
Zambia	865.73	1,306.79	0.00	0.00	0.00	0.00	865.73	-42.6	0.020
Zimbabwe	287.96	373.02	0.00	0.00	0.00	0.00	287.96	-5.5	0.303

Computer codes

This section provides the scripts and codes required to regenerate the results. Please note that ArcGIS Desktop 10.x license is needed to run ArcPy module. Also, the MATLAB Parallel Computing plus Statistics and Machine Learning toolboxes are required for running the MATLAB codes provided here.

Pre-processing the predictors' layers

The scripts provided in this sub-section show how we pre-processed the predictors assembled from the different sources and made them ready for data extraction. We refer the reader to Table A2-1 to see the corresponding pre-processing steps for each individual predictor.

Static topographic predictors including slope (degrees), plan and profile curvatures, slope length (m), and Terrain Ruggedness Index (TRI) were calculated in SAGA GIS GUI (Graphical User Interface) from CGIAR CSI SRTM 90 m Digital Elevation Database v4.1. The original DEM (Digital Elevation Model) data were resampled to 250 m and saved in three separate raster datasets named: North East, South East, and West. We downloaded these three layers, mosaicked them in ArcGIS for Desktop environment (herein we refer to its central application: ArcMap) and exported the generated global layer as a single geo-tiff. To generate the map of the topographic predictors including slope, slope length, TRI, plan, and profile curvatures, it was necessary to have the original DEM in a projected coordinates system. For computing those topographic predictors, we first projected the global DEM layer to World Mercator coordinates system (with 259.511 m spatial resolution) using ArcMap "raster project" tool. To reduce the computational load and avoid system crashes in SAGA GIS, we produced a separate DEM layer in the World Mercator coordinates system, however, at 1,000 m spatial resolution to generate the maps of slope length and TRI. For the plan and profile curvatures, the 10 parameter 3rd order polynomial method was used. Also, we used a square cell with radius of three for calculation of TRI.

Other static predictors were directly pre-processed (including projections and per-cell statistics) through ArcMap GUI and the following scripts are not applicable to those predictors. Soil texture raster datasets of clay, silt, and sand content at different depths were averaged using ArcMap "raster calculator" tool. To get an average of soil texture properties between soil surface and 100 cm depth from the available values of SoilGrids250 datasets for five standard depths of 0, 15, 30, 60, and 100 cm, we applied the trapezoidal rule as follows:

$$\textit{Soil texture property average between 0 and 100 cm} =$$

$$[(15 - 0) \times (R_{val}(15) + R_{val}(0)) + (30 - 15) \times (R_{val}(30) + R_{val}(15)) + \dots$$

$$(60 - 30) \times (R_{val}(60) + R_{val}(30)) + (100 - 60) \times (R_{val}(100) + R_{val}(60))] / (100 \times 2),$$

where R_{val} *(depth)* was the raster value at the corresponding depth.

Some predictors were originally in an .hdf format. HDF files were composed of different layers (or sub-datasets) and we required only one or two layers form those sub-datasets. An example of these kind of predictors was VIP30 v. 004 dataset and NDVI and EVI2 layers were the required sub-datasets. We used the following Python code to automatize the processes of extracting the desirable sub-dataset layers:

```
## Extract sub-dataset, we used PyCharm Python IDE (Integrated Development Environment)
## Usage: Extracting the required sub-datasets from predictors with .hdf format and saving the
## final sub-dataset as a geo-tiff.

import arcpy # Importing the ArcPy module
import os # Importing Miscellaneous operating system module required for reading the file
names in a directory
import os, fnmatch
```

```
# Setting the geo-processing environments
arcpy.env.overwriteOutput = True
arcpy.env.workspace = r"The directory of .hdf files"
arcpy.env.geographicTransformations = arcpy.SpatialReference(4326) # Setting the output
coordinates as WGS 1984

# Reading the .hdf files needed to be processed from a directory
path = r" The directory of .hdf files "
pattern = "*.hdf"
hdf_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]

for i in range(0,len(hdf_files)):
 # i is index of the .hdf file
 arcpy.ExtractSubDataset_management(hdf_files[i],"Output directory"+str(i)+".tif", "S_N")
 # Extracting the sub-dataset and saving as geo-tiff using ArcPy ExtractSubDataset_management
 # S_N is the sub-dataset number in .hdf file
```

In some cases, the original files were in an .nc format. To convert those netcdf files to raster layers we used the following Python code:

```
## Making raster layers from a netcdf file,
## Usage: This code first extracts the different temporal layers of the netcdf files and then
## saves each layer as a separate raster file in .tif format.

import arcpy # Importing the ArcPy module and spatial analysis required functions
from arcpy import env
from arcpy.sa import *
import os # Importing Miscellaneous operating system module required for reading the file
names in a directory
import os, fnmatch

# Setting the ArcPy geo-processing environments
arcpy.env.overwriteOutput = True
arcpy.env.workspace = r"The directory of the nc files"
arcpy.env.geographicTransformations = arcpy.SpatialReference(4326) # Setting the output
coordinates as WGS 1984

# Reading the nc files in the directory where they are stored
path = r" The directory of the nc files"
nc_files = [f for f in os.listdir(path) if f.endswith(".nc")]

# i is index of the nc file
for i in range(0,len(nc_files)):
   inNetCDFFile = nc_files[i]
   variable = " The name of the variable in the netcdf file"
   XDimension = "longitude" # In the netcdf file
   YDimension = "latitude" # In the netcdf file
   outRasterLayer = "Created_layer"
   bandDimmension = "" # Varies depending on the band dimension (time) name in the nc file
   dimensionValues = ""
   valueSelectionMethod = ""
   # Executing ArcPy the MakeNetCDFRasterLayer_md tool
   arcpy.MakeNetCDFRasterLayer_md(inNetCDFFile, variable, XDimension, YDimension,\
                             outRasterLayer, bandDimmension, dimensionValues,\
                              valueSelectionMethod)
   # Saveing the created layers in memory
   arcpy.SaveToLayerFile_management('Created_layer', 'Temporaty_saved_layer'+str(i))
   # Saving the created layers on the disk
   arcpy.CopyRaster_management('Temporaty_saved_layer'+str(i)+'.lyr',"Output
   location"+str(i)+'.tif')
```

After converting all predictors' datasets to raster layers, we used the ArcPy "cell-statistics" geo-processing tool to calculate the per-cell average of dynamic predictors. Temporal resolution of the predictors was different. First we generated the annual averages of each predictor. For the predictors with decadal averaging window, we calculated the average in each year from 1971 to 2018. For the predictors with 5-year averaging window, we computed annual averages from 1976. For the rest of predictors, we generated annual averages from 1980. Unfortunately, vegetation indices data including NDVI, EVI2, LAI, and FAPAR were not available for 1980. Therefore, we produced their layers by calculating an average between 1981 and 1985. For instance, we generated the raster layer of NDVI in 1980 (which was missing in the original VIP30 v. 004 dataset) by calculating the per-cell average of NDVI raster layers between 1981 and 1985. Then we computed running window averages of the predictors with decadal and five-year averaging windows (Table A2-1) from 1980 to 2018 using the following Python code. For each particular predictor and each year between 1980 and 2018, a raster layer representing the average of the corresponding predictor in the averaging window was generated. We named (labelled) these raster layers with the predictor's name as a prefix and the number of the year to which the raster layer was corresponded. Final averaged rasters of each predictor were saved in separate directories for extraction of the values and further processing.

```
## Calculation of per cell average for predictors with decadal and five-year averaging window,
## Usage: this code gets a large number of rasters in a directory and calculates the per cell
## average of the rasters
## and generates a final raster layer which is the average of input rasters.

import arcpy # Importing the ArcPy module
from arcpy import env
from arcpy.sa import * # Importing all functions from ArcPy spatial analyst toolbox
import os # Importing Miscellaneous operating system module required for reading the file
names in a directory
import os, fnmatch

# Setting the geo-processing environments
arcpy.env.workspace = r"The directory of rasters for each particular predictor"
arcpy.env.extent = "MAXOF"
arcpy.env.overwriteOutput = True
arcpy.env.geographicTransformations = arcpy.SpatialReference(4326) # Setting the output
coordinates as WGS 1984

C = "Raster name prefix"
for i in range(1980, 2019):
 # i is the index of year
 # Execution of the cell-statistics tool
 # For predictors with decadal averaging window:
  outCellStatistics = ([C + str(i) + ".tif", C + str(i-1) + ".tif", C + str(i-2) + ".tif",…, C
  + str(i-9) + ".tif"], "MEAN", "DATA")
 # For predictors with five-year averaging window:
  outCellStatistics = ([C + str(i) + ".tif", C + str(i-1) + ".tif", C + str(i-2) + ".tif",…, C
  + str(i-4) + ".tif"], "MEAN", "DATA")
 # The output of cell-statistics is temporary (saved on memory)
 # Saving the output of cell-statistics on the disk
  outCellStatistics.save("Output folder/Predictor_name_" + str(i) + ".tif")
```

Some pixels were missing in the final generated rasters of the predictors; mostly in layers of the remotely-sensed soil moisture and vegetation indices. We filled the spatial gaps (pixels with null values) in the data layers using the mean of surrounding pixels. A circle with radius of 4 from the neighbouring cells of the gap was used to calculate the mean through application of the following Python code:

```
## Filling the gaps in rasters,
## Usage: This code fills the gaps (null cells) in generated rasters for extraction of
## predictors' values.

import arcpy # Importing the ArcPy module
from arcpy import env
from arcpy.sa import * # importing the functions of spatial analyst toolbox
# Importing Miscellaneous operating system module required for reading the file names in a
directory
import os
import os, fnmatch

# Setting the geo-processing environments
arcpy.env.parallelProcessingFactor = "100%"
arcpy.env.workspace = r"The directory of rasters"
```

```
arcpy.env.extent = "MAXOF"
arcpy.CheckOutExtension("Spatial")

# Acquiring all rasters within a directory
path = r" The directory of rasters "
pattern = "*.tif"
tif_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]

# i is index of the .tif file
for i in range(0,len(tif_files)):
 string = tif_files[i] # raster layer name
 string_1 = string[0:(len(string)-4)] # Raster layer name without .tif suffix
 # Execution of the Filling. This part is a combination of ArcPy Focal Statistics and Raster
 Calculator tools
 # A circle with radius of 4 from the neighbouring cells of the gap is used to calculate the
 average and gap is filled by the average value
 Rasterfilled = Con(IsNull(string), FocalStatistics(string,NbrCircle(4,"CELL"),"MEAN","DATA"),
 string)
 # The output is temporary (saved on memory) and the following saves it on the disk
 arcpy.CopyRaster_management(Rasterfilled, r"The output directory"+string_1+".tif")
```

Extracting the predictors' values to training point feature layers

Merging of the training datasets (for ESP) from different source datasets and their pre-processing (see Methods, Data) was fully accomplished in ArcMap GUI and corresponding toolboxes. The indicator of the missing values in original training datasets was replaced by -9,999, soil layers attributes were joined to their corresponding geo-referenced profile locations, and the valid range for ESP was assumed to be 0 to 100%. Then we removed the profiles in AfSP and WISE datasets that spatially intersect the NCSS profiles and merged these three datasets into a single inventory. The final training datasets were saved as a point feature class (.shp format) file. We first projected these layers in ArcMap environment to World Mercator coordinates system to extract the values of static predictors in the World Mercator projection. We projected the point shape files instead of rasters to avoid data loss due to raster resampling.

After extraction, the two shape files were re-projected to WGS 1984 coordinates system to extract the values of other predictors. For all static predictors, the extractions were directly conducted by “extract multi values to points” tool from ArcMap “spatial analyst” toolbox. However, for the dynamic predictors, first we applied the ArcPy “select layer by attribute” tool to the point features classes and divided the points according to the year of acquisition. In detail, in each point feature class, there were some points with x- and y- spatial coordinates values representing the locations where soil EC_e and ESP were sampled. In the attribute table of these x- y- points, the year of acquisition of the sample, lower sample's depth, upper sample's depth (from the soil surface), and the measured values of EC_e or ESP for that sample were reported. We selected the samples with the same year of acquisition and exported the selected samples as new point feature classes for further processing and extraction of the dynamic predictors' values. Therefore, a total of 39 point feature layers labelled by the year of acquisition of samples were generated (since 1980). The following Python script shows the selection process:

```
## Select by attribute,
## Usage: This code selects the points in original datasets (needed for training) based on the
## year of acquisition in attribute table.
## This code splits the point feature layers of the training datasets into smaller point
## feature layers. Each smaller layer is labelled by the name of the year.

import arcpy # Importing the ArcPy module

# Setting the geo-processing environments
arcpy.env.workspace = r"The directory of training datasets"
arcpy.env.overwriteOutput = True
# Importing the original dataset feature point layer into memory
arcpy.MakeFeatureLayer_management("ESP/ECe.shp", "lyr")
# CC is the prefix of the generated point feature layers ECE_ or ESP_ for each year
```

```
CC = "" or ""
for i in range (1980,2019):
 # i is index of the year # -9999 is the index of the missing data
 arcpy.SelectLayerByAttribute_management("lyr", "NEW_SELECTION", "Year >= '"+str(i)+"' AND
 Year < '"+str(i+1)+"' AND NOT Year = '-9999'")
 arcpy.CopyFeatures_management('lyr', "Output directory/"+CC+str(i))
```

Then we extracted the values of the predictors at each year corresponding to the year of acquisition of the point feature layer using ArcPy “extract multi values to points” geo-processing tool as follows:

```
## Extract multi values to training sets' data points,
## Usage: This code extracts the values of each predictor’s raster layer (labelled by year) to
## point feature layer of the training datasets labelled with the similar year. The extracted
## values emerges in the attribute table of the point feature dataset.

import arcpy # Importing the ArcPy module
from arcpy import env
from arcpy.sa import * # Importing all functions in spatial analyst toolbox

# Setting the geo-processing environments
arcpy.CheckOutExtension("Spatial") # Checking for the spatial analyst license
arcpy.env.workspace = r"The directory of point feature layers" # The directory should also
include the dynamic predictors' raster layers
C = "the prefix of the point feature layers for each year"

# i is index of the year
for i in range(1980,2019):
   # Execution of extraction # For each predictor this loop has to iterate
   inRasterList = [["Raster layer name" + str(i) +".tif","Name of the extracted value in
   attribute table of the point feature layer"]]
   inPointFeatures = C + str(i) + ".shp"
   ExtractMultiValuesToPoints(inPointFeatures, inRasterList)
```

The extracted values of each predictor were added to the attribute table of individual years’ point feature layers. The attribute tables were composed of columns with headers named after the predictors and rows representing the sample observations. After extraction of the predictors’ values, the point features layers were merged by ArcMap “merge” tool and the final attribute tables were exported as text files (.txt format). These text files were imported to MATLAB for fitting the models and further analysis.

Model training

The prepared text files were then imported to MATLAB workspace. We calculated the linear Pearson correlation coefficients between each predictor and target variables as a univariate criterion to filter the unnecessary predictors, assuming no interaction between the predictors (Table A2-24). The Pearson correlation coefficients between the predictors and target variables were non-significant; so we retrieved all predictors for further modelling. Initially, we used MATLAB Regression and Classification Learner applications to examine the performance of different built-in models available in MATLAB Statistics and Machine Learning toolbox. We used two-part models for mapping the relation between predictors and target variables. We held out 25% of the training sets and fitted the models with default models’ hyperparameters. Tree-based ensemble models were the most suitable among other models for our regression and classification tasks (see Table A2-6).

Table A2-24: Pearson's linear correlation coefficient between the non-categorical predictors' values and target variables (EC_e or ESP). Pearson's correlation coefficients equal to -1 and +1 indicate perfect negative and positive correlations between predictor and variable, respectively. For the full name of the predictors see Table A2-1.

Predictor name	EC_e	ESP
	Pearson correlation coefficient	**Pearson correlation coefficient**
Sample's upper depth	0.083	0.120
Sample's lower depth	0.087	0.128
Elevation	-0.075	0.060
Plan curvature	-0.013	-0.006
Profile curvature	0.002	0.000
Slope	-0.121	-0.115
Slope length	0.007	0.068
Terrain Ruggedness Index	-0.118	-0.102
Fertiliser input for C3 annual crops	-0.028	-0.050
Fertiliser input for C3 perennial crops	0.055	0.024
Water table depth	-0.083	-0.018
Aspect	0.015	-0.018
Topographic index	0.119	0.106
Soil clay content	-0.166	0.085
Soil silt content	0.049	-0.123
Soil sand content	0.092	0.064
Soil-sedimentary thickness	0.133	0.138
Average rooting depth	-0.070	0.022
Diurnal temperature range	-0.040	0.140
Precipitation	-0.127	-0.209
Average temperature	-0.058	0.046
Maximum temperature	-0.067	0.073
Minimum temperature	-0.049	0.020
Root-zone soil moisture	-0.115	-0.220
PDSI	0.029	-0.064
Soil surface moisture (2 - 5 cm)	-0.091	-0.204
Evaporative stress factor	-0.054	-0.218
EVI2	-0.180	-0.254
NDVI	-0.191	-0.268
FAPAR	-0.200	-0.262
LAI	-0.165	-0.240
Wind speed	0.159	0.071
Soil surface (skin) temperature	-0.020	0.076
Soil layer one temperature	-0.012	0.101
Soil layer two temperature	-0.016	0.100
Soil layer three temperature	-0.015	0.100
Soil layer four temperature	-0.015	0.101
Potential evapotranspiration	0.047	0.175
Water deficit	0.161	0.271
Actual evapotranspiration	-0.178	-0.229

For the classification part, we used MATALB "fitcensemble" function. We ignored the slight imbalance between the classes in ESP training set. To resolve the presence of imbalance between the classes of EC_e training dataset, however, application of under-sampling, over-sampling (using Synthetic Minority Over-sampling Technique: SMOTE (Chawla et al. 2002)), and/or a combination of these two techniques was possible. Also MATLAB “fitcensemble” allowed us to modify the misclassification cost matrix to handle the imbalance in classes. We developed the following MATLAB script to inspect the effect of abovementioned solutions on performance of the fitted models by “fitcensemble” function:

```
clc;
clear;
%% Effect of misclassification cost, over-sampling, and under-sampling,
%% Usage: This code will examine the effect of altering the misclassification
%% cost of the class with lower number of samples in imbalanced binary classification. Also
%% it examines the effect of under-sampling from the class with higher number
%% of samples, over-sampling from the class with lower number of samples,
%% and a combination of these methods. For each method, combined effect of
%% manipulating the misclassification cost is also analysed. For
%% over-sampling, we have used Synthetic Minority Over-Sampling
%% Technique (SMOTE). We implemented MATLAB R2019 for running this.

ECe = readtable('Location of the ECe training dataset on the disk','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true); % Importing the table of training
     % dataset which is ECe here

%% Preparing the table
table = standardizeMissing(ECe,-9999);% Converting the cells with missing values indicator
% (-9999) to MATLAB standard NaN
table.FID = [];
table.Year = [];
table(sum(ismissing(table),2) > 0,:) = [];% Dropping the rows with missing values
edges = [0 2 100];% Setting the classes edges
table.ECe = discretize(table.ECe,edges); % Discretising the ECe values into two classes,
% saline and non-saline

%% Partitioning
% This part partitions the original dataset to training (75%) and test sets (25%)
c = cvpartition(table.ECe,'Holdout',0.25,'Stratify',true); % Data will be stratified between
% the two sets; this assures that data from both classes are available in the two final sets
idx1 = test(c);
idx2 = training(c);
Test = table(idx1(:)==1,:);
Training = table(idx2(:)==1,:);

%% Preparing required tables for training and validation
% Categorizing the categorical variables in the test set
Test.Main_litho = categorical(Test.Main_litho);
Test.WRB = categorical(Test.WRB);
Test.LC = categorical(Test.LC);
TrueLabels = Test.ECe;
Class_1 = Training(Training.ECe == 1,:);
Class_1_refilled = Training(Training.ECe == 1,:);
Class_2 = Training(Training.ECe == 2,:);
% Categorizing the categorical variables in the training set
Training.Main_litho = categorical(Training.Main_litho);
Training.WRB = categorical(Training.WRB);
Training.LC = categorical(Training.LC);

%% Model Training
%% Calculating the classification accuracy metrics for different misclassification costs
%% 'i' for the misclassification cost

Row = 1;
Accuracy_Metrics = zeros(13,10);
for i = 1:0.25:4
% Each iteration of 'i' changes the misclassification cost of the class with lower number of
% samples in fitcensemble function misclassification cost matrix
% For hyperparameter optimisation, 130 iterations are conducted to evaluate the objective
% function. Holdout set (with %25 held out) was used to evaluate the objective function
% 'ens' is the object of the final trained model
  ens = fitcensemble(Training,'ECe','Cost',[0 1;i 0],'OptimizeHyperparameters',...
  {'Method','LearnRate','NumLearningCycles','MinLeafSize','MaxNumSplits',...
  'NumVariablesToSample','SplitCriterion'},'HyperparameterOptimisationOptions',...
struct('Holdout',.25,'UseParallel',true,'MaxObjectiveEvaluations',130,'Repartition',true,'Show
Plots',false,'Verbose',0));
    %%Obtaining validation metrics
    Predictedlabels = predict(ens,Test);
    % Tp = True Positive, Fn = False Negative , Fp = False Positive, Tn = True Negative
```

```
    C = confusionmat(TrueLabels,Predictedlabels); Tp = C(1,1); Fn = C(1,2); Fp = C(2,1); Tn =
    C(2,2);
    Accuracy_Metrics(Row,1) = i;
    Accuracy_Metrics(Row,2) = Tp; Accuracy_Metrics(Row,3) = Fn;
    Accuracy_Metrics(Row,4) = Fp; Accuracy_Metrics(Row,5) = Tn;
    Accuracy_Metrics(Row,6) = loss(ens,Test,'ECe'); % Binary misclassification loss
    Accuracy_Metrics(Row,7) = (Tp+Tn)/(Tp+Fp+Fn+Tn)*100; % Binary classification accuracy
    Accuracy_Metrics(Row,8) = Tp/(Tp+Fp); % Precision
    Accuracy_Metrics(Row,9) = Tp/(Tp+Fn); % Recall
    %MCC (Matthews Correlation Coefficient) for binary imbalanced classification
    Accuracy_Metrics(Row,10) = ((Tp*Tn)-(Fp*Fn))/sqrt((Tp+Fp)*(Tp+Fn)*(Tn+Fp)*(Tn+Fn));
    Row = Row + 1;
end
% Saving results as a table
Accuracy_Metrics = array2table(Accuracy_Metrics,'VariableNames',{'Cost_2_1' 'Tp' 'Fn'...
'Fp''Tn' 'Loss_Classification_error' 'Accuracy' 'Precision' 'Recall' 'MCC'});
writetable(Accuracy_Metrics, 'Output directory\output file name.txt');

%% Calculating the classification accuracy metrics for the extent of oversampling and
%% misclassification cost; 'i' for the number of increased samples, 'j' for the
%% misclassification cost. We have used SMOTE (Synthetic Minority Over-sampling Technique) for
%% generation of synthetic samples. Each iteration of 'i' produces 2500 new samples using
%% the feature space between the 2500 random selected samples (without replacement) and
%% their nearest neighbours

Row = 1;
Accuracy_Metrics = zeros(28,11);
Oversampling_rate = 2500;
Class_2_matrix = table2array(Class_2);
for i = 1:4
    % Random sample selection from class 2 (saline) without replacement and copying into a
    % new matrix(named matrix)
    y = randsample(size(Class_2_matrix,1),Oversampling_rate);
    matrix = Class_2_matrix(y,:);
    z = 1;
    % 'inc' is the number of synthetic samples which will be made in the feature space between
    % the two nearest neighbours
    inc = 1;
    % Augmented matrix is the matrix of new generated samples
    Augmented_matrix = zeros(size(matrix,1)*inc,44);
    for ii = 1:size(matrix,1)
       % Finding the nearest neighbours of each query row of the selected matrix of samples
       % in Class 2 (saline)
        Index = knnsearch(Class_2_matrix,matrix(ii,:),'K',2);
        % Generation of samples between the two nearest neighbours using proposed
        % interpolation method in SMOTE
        for jj = z:inc + (z-1)
            Augmented_matrix(jj,1:2) =
            Class_2_matrix(ii,1:2)+rand*(Class_2_matrix(Index(1,2),1:2)-...
            Class_2_matrix(ii,1:2));
            % The target variable was the third column in the training set
            Augmented_matrix(jj,3) = 2;
            Augmented_matrix(jj,4:end) = ...
            Class_2_matrix(ii,4:end)+rand*(Class_2_matrix(Index(1,2),4:end)-...
            Class_2_matrix(ii,4:end));
            % Categorical variables of the generated sample are the same as the original
            % sample (not eligible for interpolation)
            Augmented_matrix(jj,20) = Class_2_matrix(Index(1,2),20);
            Augmented_matrix(jj,21) = Class_2_matrix(Index(1,2),21);
            Augmented_matrix(jj,44) = Class_2_matrix(Index(1,2),44);
        end
        z = jj+1;
    end
    % Converting the augmented matrix to a table
    Augmented_matrix = array2table(Augmented_matrix,'VariableNames',...
    {'upper_dept' 'lower_dept' 'ECe' 'Elevation' 'Pla_cur' 'Pro_cur' 'Slope' ...
    'Slope_Leng' 'TRI' 'c3ann' 'c3per' 'WTD' 'Aspect' 'Topo_index' 'Clay' 'Silt' ...
    'Sand' 'Soil_thick' 'Root_D' 'WRB' 'Main_litho' 'dtr' 'Pre' 'T_ave' 'T_max' 'T_min' ...
    'S_mo' 'PDSI' 'Sat_SM' 'Gleam_S' 'EVI' 'NDVI' 'FAPAR' 'LAI' 'Wind_S' 'Skin_T' ...
    'S_T_1' 'S_T_2' 'S_T_3' 'S_T_4' 'Pet' 'Def' 'aet' 'LC'});
    Augmented_matrix.Main_litho = categorical(Augmented_matrix.Main_litho);
    Augmented_matrix.WRB = categorical(Augmented_matrix.WRB);
    Augmented_matrix.LC = categorical(Augmented_matrix.LC);
    % Adding the augmented matrix to the original training set
    if i == 1
        Old_Augmented_matrix = [];
    end
```

```
    Tbl = [Training;Old_Augmented_matrix;Augmented_matrix];
    % To keep the previous generated samples in the next iteration:
    Old_Augmented_matrix = [Old_Augmented_matrix;Augmented_matrix];
    for j = 1:0.5:4
        % Each iteration of 'j' examines the effect of misclassification cost
        % of the class with lower number of samples on the performance of
        % the fitted model
        ens = fitcensemble(Tbl,'ECe','Cost',[0 1;i 0],'OptimizeHyperparameters',...
{'Method','LearnRate','NumLearningCycles','MinLeafSize','MaxNumSplits',...
'NumVariablesToSample','SplitCriterion'},'HyperparameterOptimisationOptions',...
struct('Holdout',.25,'UseParallel',true,'MaxObjectiveEvaluations',130,...
'Repartition',true,'ShowPlots',false,'Verbose',0));
        %% Obtaining the validation metrics
        Predictedlabels = predict(ens,Test);
        C = confusionmat(TrueLabels,Predictedlabels); Tp = C(1,1); Fn = C(1,2); Fp = C(2,1);
        Tn = C(2,2);
        Accuracy_Metrics(Row,1) = i*Oversampling_rate;
        Accuracy_Metrics(Row,2) = j;
        Accuracy_Metrics(Row,3) = Tp; Accuracy_Metrics(Row,4) = Fn;
        Accuracy_Metrics(Row,5) = Fp; Accuracy_Metrics(Row,6) = Tn;
        Accuracy_Metrics(Row,7) = loss(ens,Test,'ECe');
        Accuracy_Metrics(Row,8) = (Tp+Tn)/(Tp+Fp+Fn+Tn)*100;
        Accuracy_Metrics(Row,9) = Tp/(Tp+Fp);
        Accuracy_Metrics(Row,10) = Tp/(Tp+Fn);
        Accuracy_Metrics(Row,11) = ((Tp*Tn)-(Fp*Fn))/sqrt((Tp+Fp)*(Tp+Fn)*(Tn+Fp)*(Tn+Fn));
        Row = Row + 1;
    end
end
% Saving the results as a table
Accuracy_Metrics = array2table(Accuracy_Metrics,'VariableNames',...
{'Augmented_samples_Num' 'Cost_2_1' 'Tp' 'Fn' 'Fp' 'Tn'...
 'Loss_Classification_error' 'Accuracy' 'Precision' 'Recall' 'MCC'});
writetable(Accuracy_Metrics, 'Output directory\output file name.txt');

%% Calculating the classification accuracy metrics for different under-sampling rates and
%% misclassification cost. 'i' the number of decreased samples, 'j' for the misclassification
%% cost, under-sampling rate is 2500 samples in each iteration

Row = 1;
Accuracy_Metrics = zeros(28,11);
Undersampling_rate = 2500;
for i = 1:4
    % Random undersampling without replacement
    Class_1(randsample(height(Class_1),Undersampling_rate),:) = [];
    Tbl = [Class_1;Class_2];
    Tbl.Main_litho = categorical(Tbl.Main_litho);
    Tbl.WRB = categorical(Tbl.WRB);
    Tbl.LC = categorical(Tbl.LC);
    for j = 1:0.5:4
        ens = fitcensemble(Tbl,'ECe','Cost',[0 1;j 0],'OptimizeHyperparameters',...
{'Method','LearnRate','NumLearningCycles','MinLeafSize','MaxNumSplits','NumVariablesToSample',
'SplitCriterion'},'HyperparameterOptimisationOptions',...
struct('Holdout',.25,'UseParallel',true,'MaxObjectiveEvaluations',130,'Repartition',true,'Show
Plots',false,'Verbose',0));
        %% Obtaining validation metrics
        Predictedlabels = predict(ens,Test);
        C = confusionmat(TrueLabels,Predictedlabels); Tp = C(1,1); Fn = C(1,2); Fp = C(2,1);
        Tn = C(2,2);
        Accuracy_Metrics(Row,1) = i*Undersampling_rate;
        Accuracy_Metrics(Row,2) = j;
        Accuracy_Metrics(Row,3) = Tp; Accuracy_Metrics(Row,4) = Fn;
        Accuracy_Metrics(Row,5) = Fp; Accuracy_Metrics(Row,6) = Tn;
        Accuracy_Metrics(Row,7) = loss(ens,Test,'ECe');
        Accuracy_Metrics(Row,8) = (Tp+Tn)/(Tp+Fp+Fn+Tn)*100;
        Accuracy_Metrics(Row,9) = Tp/(Tp+Fp);
        Accuracy_Metrics(Row,10) = Tp/(Tp+Fn);
        Accuracy_Metrics(Row,11) = ((Tp*Tn)-(Fp*Fn))/sqrt((Tp+Fp)*(Tp+Fn)*(Tn+Fp)*(Tn+Fn));
        Row = Row + 1;
    end
end
Accuracy_Metrics = array2table(Accuracy_Metrics,'VariableNames',{'Removed_samples_Num'
'Cost_2_1' 'Tp' 'Fn' 'Fp' 'Tn'...
'Loss_Classification_error' 'Accuracy' 'Precision' 'Recall' 'MCC'});
writetable(Accuracy_Metrics, 'Output directory\output file name.txt');

%% Calculating the classification accuracy metrics for combined under-sampling and over-
%% sampling and misclassification cost change. 'i' for the iteration of decreased and
```

```
%% increased samples,'j' for misclassification cost, under-sampling rate is 1500 samples in
%% each iteration. Oversampling rate is 1500 for each iteration

Row = 1;
Accuracy_Metrics = zeros(36,12);
Oversampling_rate = 1500;
Class_2_matrix = table2array(Class_2);
Undersampling_rate = 1500;
for i = 1:4
    % Random oversampling with replacement
    y = randsample(size(Class_2_matrix,1),Oversampling_rate);
    matrix = Class_2_matrix(y,:);
    z = 1;
    inc = 1;
    Augmented_matrix = zeros(size(matrix,1)*inc,44);
    for ii = 1:size(matrix,1)
        Index = knnsearch(Class_2_matrix,matrix(ii,:),'K',2);
        for jj = z:inc + (z-1)
            Augmented_matrix(jj,1:2) = ...
            Class_2_matrix(ii,1:2)+rand*(Class_2_matrix(Index(1,2),1:2)- ...
            Class_2_matrix(ii,1:2));
            Augmented_matrix(jj,3) = 2;
            Augmented_matrix(jj,4:end) = ...
            Class_2_matrix(ii,4:end)+rand*(Class_2_matrix(Index(1,2),4:end)- ...
            Class_2_matrix(ii,4:end));
            Augmented_matrix(jj,20) = Class_2_matrix(Index(1,2),20);
            Augmented_matrix(jj,21) = Class_2_matrix(Index(1,2),21);
            Augmented_matrix(jj,44) = Class_2_matrix(Index(1,2),44);
        end
        z = jj+1;
    end
    Augmented_matrix = array2table(Augmented_matrix,'VariableNames',...
    {'upper_dept' 'lower_dept' 'ECe' 'Elevation' 'Pla_cur' 'Pro_cur' 'Slope' ...
     'Slope_Leng' 'TRI' 'c3ann' 'c3per' 'WTD' 'Aspect' 'Topo_index' 'Clay' 'Silt' ...
     'Sand' 'Soil_thick' 'Root_D' 'WRB' 'Main_litho' 'dtr' 'Pre' 'T_ave' 'T_max' 'T_min' ...
     'S_mo' 'PDSI' 'Sat_SM' 'Gleam_S' 'EVI' 'NDVI' 'FAPAR' 'LAI' 'Wind_S' 'Skin_T' ...
     'S_T_1' 'S_T_2' 'S_T_3' 'S_T_4' 'Pet' 'Def' 'aet' 'LC'});
    % Random_under-sampling_without_replacement
    Class_1_refilled(randsample(height(Class_1_refilled),Undersampling_rate),:) = [];
    if i == 1
        Old_Augmented_matrix = [];
    end
    Tbl = [Class_1_refilled;Class_2;Old_Augmented_matrix;Augmented_matrix];
    Tbl.Main_litho = categorical(Tbl.Main_litho);
    Tbl.WRB = categorical(Tbl.WRB);
    Tbl.LC = categorical(Tbl.LC);
    Old_Augmented_matrix = [Old_Augmented_matrix;Augmented_matrix];
    for j = 1:0.5:5
        ens = fitcensemble(Tbl,'ECe','Cost',[0 1;j 0],'OptimizeHyperparameters',...
{'Method','LearnRate','NumLearningCycles','MinLeafSize','MaxNumSplits','NumVariablesToSample',
'SplitCriterion'},'HyperparameterOptimisationOptions',...
struct('Holdout',.25,'UseParallel',true,'MaxObjectiveEvaluations',130,'Repartition',true,'Show
Plots',false,'Verbose',0));
        %% Obtaining validation metrics
        Predictedlabels = predict(ens,Test);
        C = confusionmat(TrueLabels,Predictedlabels); Tp = C(1,1); Fn = C(1,2); Fp = C(2,1);
        Tn = C(2,2);
        Accuracy_Metrics(Row,1) = i*Undersampling_rate;
        Accuracy_Metrics(Row,2) = i*Oversampling_rate;
        Accuracy_Metrics(Row,3) = j;
        Accuracy_Metrics(Row,4) = Tp; Accuracy_Metrics(Row,5) = Fn;
        Accuracy_Metrics(Row,6) = Fp; Accuracy_Metrics(Row,7) = Tn;
        Accuracy_Metrics(Row,8) = loss(ens,Test,'ECe');
        Accuracy_Metrics(Row,9) = (Tp+Tn)/(Tp+Fp+Fn+Tn)*100;
        Accuracy_Metrics(Row,10) = Tp/(Tp+Fp);
        Accuracy_Metrics(Row,11) = Tp/(Tp+Fn);
        Accuracy_Metrics(Row,12) = ((Tp*Tn)-(Fp*Fn))/sqrt((Tp+Fp)*(Tp+Fn)*(Tn+Fp)*(Tn+Fn));
        Row = Row + 1;
    end
end
% Saving results as a table
Accuracy_Metrics = array2table(Accuracy_Metrics,'VariableNames',{'Removed_samples_Num'
'Augmented_samples_Num' 'Cost_2_1' 'Tp' 'Fn' 'Fp' 'Tn'...
'Loss_Classification_error' 'Accuracy' 'Precision' 'Recall' 'MCC'});
writetable(Accuracy_Metrics, 'Output directory\output file name.txt');
```

The results (Table A2-25 to Table A2-28) showed that none of the imbalance handling techniques were effective in improving the performance of the final fitted binary classification. Thus, we just set the misclassification cost of the minority (or saline class) to be two since the number of samples in the saline class were half of the non-saline class.

Table A2-25: Effect of variation of the misclassification cost on performance of the binary classifier for saline/non-saline classification task.

Misclassification cost	T_p[a]	F_n[b]	F_p[c]	T_n[d]	Classification error	Accuracy (%)	Precision	Recall	MCC[e]
1.00	6,580	579	775	2,812	0.126	87.40	0.895	0.919	0.713
1.25	6,512	647	816	2,771	0.143	86.39	0.889	0.910	0.691
1.50	6,450	709	725	2,862	0.143	86.66	0.899	0.901	0.700
1.75	6,526	633	618	2,969	0.128	88.36	0.913	0.912	0.739
2.00	6,579	580	739	2,848	0.144	87.73	0.899	0.919	0.721
2.25	6,426	733	496	3,091	0.121	88.56	0.928	0.898	0.748
2.50	6,244	915	450	3,137	0.126	87.30	0.933	0.872	0.727
2.75	5,964	1,195	433	3,154	0.140	84.85	0.932	0.833	0.684
3.00	6,539	620	636	2,951	0.141	88.31	0.911	0.913	0.737
3.25	6,470	689	591	2,996	0.139	88.09	0.916	0.904	0.734
3.50	6,152	1,007	489	3,098	0.138	86.08	0.926	0.859	0.702
3.75	6,199	960	480	3,107	0.134	86.60	0.928	0.866	0.712
4.00	6,482	677	659	2,928	0.154	87.57	0.908	0.905	0.721

[a] True positive
[b] False negative
[c] False positive
[d] True negative
[e] Matthews Correlation Coefficient

Table A2-26: Effect of over-sampling of the under-represented class (saline class) using SMOTE technique and variation of the misclassification cost on performance of the binary classifier for saline/non-saline classification task.

Number of augmented samples	Misclassification cost	T_p	F_n	F_p	T_n	Classification error	Accuracy (%)	Precision	Recall	*MCC*
2,500	1.00	6,341	818	762	2,825	0.152	85.30	0.893	0.886	0.671
2,500	1.50	5,990	1,169	766	2,821	0.187	81.99	0.887	0.837	0.608
2,500	2.00	6,271	888	449	3,138	0.125	87.56	0.933	0.876	0.731
2,500	2.50	5,990	1,169	381	3,206	0.129	85.58	0.940	0.837	0.701
2,500	3.00	6,365	794	518	3,069	0.133	87.79	0.925	0.889	0.732
2,500	3.50	5,916	1,243	340	3,247	0.120	85.27	0.946	0.826	0.699
2,500	4.00	5,644	1,515	289	3,298	0.118	83.21	0.951	0.788	0.671
5,000	1.00	6,494	665	675	2,912	0.133	87.53	0.906	0.907	0.719
5,000	1.50	6,446	713	638	2,949	0.141	87.43	0.910	0.900	0.719
5,000	2.00	6,430	729	572	3,015	0.136	87.89	0.918	0.898	0.731
5,000	2.50	6,062	1,097	396	3,191	0.126	86.11	0.939	0.847	0.709
5,000	3.00	6,458	701	629	2,958	0.151	87.62	0.911	0.902	0.723
5,000	3.50	6,300	859	538	3,049	0.142	87.00	0.921	0.880	0.716
5,000	4.00	6,348	811	666	2,921	0.167	86.26	0.905	0.887	0.694
7,500	1.00	6,517	642	689	2,898	0.137	87.61	0.904	0.910	0.721
7,500	1.50	6,403	756	575	3,012	0.136	87.61	0.918	0.894	0.725
7,500	2.00	6,311	848	687	2,900	0.164	85.72	0.902	0.882	0.683
7,500	2.50	5,863	1,296	468	3,119	0.147	83.58	0.926	0.819	0.660
7,500	3.00	6,281	878	578	3,009	0.150	86.45	0.916	0.877	0.703
7,500	3.50	6,377	782	586	3,001	0.150	87.27	0.916	0.891	0.718
7,500	4.00	6,357	802	523	3,064	0.138	87.67	0.924	0.888	0.729
10,000	1.00	6,486	673	701	2,886	0.144	87.21	0.902	0.906	0.712
10,000	1.50	6,299	860	579	3,008	0.145	86.61	0.916	0.880	0.706
10,000	2.00	6,250	909	827	2,760	0.195	83.85	0.883	0.873	0.639
10,000	2.50	5,891	1,268	400	3,187	0.131	84.48	0.936	0.823	0.681
10,000	3.00	5,740	1,419	354	3,233	0.124	83.50	0.942	0.802	0.669
10,000	3.50	5,355	1,804	281	3,306	0.118	80.60	0.950	0.748	0.632
10,000	4.00	5,676	1,483	389	3,198	0.129	82.58	0.936	0.793	0.651

Table A2-27: Effect of random under-sampling from the under-represented class (saline class) and variation of the misclassification cost on performance of the binary classifier for saline/non-saline classification task.

Number of removed samples	Misclassification cost	T_p	F_n	F_p	T_n	Classification error	Accuracy (%)	Precision	Recall	*MCC*
2,500	1.00	6,592	566	692	2,896	0.120	88.29	0.905	0.921	0.735
2,500	1.50	6,381	777	555	3,033	0.130	87.60	0.920	0.891	0.726
2,500	2.00	6,186	972	478	3,110	0.134	86.51	0.928	0.864	0.710
2,500	2.50	5,629	1,529	616	2,972	0.189	80.04	0.901	0.786	0.588
2,500	3.00	6,107	1,051	451	3,137	0.134	86.02	0.931	0.853	0.703
2,500	3.50	5,226	1,932	267	3,321	0.140	79.54	0.951	0.730	0.619
2,500	4.00	6,505	653	584	3,004	0.141	88.49	0.918	0.909	0.743
5,000	1.00	6,436	722	633	2,955	0.131	87.39	0.910	0.899	0.718
5,000	1.50	5,859	1,299	674	2,914	0.185	81.64	0.897	0.819	0.609
5,000	2.00	6,037	1,121	394	3,194	0.130	85.90	0.939	0.843	0.706
5,000	2.50	6,260	898	527	3,061	0.139	86.74	0.922	0.875	0.711
5,000	3.00	6,073	1,085	467	3,121	0.137	85.56	0.929	0.848	0.694
5,000	3.50	6,260	898	544	3,044	0.144	86.58	0.920	0.875	0.707
5,000	4.00	6,197	961	443	3,145	0.126	86.93	0.933	0.866	0.720
7,500	1.00	6,256	902	499	3,089	0.132	86.96	0.926	0.874	0.717
7,500	1.50	6,050	1,108	479	3,109	0.143	85.23	0.927	0.845	0.687
7,500	2.00	5,644	1,514	431	3,157	0.156	81.90	0.929	0.788	0.636
7,500	2.50	6,155	1,003	473	3,115	0.135	86.26	0.929	0.860	0.706
7,500	3.00	5,623	1,535	582	3,006	0.178	80.30	0.906	0.786	0.595
7,500	3.50	5,115	2,043	261	3,327	0.130	78.56	0.951	0.715	0.605
7,500	4.00	5,285	1,873	232	3,356	0.113	80.41	0.958	0.738	0.636
10,000	1.00	6,006	1,152	502	3,086	0.151	84.61	0.923	0.839	0.675
10,000	1.50	5,722	1,436	402	3,186	0.149	82.90	0.934	0.799	0.655
10,000	2.00	5,989	1,169	454	3,134	0.139	84.90	0.930	0.837	0.684
10,000	2.50	5,279	1,879	256	3,332	0.129	80.13	0.954	0.737	0.629
10,000	3.00	5,194	1,964	247	3,341	0.123	79.42	0.955	0.726	0.620
10,000	3.50	5,713	1,445	322	3,266	0.116	83.56	0.947	0.798	0.673
10,000	4.00	6,034	1,124	491	3,097	0.141	84.97	0.925	0.843	0.682

Table A2-28: Effect of combined random under-sampling and over-sampling (using SMOTE technique) and variation of the misclassification cost on performance of the binary classifier for saline/non-saline classification task.

Number of removed samples	Number of augmented samples	Misclassification cost	T_p	F_n	F_p	T_n	Classification error	Accuracy (%)	Precision	Recall	*MCC*
1,500	1,500	1.00	6,521	637	610	2,978	0.120	88.40	0.914	0.911	0.740
1,500	1,500	1.50	6,267	891	571	3,017	0.141	86.39	0.916	0.876	0.702
1,500	1,500	2.00	5,914	1,244	468	3,120	0.150	84.07	0.927	0.826	0.668
1,500	1,500	2.50	5,942	1,216	453	3,135	0.143	84.47	0.929	0.830	0.676
1,500	1,500	3.00	6,144	1,014	452	3,136	0.131	86.36	0.931	0.858	0.709
1,500	1,500	3.50	5,584	1,574	574	3,014	0.179	80.01	0.907	0.780	0.591
1,500	1,500	4.00	6,256	902	433	3,155	0.122	87.58	0.935	0.874	0.733
1,500	1,500	4.50	6,299	859	493	3,095	0.133	87.42	0.927	0.880	0.726
1,500	1,500	5.00	6,147	1,011	484	3,104	0.136	86.09	0.927	0.859	0.702
3,000	3,000	1.00	6,387	771	677	2,911	0.142	86.53	0.904	0.892	0.699
3,000	3,000	1.50	6,157	1,001	562	3,026	0.149	85.46	0.916	0.860	0.685
3,000	3,000	2.00	6,100	1,058	534	3,054	0.148	85.19	0.920	0.852	0.682
3,000	3,000	2.50	5,761	1,397	390	3,198	0.139	83.37	0.937	0.805	0.664
3,000	3,000	3.00	5,449	1,709	290	3,298	0.130	81.40	0.949	0.761	0.643
3,000	3,000	3.50	5,551	1,607	323	3,265	0.127	82.04	0.945	0.775	0.649
3,000	3,000	4.00	5,703	1,455	591	2,997	0.174	80.96	0.906	0.797	0.605
3,000	3,000	4.50	5,544	1,614	340	3,248	0.125	81.82	0.942	0.775	0.644
3,000	3,000	5.00	6,145	1,013	540	3,048	0.149	85.55	0.919	0.858	0.689
4,500	4,500	1.00	6,299	859	551	3,037	0.136	86.88	0.920	0.880	0.713
4,500	4,500	1.50	6,081	1,077	471	3,117	0.139	85.59	0.928	0.850	0.694
4,500	4,500	2.00	6,057	1,101	592	2,996	0.161	84.25	0.911	0.846	0.661
4,500	4,500	2.50	5,355	1,803	313	3,275	0.138	80.31	0.945	0.748	0.624
4,500	4,500	3.00	6,029	1,129	411	3,177	0.126	85.67	0.936	0.842	0.700
4,500	4,500	3.50	5,372	1,786	309	3,279	0.126	80.50	0.946	0.750	0.628
4,500	4,500	4.00	6,155	1,003	554	3,034	0.151	85.51	0.917	0.860	0.687
4,500	4,500	4.50	6,006	1,152	630	2,958	0.173	83.42	0.905	0.839	0.644
4,500	4,500	5.00	6,069	1,089	484	3,104	0.138	85.36	0.926	0.848	0.689
6,000	6,000	1.00	6,214	944	543	3,045	0.142	86.16	0.920	0.868	0.700
6,000	6,000	1.50	5,702	1,456	470	3,118	0.159	82.08	0.924	0.797	0.635
6,000	6,000	2.00	5,631	1,527	366	3,222	0.137	82.38	0.939	0.787	0.650
6,000	6,000	2.50	5,361	1,797	331	3,257	0.135	80.20	0.942	0.749	0.620
6,000	6,000	3.00	6,108	1,050	559	3,029	0.154	85.03	0.916	0.853	0.678
6,000	6,000	3.50	5,021	2,137	234	3,354	0.114	77.94	0.955	0.701	0.600
6,000	6,000	4.00	5,844	1,314	725	2,863	0.199	81.03	0.890	0.816	0.594
6,000	6,000	4.50	6,079	1,079	604	2,984	0.165	84.34	0.910	0.849	0.662
6,000	6,000	5.00	6,020	1,138	441	3,147	0.129	85.31	0.932	0.841	0.692

One of the challenges in fitting the classification and regression models to target variables was optimisation of the hyperparameters. A hyperparameter is a parameter whose value should be set before launching the training process of a machine learning model. To handle this during the model training, we used MATALB hyperparameter optimiser which applies Bayesian optimisation algorithm to estimate the optimal hyperparameters. Since the Bayesian optimisation algorithm used for optimizing the objective function depends on the runtime (it avoids the areas with high run time in each iteration), results of the hyperparameter tuning jobs were not reproducible. Therefore, according to our computational resources we repeated the trainings 30 times and acquired confidence intervals for the mean of optimised hyperparameters using bootstrapping technique; we used bootstrapping because it was not possible to determine the exact distribution of optimised hyperparameters with only 30 iterations. We applied 10-fold cross validation to calculate the accuracy metrics for trained models and similarly the confidence intervals of mean was reported to show the performance of the models. The following MATLAB code was used to fit an ensemble of classification trees on training datasets, optimise hyperparameters, and bootstrapping the results to calculate 95% confidence intervals of the mean:

```
clc;
clear;
%% Training ensemble of regression trees for predicting ECe or ESP,
%% Usage: This script returns the tuned fitcensemble hyperparameters and
%% accuracy metrics calculated on the holdout set for 30 iterations using fitcensemble
%% function in order to calculate the confidence intervals using bootstrapping technique.
%% Due to non-repeatable nature of the hyperparameter optimisation jobs, training
%% jobs are repeated 30 times and using bootstrapping, 95% confidence intervals for
%% hyperparameters and accuracy metrics are calculated. Accuracy metrics include: Binomial
%% deviance loss, Misclassification error, Accuracy,Precision, Recall, and MCC (Matthews
%% Correlation Coefficient, useful for binary imbalanced classification).

%% Classification using ensemble of trees (fitcensemble)
% Here ECe is the target variable; however for ESP, all ECe values should be replaced by ESP

% Importing the original dataset
ECe = readtable(Location of the training datasets','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);

% Pre-processing the original dataset
table = standardizeMissing(ECe,-9999);
table.FID = [];
table.Year = [];
table(sum(ismissing(table),2) > 0,:) = [];% Dropping the rows with missing values
% Categorizing the categorical variables in the training set
table.Main_litho = categorical(table.Main_litho);
table.WRB = categorical(table.WRB);
table.LC = categorical(table.LC);
% Classifying the ECe values
edges = [0 2 100];  % for ESP: edges = [0 1 100]
table.ECe = discretize(table.ECe,edges);

% Pre-allocating memory to variables with increasing size in each iteration
Num_learning_cycles = zeros(30,1); Learn_rate = zeros(30,1); Min_leaf_size = zeros(30,1);
Max_num_splits = zeros(30,1); Num_variables_to_sample = zeros(30,1); Binomial_deviance_loss =
zeros(30,1);
Mis_classification_error = zeros(30,1); Accuracy = zeros(30,1); Precision = zeros(30,1);
Recall = zeros(30,1); MCC  = zeros(30,1); MinObjective = zeros(30,1);

% Training: This loop repeats the fitting of the classification model 30 times
for i = 1:30
    % We used holdout method with the maximum 130 objective function evaluations to
    % optimise the ensemble hyperparameters
    % 'ens' is the object of the final trained model
    % The misclassification cost for saline class is set to be 2
    % Note the misclassification cost matrix was the default for ESP training dataset
    ens = fitcensemble(table,ECe','Cost',[0 1;2 0],
'ScoreTransform','logit','OptimizeHyperparameters',...
{'Method','LearnRate','NumLearningCycles','MinLeafSize','MaxNumSplits','NumVariablesToSample',
'SplitCriterion'},'HyperparameterOptimisationOptions',struct('Holdout',.25,'UseParallel',true,
'MaxObjectiveEvaluations',130,'ShowPlots',true,'Repartition',true));
```

```
    % Saving the created model objects
    save(strcat('output folder\ens_',num2str(i)),'ens');
end
Truelabels = table.ECe;
% This loop cross-validates the fitted models using 10-fold cross validation technique
parfor i = 1:30
    % Loading the saved model objects
    ens = load(strcat('Output location from the previous loop\ens_',num2str(i)));
    % Acquiring hyperparameter tuning job results
    MinObjective(i,1) = ens.ens.HyperparameterOptimisationResults.MinObjective;
    Num_learning_cycles(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,2));
    Learn_rate(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,3));
    Min_leaf_size(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,4));
    Max_num_splits(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,5));
    Num_variables_to_sample(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,7));
    % Validation and Acquiring accuracy metrics
    cvens = crossval(ens.ens,'Kfold',10);
    Predictedlabels = kfoldPredict(cvens);
    C = confusionmat(Truelabels,Predictedlabels);
    % Tp = True Positive, Fn = False Negative , Fp = False Positive, Tn = True Negative
    Tp = C(1,1); Fn = C(1,2); Fp = C(2,1); Tn = C(2,2);
    Binomial_deviance_loss(i,1) = kfoldLoss(cvens,'Lossfun','binodeviance');
    Mis_classification_error(i,1) = kfoldLoss(cvens);% Binary misclassification loss
    Accuracy(i,1) = (Tp+Tn)/(Tp+Fp+Fn+Tn)*100;% Binary classification accuracy
    Precision(i,1) = Tp/(Tp+Fp);% Precision
    Recall(i,1) = Tp/(Tp+Fn);% Recall
    % MCC (Matthews Correlation Coefficient) for binary imbalanced classification
    MCC(i,1) = ((Tp*Tn)-(Fp*Fn))/sqrt((Tp+Fp)*(Tp+Fn)*(Tn+Fp)*(Tn+Fn));
end

% Exporting the output into a table
Statistics = [Num_learning_cycles Learn_rate Min_leaf_size Max_num_splits
Num_variables_to_sample Binomial_deviance_loss Mis_classification_error Accuracy Precision
Recall MCC MinObjective];
Statistics_table = array2table(Statistics,'VariableNames',{'Num_learning_cycles'...
     'Learn_rate' 'Min_leaf_size' 'Max_num_splits' 'Num_variables_to_sample'...
     'Binomial_deviance_loss' 'Mis_classification_error' 'Accuracy' 'Precision'...
     'Recall' 'MCC' 'MinObjective'});

% Saving the obtained statistics
writetable(Statistics_table, 'Output directory\output file name.txt');

%% Bootstrapping

% Computing the 95% confidence intervals of the mean for the statistics calculated in the
% above loop using 1000 bootstrap iterations. bootci creates each bootstrap sample by sampling
% with replacement from the rows of the data arguments and computes the confidence interval by
% bias corrected and accelerated percentile method

opt = statset('UseParallel',true);
ci = bootci(1000,{@nanmean,Statistics},'type','bca','Options',opt);
ci = array2table(ci,'VariableNames',{'Num_learning_cycles' 'Learn_rate' 'Min_leaf_size'...
     'Max_num_splits' 'Num_variables_to_sample' 'Binomial_deviance_loss'...
     'Mis_classification_error' 'Accuracy' 'Precision' 'Recall' 'MCC' 'MinObjective'});

% Exporting the output into a table
writetable(ci, 'Output directory\output file name.txt');
```

Likewise, the following script in MATLAB was used to train an ensemble of regression trees (using MATLAB “fitrensemble” function) on each class of training datasets, optimise function’s hyperparameters, and bootstrapping the results to calculate 95% confidence intervals of the mean for hyperparameters and accuracy metrics:

```
clc;
clear;
%% Fitting an ensemble of regression trees to each class,
% This script returns the tuned fitrensemble hyperparameters and
```

```
% accuracy metrics calculated by 10-fold cross validation for 30 iterations using fitrensemble
% function and calculates the confidence intervals for hyperparameters using bootstrapping
% technique. Hyperparameters tuning job is conducted after log-transformation of the response
% variable. Accuracy metrics including mean squared error (mse), mean absolute error (mae),
% and NSE which ia a specifc definition of coefficient of determination (shown as R_squared in
% this script) are computed for both logarithmic and non-logarithmic spaces.


%% Regression using ensemble of trees (fitrensemble)
% Note this is regression using the ensemble of trees (fitrensemble) on ECe as
% a target variable and for ESP, all variables shown by ECe should be replaced by ESP

% Importing the original dataset
ECe = readtable('Training set file location','FileType',...
                                'text','Delimiter',',','PreserveVariableNames',true);

% Pre-processing the original dataset
table = standardizeMissing(ECe,-9999);
table.FID = [];
table.Year = [];
table(sum(ismissing(table),2) > 0,:) = [];% Dropping the rows with missing values
% Classifying the table into two parts based on the values of target variable
edges = [0 2 100]; % edges = [0 1 100] for ESP
table.W = discretize(table.ECe,edges); % W would be 1 or 2 and is indicator of class
table = table(table.W(:) == 2,:); % Removing the first class (non-saline) to do the regression
% job on saline class; Similarly by setting table = table(table.W(:) == 1,:), second class
% (saline) can be removed and regression can be done on the remaining class
table.W = [];
table.ECe = log10(table.ECe); Transforming the target variable to logarithmic scale; for
% regression on the non-saline class, logarithm transformation of the values that
% are 0 (zero) is not possible. Therefore, first a constant
% (here one) should be added to the target variable values and then transform the target to
% the logarithmic scale (table.ECe = log10(table.ECe + 1))
% Categorizing the categorical variables in the training set
table.Main_litho = categorical(table.Main_litho);
table.WRB = categorical(table.WRB);
table.LC = categorical(table.LC);

% Pre-allocating memory to variables with increasing size in each iteration
Num_learning_cycles = zeros(30,1); Learn_rate = zeros(30,1); Min_leaf_size = zeros(30,1);
Max_num_splits = zeros(30,1); Num_variables_to_sample = zeros(30,1);
mse_log = zeros(30,1);mae_log = zeros(30,1); R_squared_log = zeros(30,1);
mse = zeros(30,1);mae = zeros(30,1);R_squared = zeros(30,1);MinObjective = zeros(30,1);

for i = 1:30
    % Training and hyperparameter tuning job
    % We used holdout method (25% held out) with 130 objective function evaluations to
    % optimise the ensemble hyperparameters
    % 'ens' is the object of the final trained model
    ens = fitrensemble(table,'ECe','Method','LSBoost','OptimizeHyperparameters',...
     {'NumLearningCycles','LearnRate','MinLeafSize','MaxNumSplits','NumVariablesToSample'},...
      'HyperparameterOptimisationOptions',struct('Holdout',.25,'UseParallel',true,...
      'MaxObjectiveEvaluations',130,'Repartition',true,'ShowPlots',true,'Verbose',1));
    % Acquiring and saving the hyperparameter tuning job results on the disk
    save(strcat('Output directory\ens_',... num2str(i)),'ens');
end

% Validation and Acquiring accuracy metrics
ytrue_log = table.ECe;
% Back-transformation of the predicted values from logarithmic scale
ytrue = 10.^(table.ECe); % ytrue = 10.^(table.ECe)-1 for the non-saline class
% This loop cross-validates the fitted models using 10-fold cross validation
parfor i = 1:30
    % Loading the saved model objects
    ens =
    load(strcat('Output location from the previous loop\ens_'\ens_',...
    num2str(i)));
    % Acquiring hyperparameter tuning job results
    MinObjective(i,1) = ens.ens.HyperparameterOptimisationResults.MinObjective;
    Num_learning_cycles(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,1));
    Learn_rate(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,2));
    Min_leaf_size(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,3));
    Max_num_splits(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,4));
```

```
    Num_variables_to_sample(i,1) = ...
table2array(ens.ens.HyperparameterOptimisationResults.XAtMinObjective(1,5));
    % Validation of the trained ensembles using 10-fold cross validation
    cvens = crossval(ens.ens,'Kfold',10);
    mse_log(i,1) = kfoldLoss(cvens,'mode','average');% Mean square error in logarithmic scale
    yfit_log = kfoldPredict(cvens);
    mae_log(i,1) = mean(abs(yfit_log - ytrue_log));% Mean absolute error in logarithmic scale%
    % NSE in the logarithmic scale
    R_squared_log(i,1) = 1 - sum((ytrue_log - yfit_log).^2)/sum((ytrue_log - ...
    mean(ytrue_log)).^2);
    yfit = 10.^(yfit_log); % ytrue = 10.^(table.ECe)-1 for the non-saline class
    mse(i,1) = mean((ytrue - yfit).^2); % Mean square error
    mae(i,1) = mean(abs(yfit - ytrue)); % Mean absolute error
    R_squared(i,1) = 1 - sum((ytrue - yfit).^2)/sum((ytrue - mean(ytrue)).^2);
    % NSE
end
rmse_log = sqrt(mse_log);% Root mean square error in the logarithmic scale
rmse = sqrt(mse);% Root mean square error

% Exporting the output into a table
Statistics = [Num_learning_cycles Learn_rate Min_leaf_size Max_num_splits
Num_variables_to_sample mse_log rmse_log mae_log R_squared_log mse rmse mae R_squared
MinObjective];
Statistics_table = array2table(Statistics,'VariableNames',{'Num_learning_cycles' 'Learn_rate'
     'Min_leaf_size' 'Max_num_splits' 'Num_variables_to_sample' 'mse_log' 'rmse_log'...
     'mae_log' 'R_squared_log' 'mse' 'rmse' 'mae' 'R_squared' 'MinObjective'});

% Saving the table on the disk
writetable(Statistics_table, 'Output directory\output file name.txt');

%% Bootstrapping

% Computing the 95% confidence intervals of the mean for the statistics calculated in
% above loop using 1000 bootstrap iterations. bootci creates each bootstrap sample by sampling
% with replacement from the rows of the data arguments and computes the confidence interval by
% bias corrected and accelerated percentile method.
opt = statset('UseParallel',true);
ci = bootci(1000,{@nanmean,Statistics},'type','bca','Options',opt);
ci = array2table(ci,'VariableNames',{'Num_learning_cycles' 'Learn_rate' 'Min_leaf_size'...
    'Max_num_splits' 'Num_variables_to_sample' 'mse_log' 'rmse_log'...
    'mae_log' 'R_squared_log' 'mse' 'rmse' 'mae' 'R_squared' 'MinObjective'});

%Exporting the output into a table
writetable(ci, 'Output directory\output file name.txt');
```

For both classification and regression jobs, increasing the number of weak learners (number of learning cycles) did not improve the performance of the ensembles. Among trained classifiers, the one with highest *MCC* and among regressions within each class, the one with highest *NSE* were selected for the rest of analysis, which means a total of six models, two for classification and four for per-class regression jobs.

Generation of soil mask and spatio-temporal predictions

Through applying the trained models to a global soil mask, we created the global maps of surface soil salinity and sodicity (0 - 30 cm) at 0.008333333° spatial resolution. To create that soil mask, first we re-projected the 2014 MODIS land cover map from sinusoidal coordinates system to WGS 1984 using the ArcMap “raster project” tool. During the re-projection, we also resampled the map (with 0.004° spatial resolution) to 0.008333333° resolution (which was our desirable resolution) by the nearest neighbour resampling method to minimise the data loss during resampling and re-projection steps. Then we masked out the pixels labelled as water bodies, permanent wetlands, urban and built-up lands, and permanent snow and ice (numbers 11, 13, 15, and 17 in the map’s IGBP legend) using the “mask function” available in ArcMap image analysis window. During exporting the generated layer, the lower and upper extents were set to be -55 and 55, respectively. Using ArcMap “raster split tool”, we split the soil mask to smaller tiles so that the final smaller rasters were of maximum 3,600 pixels (60 rows and 60 columns). A total of 50,687 raster tiles were generated and converted to point feature classes in the World Mercator coordinates system using the following Python script:

```
## Raster to point conversion, we used PyCharm Python IDE (Integrated Development Environment)
## Usage: This code converts raster layers to point feature layers.

import arcpy # Importing the ArcPy module
import multiprocessing #Importing multi-processing module
from multiprocessing import Process
import os # Importing Miscellaneous operating system module required for reading the file
names in a directory
import os, fnmatch

# Setting the geo-processing environments
arcpy.env.overwriteOutput = True
arcpy.env.workspace = r"Directory of the raster tiles created form splitting job"
arcpy.env.extent = "MAXOF"
# Setting the output coordinates system as World Mercator
arcpy.env.geographicTransformations = arcpy.SpatialReference(54004)

# Reading all raster files in a directory
path = r" Directory of the raster tiles created form splitting job "
pattern = "*.tif"
Tiffs = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]

# Defining the function that will be passed to child processes
# Function arguments: ini, end
def cell(ini,end):
 jj = range(ini,end)# Indicator of the raster file
 for j in jj:
   inRaster = Tiffs[j]
   string = Tiffs[j]
   outPoint = "Output directory of point feature layers/"+string[0:4]+"_"+str(j)+".shp"
   field = "VALUE"
   # Execution of the ArcPy raster to point tool
   arcpy.RasterToPoint_conversion(inRaster, outPoint, field)

if __name__ == '__main__':
 count = 0
 processes = []
 for i in range(0,number of system cores):
   ini = count
   end = count + The number of rasters that should be converted by each core
   process = Process(target=cell, args=(ini,end,))
   processes.append(process)
   process.start()
   count = end
```

The output point feature classes were in the World Mercator coordinates system. The values of static predictors in the World Mercator coordinates system were then extracted to the points in the generated point feature classes as follows:

```
# Extracting static raster values to points in World Mercator coordinates system,
# Usage: Extracts the cells' values of multiple rasters as attributes in
# the output point feature classes. Requirements: Spatial Analyst Extension.

import arcpy # Importing the ArcPy module
from arcpy import env
from arcpy.sa import *
import multiprocessing #Importing multi-processing module
from multiprocessing import Process
import time
import os
import os, fnmatch

# Setting the geo-processing environments
arcpy.CheckOutExtension("Spatial")
arcpy.env.extent = "MAXOF"
arcpy.env.overwriteOutput = True
arcpy.env.workspace = r"Location of the point feature layers" # Raster layers must be in the
# same directory

# Reading all point feature layers in a directory
path = r" Location of the point feature layers "
```

```
pattern = "*.shp"
shape_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]

# Defining the function that will be passed to child processes
# Function arguments: ini, end
def cell(ini,end):
 jj = range(ini,end)
 for j in jj:
   # j is the indicator for shape files
   inRasterList = [["Global_DEM.tif","Elevation"],["Global_Plan.tif","Pla_cur"],
   ["Global_Profile.tif","Pro_cur"],["Global_Slope.tif","Slope"],
   ["Slope Length.sdat","Slope_Leng"],["Global_Terrain Ruggedness Index (TRI).sdat","TRI"]]
   # The value of the cell will be calculated from the adjacent cells with
   # valid values using bilinear interpolation
   inPointFeatures = shape_files[j]
   ExtractMultiValuesToPoints(inPointFeatures, inRasterList, "BILINEAR")

if __name__ == '__main__':
 processes = []
 count = 0
 for i in range(0,number of system cores):
   # To do each child process in a different temporary folder:
   time.sleep(1.1)
   newTempDir = r"C:\temp\gptmpenvr_" + time.strftime('%Y%m%d%H%M%S') + str(i)
   os.mkdir(newTempDir)
   os.environ["TEMP"] = newTempDir
   os.environ["TMP"] = newTempDir
   ini = count
   end = count + The number of point features that should be processed by each core
   process = Process(target = cell, args = (ini,end,))
   processes.append(process)
   process.start()
   count = end
```

After extracting the static Mercator predictors' values, the point feature classes were again projected to the WGS 1984 coordinates system using the following Python script:

```
# Projecting point feature classes to WGS 1984 geographic coordinates.

import multiprocessing # Importing the ArcPy module
from multiprocessing import Process #Importing multi-processing module
import arcpy
import os
import os, fnmatch

# Setting the geo-processing environments
arcpy.env.workspace = r"Location of the point feature classes"

# Reading all point feature layers in a directory
path = r" Location of the point feature classes "
pattern = "*.shp"
shape_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]
# Setting the output coordinates system object (WGS 1984)
sr = arcpy.SpatialReference(4326)

# Defining the function that will be passed to child processes
# Function arguments: ini, end
def cell(ini,end):
  jj = range(ini,end)
  for j in jj:
    # j is the indicator for shape files
    string = shape_files[j]
    output_feature_class = r"Location of the projected point feature classes/"+string+".shp"
    arcpy.Project_management(shape_files[j], output_feature_class, sr)

if __name__ == '__main__':
 processes = []
 count = 0
 for i in range(0,number of system cores):
   ini = count
   end = count + The number of point feartures that should be re-projected by each core
   process = Process(target = cell, args = (ini,end,))
   processes.append(process)
   process.start()
   count = end
```

Similar to extraction of the rasters values to points in the Mercator coordinates, we drew the information from the rasters in geographic coordinates system and attributed to the points. For static predictors in the WGS 1984:

```
# Extract static predictors’ raster values to points in WGS 1984 coordinate system,
# Usage: Extracts the cells’ values of multiple rasters as attributes in
# the output point feature classes. Requirements: Spatial Analyst Extension.

import arcpy # Importing the ArcPy module
from arcpy import env
from arcpy.sa import *
import multiprocessing # Importing multi-processing module
from multiprocessing import Process
import time
import os
import os, fnmatch

# Setting the geo-processing environments
arcpy.env.workspace = r"Directory of the point feature classes"
arcpy.CheckOutExtension("Spatial")
arcpy.env.extent = "MAXOF"

# Reading all point feature layers in a directory
path = r" Directory of the point feature classes "
pattern = "*.shp"
shape_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]

# Defining the function that will be passed to child processes
# Function arguments: ini, end
def cell(ini,end):
 for j in range(ini,end):
  # j is the indicator for shape files
  inRasterList1 = [["fertl_c3ann_Layer.tif","c3ann"],["fertl_c3per_Layer.tif","c3per"],
  ["Global_water_table.tif","WTD"],
  ["Global_Aspect.tif","Aspect"],["Topographic_index.tif","Topo_index"],
  ["Clay.tif","Clay"],
  ["Silt.tif","Silt"],["Sand.tif","Sand"],
  ["average_soil_and_sedimentary-deposit_thickness.tif","Soil_thick"],
  ["95ecosys_rootdepth.tif","95_Root_D"]]
  inRasterList2 = [["WRB.tif","WRB"],["Main_lithological_units_geographic.tif","Main_litho"]]
  inPointFeatures = shape_files[j]
  # The value of the cell will be calculated from the adjacent cells with valid values using
  # bilinear interpolation
  ExtractMultiValuesToPoints(inPointFeatures,inRasterList1,"BILINEAR")
  # No interpolation will be applied to the categorical variables
  ExtractMultiValuesToPoints(inPointFeatures,inRasterList2,"NONE")

if __name__ == '__main__':
 processes = []
 count = 0
 for i in range(0,number of system cores):
       # To do each child process in a different temporary folder
       time.sleep(1.1)
       newTempDir = r"C:\temp\gptmpenvr_" + time.strftime('%Y%m%d%H%M%S') + str(i)
       os.mkdir(newTempDir)
       os.environ["TEMP"] = newTempDir
       os.environ["TMP"] = newTempDir
       ini = count
       end = count + The number of point features that should be processed by each core
       process = Process(target = cell, args = (ini,end,))
       processes.append(process)
       process.start()
       count = end
```

And for extraction of dynamic predictors in the WGS 1984 to attribute tables of the point feature classes:

```
# Extract dynamic predictors’ raster values to points in WGS 1984 coordinate system,
# Usage: Extracts the cells’ values of multiple rasters as attributes in
# the output point feature classes. Requirements: Spatial Analyst Extension.

import arcpy # Importing the ArcPy module
from arcpy import env
from arcpy.sa import *
```

```
import multiprocessing # Importing multi-processing module
from multiprocessing import Process
import time
import os
import os, fnmatch

# Setting the geo-processing environments
arcpy.env.workspace = r" Directory of the point feature classes"
arcpy.CheckOutExtension("Spatial")
arcpy.env.extent = "MAXOF"

# Reading all point feature layers in a directory
path = r" Directory of the point feature classes"
pattern = "*.shp"
shape_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]

# Defining the function that will be passed to child processes
# Function arguments: ini, end
def cell(ini,end):
 for j in range(ini,end):
  # j is the indicator for shape files
  inPointFeatures = shape_files[j]
  Year = 1
  for i in range(1980,2019):
    # i is the indicator for year
    inRasterList1 = [["dtr_mean_"+str(i)+".tif","dtr_"+str(Year)],
    ["pre_mean_"+str(i)+".tif","Pre_"+str(Year)],
    ["tmp_mean_"+str(i)+".tif","T_ave_"+str(Year)],
    ["tmx_mean_"+str(i)+".tif","T_max_"+str(Year)],
    ["tmn_mean_"+str(i)+".tif","T_min_"+str(Year)],
    ["Soil_moisture_mean_"+str(i)+".tif","S_mo_"+str(Year)],
    ["PDSI_mean_"+str(i)+".tif","PDSI_"+str(Year)],
    ["SM_"+str(i)+"_smoothed.tif","Sat_SM_"+str(Year)],
    ["Gleam_S_"+str(i)+".tif","Gleam_S_"+str(Year)],
    ["Mod_EVI_"+str(i)+".tif","EVI_"+str(Year)],
    ["Mod_NDVI_"+str(i)+".tif","NDVI_"+str(Year)],
    ["FAPAR_"+str(i)+"_smoothed.tif","FAPAR_"+str(Year)],
    ["LAI_"+str(i)+"_smoothed.tif","LAI_"+str(Year)],
    ["WS"+str(i)+".tif","Wind_S_"+str(Year)],
    ["Skin_temp_"+str(i)+".tif","Skin_T_"+str(Year)],
    ["Soiltemp1"+str(i)+".tif","S_T_1_"+str(Year)],
    ["Soiltemp2_"+str(i)+".tif","S_T_2_"+str(Year)],
    ["Soiltemp3_"+str(i)+".tif","S_T_3_"+str(Year)],
    ["Soiltemp4"+str(i)+".tif","S_T_4_"+str(Year)],
    ["pet_mean_"+str(i)+".tif","Pet_"+str(Year)],
    ["def_mean_"+str(i)+".tif","Def_"+str(Year)],
    ["aet_mean_"+str(i)+".tif","aet_"+str(Year)]]
    inRasterList2 = [["Land_cover_"+str(i)+".tif","LC_"+str(Year)]]
    # The value of the cell will be calculated from the adjacent cells with valid values
    # using bilinear interpolation
    ExtractMultiValuesToPoints(inPointFeatures,inRasterList1,"BILINEAR")
    # No interpolation will be applied to the categorical variables
    ExtractMultiValuesToPoints(inPointFeatures,inRasterList2,"NONE")
    Year = Year+1

if __name__ == '__main__':
 # To do each child process in a different temporary folder
 count = 0
 processes = []
 for i in range(0, number of system cores):
   time.sleep(1.1)
   newTempDir = r"C:\temp\gptmpenvr_" + time.strftime('%Y%m%d%H%M%S') + str(i)
   os.mkdir(newTempDir)
   os.environ["TEMP"] = newTempDir
   os.environ["TMP"] = newTempDir
   ini = count
   end = count + The number of point features that should be processed by each core
   process = Process(target = cell, args = (ini,end,))
   processes.append(process)
   process.start()
   count = end
```

To more efficiently handle the size of the point feature classes and increasing the predictors' value extraction speed, we copied all 50,687 files (with WGS 1984 projection) in four different directories and in each directory, we extracted the values of the dynamic predictors for a decade since 1980. 1980 - 1987 in the first directory, 1988 - 1998 in the second, 1999 - 2008 in the third, and 2009 - 2018 in the fourth directory. Data of static predictors in the geographic coordinates (WGS 1984) were attributed to the point feature classes in the first directory. So in total, we had five point feature classes for each spatial tile generated from the soil mask: one with data of static predictors in the Mercator coordinates, and four for the rest of predictors in the WGS 1984. The attribute tables of point feature layers were then merged, data of each individual year was extracted, and for each spatial tile, 39 comma delimited tables in .txt format were exported to 39 different directories using the following script in MATLAB (Mapping Toolbox license is required for MATLAB “shaperead” function):

```
clc;
clear;
%% Shape file to text file converter,
%% The global soil mask was split to 50,687 tiles and the smaller tiles
%% were then converted to the point feature layers to extract the values of the predictors
%% at each pixel. This script converts the attribute tables of the shapefiles (n = 50,687)
%% to comma delimited tables with .txt format importable by MATLAB for
%% further processing and making predictions.

% Setting the directory of shapefiles
Shape_files_Merc = dir('Directory of point feature classes in the World Mercator...
system\*.shp');
Shape_files_1 = dir('First directory \*.shp');
Shape_files_2 = dir('Second directory \*.shp');
Shape_files_3 = dir('Third directory \*.shp');
Shape_files_4 = dir('Fourth directory \*.shp');

% Reading the shapefiles and merging, we kept X and Y coordinate values only from the first
table in geographic coordinates system (WGS 1984)
parfor i = 1:50687
    S_Merc = shaperead(strcat(Shape_files_Merc(i).folder,'\',Shape_files_Merc(i).name));
    T_Merc = struct2table(S_Merc);
    T_Merc.Geometry = []; T_Merc.X = []; T_Merc.Y = []; T_Merc.pointid = [];
    T_Merc.grid_code  = [];
    S_1 = shaperead(strcat(Shape_files_1(i).folder,'\',Shape_files_1(i).name));
    T_1 = struct2table(S_1);
    T_1.Geometry = []; T_1.pointid = []; T_1.grid_code = [];
    S_2 = shaperead(strcat(Shape_files_2(i).folder,'\',Shape_files_2(i).name));
    T_2 = struct2table(S_2);
    T_2.Geometry = []; T_2.X = []; T_2.Y = []; T_2.pointid = []; T_2.grid_code = [];
    S_3 = shaperead(strcat(Shape_files_3(i).folder,'\',Shape_files_3(i).name));
    T_3 = struct2table(S_3);
    T_3.Geometry = []; T_3.X = []; T_3.Y = []; T_3.pointid = []; T_3.grid_code = [];
    S_4 = shaperead(strcat(Shape_files_4(i).folder,'\',Shape_files_4(i).name));
    T_4 = struct2table(S_4);
    T_4.Geometry = []; T_4.X = []; T_4.Y = []; T_4.pointid = []; T_4.grid_code = [];
    table = [T_Merc T_1 T_2 T_3 T_4];

    % Extracting data of individual years from 1980
    for j = 1:39
            T = table(:,{'Elevation','Pla_cur','Pro_cur','Slope',...
            'Slope_Leng','TRI','X','Y','c3ann','c3per','WTD',...
            'Aspect','Topo_index','Clay','Silt','Sand','Soil_thick',...
            'x95_Root_D','WRB','Main_litho',...
            strcat('dtr_',num2str(j)),strcat('Pre_',num2str(j)),...
            strcat('T_ave_',num2str(j)),strcat('T_max_',num2str(j)),...
            strcat('T_min_',num2str(j)),strcat('S_mo_',num2str(j)),...
            strcat('PDSI_',num2str(j)),strcat('Sat_SM_',num2str(j)),...
            strcat('Gleam_S_',num2str(j)),strcat('EVI_',num2str(j)),...
            strcat('NDVI_',num2str(j)),strcat('FAPAR_',num2str(j)),...
            strcat('LAI_',num2str(j)),strcat('Wind_S_',num2str(j)),...
            strcat('Skin_T_',num2str(j)),strcat('S_T_1_',num2str(j)),...
            strcat('S_T_2_',num2str(j)),strcat('S_T_3_',num2str(j)),...
            strcat('S_T_4_',num2str(j)),strcat('Pet_',num2str(j)),...
            strcat('Def_',num2str(j)),strcat('aet_',num2str(j)),strcat('LC_',num2str(j))});
            T.Properties.VariableNames = {'Elevation','Pla_cur','Pro_cur','Slope',...
            'Slope_Leng','TRI','X','Y','c3ann','c3per','WTD',...
            'Aspect','Topo_index','Clay','Silt','Sand','Soil_thick',...
            'dtr','Pre','T_ave','T_max','T_min',...
            'S_mo','PDSI','Sat_SM','Gleam_S','EVI','NDVI','FAPAR','LAI','Wind_S','Skin_T',...
```

```
            'S_T_1','S_T_2','S_T_3','S_T_4','Pet','Def','aet','LC'};
            % Saving the final extracted tables in 39 different directories
            % representing 39 years since 1980
            writetable(T,strcat('The output directory on the disk\',num2str(1979 + j),'\',...
            num2str(1979 + j),'_',num2str(i),'.txt'));
   end
end
```

To create 39 folders in Microsoft Windows 10 (the name of each folder was the number of the corresponding year), we used the following code in the command prompt:

```
@echo off
Driver’s name (A capitalized letter like C or D):
for /l %%i in (1980,1,2018) do (md The desirable directory\%%i\%%j)
```

Model deployment

Following the extraction of predictors’ values to points (we needed to make predictions of the soil salinity/sodicity for those points), we had 39 folders (representing 1980 to 2018) and in each folder there were 50,687 tables saved in .txt format. Each individual table was representative of a spatial tile created from the original soil mask, with 43 columns (2 for the x- and y- coordinates and 41 for the predictor values) and maximum of 3,600 rows. x- y- values were the coordinates of grids in the WGS 1984 coordinates system. For all observations, the sample’s upper and lower depths were set to be zero and 30, respectively.

Tree based regression and classification models can handle the missing data by default; however, predictions for rows, which had more than five missing values of predictors were set as no data value. The indicator of no data value was 255. Each row in the tables was representative of a pixel from the original soil mask raster layer. The predictions made by models for each row, in addition to x- and y- values were later used to generate raster layers. The spatial resolution of these rasters was the same as original soil mask layer. We needed the area of each pixel to do zonal statistics and computing the area of different salinity classes at the country, biome, land cover, and climate levels. We directly calculated the area of each pixel in the WGS 1984 coordinates system from the x- and y- coordinates of input tables.

Additionally, for the classification step of the two-part models, we produced pixel-level scaled Shannon Entropy Index (Shannon 1949) (H_s) to identify the certainty of the classifier in binary prediction of classes. For each particular class, the binary classifier returns a score indicating the probability that the predicted label comes from that class and the final predicted label is the class with the highest score. We transformed these scores to probability (a value between zero and one) using the “logit” transformation and computed the binary H_s by $H_s = -(p\ (1) \times log_2 p\ (1) + p\ (2) \times log_2 p\ (2))$; where $p\ (1)$ and $p\ (2)$ were the per-class probabilities and log_2 was the logarithm with base 2. H_s shows the ambiguity in the model predictions and is a different concept from the validity of the predictions. Even with a zero H_s, the predicted labels can be false and therefore, H_s must not be used instead of the accuracy metrics for inspecting the validity of model predictions. The following script shows how we deployed the trained models to the tables and calculated each pixel’s H_s and area in MATLAB:

```
clc;
clear;
%% Model deployment for making predictions from the new data,
%% This code gets tables of predictors as an input and returns vectors of the predicted values
%% and X and Y coordinates for each row. The size of the output vectors is equal to the number
%% of rows in the input tables (observations). Each row or observation in the input table is
%% representative of a pixel in the original soil mask. This code also directly calculates the
%% area of each pixel.

% This code predicts the values of ECe; variables need to be changed to ESP to make
predictions for ESP
```

```
% Loading the best trained classification and per-class regression models
load('Location of the fitted ensemble object on the disk\ens_13')
Classification = compact(ens); % Function compact removes unnecessary data from the fitted
% model object (ens)
clear ens;
load('Location of the fitted ensemble object on the disk \ens_2')
Regression_1 = compact(ens); % Fitted regression model on the saline class
clear ens;
load('Location of the fitted ensemble object on the disk \ens_20')
Regression_2 = compact(ens); % Fitted regression model on the non-saline class
clear ens;

% Setting constant parameters required for calculation of the pixel area
a = 6378137;
b = 6356752.3142;
% e = sqrt(1 - (b/a)^2) = 0.08181919084296;
e = 0.08181919084296;
cell_size = 0.008333333;
edges = [0 4 8 16 100]; % Required for classifying the final predictions
% for ESP: edges = [0 6 15 30 100]

for ii = 1980:2018
    % 'ii' is the index of year. Input tables for each year are recorded in a separate
    % directory
    Text_files = dir(strcat('directory of the text files on the disk\',num2str(ii),'\*.txt'));
    parfor i = 1:numel(Text_files) % 'i' is the index of the input table
        % Preparing the tables
        T = readtable(strcat(Text_files(i).folder,'\',Text_files(i).name),'FileType',...
            'text','Delimiter',',','PreserveVariableNames',true);
        T = standardizeMissing(T,-9999); % Standardizing the missing values for MATLAB
        T = fillmissing(T,'nearest'); % Filling the missing values
        T.upper_dept = zeros(height(T),1); % Adding upper depth to the samples’ attributes
        T.lower_dept = 30.*ones(height(T),1); % Adding lower depth to the samples’ attributes
        % Categorizing the categorical variables in the training set
        T.WRB = categorical(T.WRB);
        T.LC = categorical(T.LC);
        T.Main_litho = categorical(T.Main_litho);

        % Predicting the labels for each class (classifying to saline and non-saline classes)
        % 'predict' function also returns a matrix of the classification scores
        % indicating the likelihood that the predicted label comes from a particular class
        [T.ECe,score] = predict(Classification,T);

        % Calculating scaled Shannon Entropy Index (Hs) using the per-class probability maps
        % For ESP the score matrix must first back-transform to probability
        % using: score = log(score./(1-score)); This is because
        % for ESP the classifier uses 'Bag' method and the returned
        % scores by this method are originally probabilities (values between 0 and 1);
        % however, during the training process, we set the score transformation to
        % be 'logit' and this transforms the scores to the values out of the
        % range of 0 and 1
        Hs = -(score(:,1).*log2(score(:,1)) + score(:,2).*log2(score(:,2)));
        Hs(sum(ismissing(T),2) > 5) = 255; % Setting the rows with more than five missing
        % values as no data
        % The indicator of the no data values is 255 here
        Hs = fillmissing(Hs,'constant',255);

        T.ECe(T.ECe == 1) = 10.^(predict(Regression_2,T(T.ECe == 1,:)))-1; % Making
        % predictions for the non-saline class
        T.ECe(T.ECe == 2) = 10.^(predict(Regression_1,T(T.ECe == 2,:))); % Making predictions
        % for the saline class
        T.ECe(sum(ismissing(T),2) > 5) = 255; % Setting the rows with more than five missing
        % values as no data
        % The indicator of no data values is 255 here
        T.ECe(T.ECe < 0) = 0; % Setting the negative predictions as zero
        T.ECe = fillmissing(T.ECe,'constant',255);

        %%%%%%%  Calculating m^2 area of a WGS 1984 square pixel  %%%%%%
        % Adapted from: https://gis.stackexchange.com/a/127327/2397
        % Parameters:
        % cell_size (float): Pixel size in the Geographic coordinates (WGS 1984)
        % Returns: Area of square pixel of side length cell_size in m^2
        f_up = deg2rad(T.Y + cell_size/2);
        f_down = deg2rad(T.Y - cell_size/2);
        zm_up = (1 - e*sin(f_up));
        zp_up = (1 + e*sin(f_up));
```

```
        area_up = pi * b^2 * (log(zp_up./zm_up)/(2*e) + sin(f_up)./(zp_up.*zm_up));
        zm_down = (1 - e*sin(f_down));
        zp_down = (1 + e*sin(f_down));
        area_down = pi * b^2 * (log(zp_down./zm_down)/(2*e) + ...
        sin(f_down)./(zp_down.*zm_down));
        cell_area = cell_size/360.*(area_up - area_down);
        grid = [T.X T.Y T.ECe Hs cell_area];

        % Exporting the predictions, scores, and calculated areas for each
        % observation (pixel) as a table
        T_result = array2table(grid,'VariableNames', {'X' 'Y' 'ECe' 'Hs' 'Area'});
        writetable(T_result,strcat('Output ...
        directory\',num2str(ii),'\1\','ECe_',Text_files(i).name));
        % Here we divide the output tables into four parts:
        % non-saline, slightly saline, moderately saline, and extremely saline
        % based on the predicted values of ECe and save each part separately. These are needed
        % later to do zonal statistics in ArcPy. From the variables of tables, only the area
        % of each pixel was required. Similar to this was conducted
        % for ESP and sodicity
        T_result.ECe = discretize(T_result.ECe,edges);
        % edges of the classes are defined before 'ii' loop
        T_result.Hs = [];
        for j = 2:4
            Table = T_result(T_result.ECe == j,:);
            Table.ECe = [];
            if height(Table) ~= 0 % Removing the tables without any record
                writetable(Table,strcat('Output directory\', ...
                num2str(ii),'\',num2str(j),'\','ECe_', ...
                num2str(j),'_',Text_files(i).name));
            end
        end
    end
end
```

The results of applying the trained models to input tables including each row's (point/pixel) area, x-, y-, H_s, and the corresponding predictions were then exported as new comma delimited tables in .txt format. So the output was 39 folders with 50,687 text files within each folder. In addition, we separated the predicted EC$_e$ and ESP values into four smaller bins. For salinity: 0 - 4 dS m^{-1}, 4 - 8 dS m^{-1}, 8 - 16 dS m^{-1}, and more than 16 dS m^{-1} and for sodicity: 0 - 6%, 6 - 15%, 15 - 30%, and more than 30%. Each bin included the left bin edge. According to these bins, we generated smaller sub-tables including only the values of x-, y-, and pixel area. We needed these later to calculate the per-class salinity and sodicity areas at the country, biome, land cover, and climate levels (see "Zonal statistics" section).

Trend analysis

As mentioned earlier, within each of 39 folders we had 50,687 output tables. We used the values of predictions for target variables from those output tables to create annual time series of EC$_e$ and ESP between 1980 and 2018. By fitting a linear model to these time series, we generated different layers including trends of soil salinity variation since 1980, likelihood of soils with EC$_e$ 4 ≥ dS m^{-1} or ESP ≥ 6%, and change in the likelihood of soils with EC$_e$ 4 ≥ dS m^{-1} or ESP ≥ 6%. The calculated coefficients (slopes) for locations with $p \geq 0.05$ were set as no data value. The trend values and x- y- coordinates were then converted to raster datasets and mosaicked to generate the variation of global longitude-latitude grid cells of EC$_e$ and ESP at 30" spatial resolution. Additionally, for the classification step of the two-part models, we produced 39-year mean of the pixel-level H_s (Figure A2-23). Also we generated other layers including the average of EC$_e$/ESP values between 1980 and 2018, and standard deviation of the predictions between 1980 and 2018. The following MATLAB code shows how we performed the trend analysis and computed the statistical layers:

```
clc;
clear;
%% Trend analysis,
%% This script returns the tables required for generation of the final raster layers.
```

```
%% It reads the predicted values from models and does trend analysis.
%% Also it computes mean, standard deviation, and scaled Shannon Entropy Index
%% of the predictions from 1980 to 2018. The code first reads the corresponding 39 tables from
%% each of 39 folders (representing the individual years between 1980 and 2018); each table
%% contains the X, Y, and predicted salinity values and is representative of a tile from the
%% original soil mask. Then puts 39 predictions in a matrix with 39 columns and rows equal to
%% the size of input tables (all 39 tables must have the same number of rows) and does trend
%% analysis for each row of the matrix. This processes will be repeated 50687 times to cover
%% all tables in all 39 directories.

% Note this code is generated for ECe; variables should be replaced by ESP for soil sodicity

parfor i = 1:50687
    % To have X and Y coordinates
    % 'i' is the index of tables in each folder
    table = readtable(strcat('Directory of the folder containing the tables of
        1980',num2str(i),'.txt'),'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);
    % Pre-allocating memory to variables with varying size in each iteration
    tile_ECe = zeros(height(table),39);
    tile_Hs = zeros(height(table),39);
    tile_Area = zeros(height(table),39);

    jj = 1; % 'jj' is the index of the column in matrix created form the 39
    % individual years' tables
    for j = 1980:2018
        % 'j' is the index of year
        T = readtable(strcat('The directory where the output tables of the model deployment
        are saved\ECe\',num2str(j),'\1\ECe_',num2str(j),'_',num2str(i),'.txt'),'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);
        % The predictions form 39 tables imported form 39 directories make a
        % matrix here with 39 columns
        T = standardizeMissing(T,255);
        tile_ECe(:,jj) = T.ECe;
        tile_Hs (:,jj) = T.Hs;
        tile_Area (:,jj) = T.Area;
        jj = jj+1;
    end

    % Pre-allocating memory to variables with varying size in each iteration
    Coeff_value = zeros(height(table),1);
    P_value = zeros(height(table),1);
    Hs_mean = zeros(height(table),1);
    Mean = zeros(height(table),1);
    Std = zeros(height(table),1);
    Frequency_2 = zeros(height(table),1);
    Frequency_4 = zeros(height(table),1);
    Frequency_change_2 = zeros(height(table),1);
    Frequency_change_4 = zeros(height(table),1);

    % To each row of the created matrix, a linear model is fitted
    for ii = 1:height(table)
        mdl = fitlm(1980:2018,tile_ECe(ii,1:39)); % mdl is the object created from the linear
        % model fitting
        % Acquiring the slope coefficient and p-value for the t-statistic of the hypothesis
        % test that the corresponding coefficient is equal to zero or not
        % from the linear regression object
        mdl_Coefficients = table2array(mdl.Coefficients);
        Coeff_value(ii,1) = mdl_Coefficients(2,1);
        P_value(ii,1) = mdl_Coefficients(2,4);
        % Calculation of other required statistics
        Hs_mean(ii,1) = nanmean(tile_Hs(ii,:)); % Average of the scaled Shannon Entropy Index
        % between 39 years
        Mean(ii,1) = nanmean(tile_ECe(ii,:)); % Average of the predicted ECe values between 39
        % years
        Std(ii,1) = nanstd(tile_ECe(ii,:)); % Standard deviation of the predicted ECe values
        % between 39 years
        % Computing the frequency of happening saline soil assuming saline
        % soil has an ECe value more than 2 dS/m (or happening sodic soil with ESP value more
        % than 6% for ESP)
        Frequency_2(ii,1) = nansum(tile_ECe(ii,:) >= 2)/numel(1980:2018);
        % Computing the frequency of happening saline soil assuming saline
        % soil has an ECe value more than 4 dS/m
        Frequency_4(ii,1) = nansum(tile_ECe(ii,:) >= 4)/numel(1980:2018);
        % Computing the change in frequency of happening saline soil assuming saline
        % soil has an ECe value more than 4 dS/m (or happening sodic soil with ESP value more
        % than 6% for ESP)
```

```
        Frequency_change_4(ii,1) = log(((nansum(tile_ECe(ii,21:39) >= 4) + 0.5)...
        /numel(2000:2018))/((nansum(tile_ECe(ii,2:20) >= 4) + 0.5)/numel(1981:1999)));
    End

    % Replacing the missing values with no data value indicators
    Coeff_value = fillmissing(Coeff_value,'constant',-9999);
    P_value = fillmissing(P_value,'constant',255);
    Hs_mean = fillmissing(Hs_mean,'constant',255);
    Mean = fillmissing(Mean,'constant',255);
    Std = fillmissing(Std,'constant',255);
    Frequency_2 = fillmissing(Frequency_2,'constant',255);
    Frequency_4 = fillmissing(Frequency_4,'constant',255);
    Frequency_change_2 = fillmissing(Frequency_change_2,'constant',255);
    Frequency_change_2(Frequency_change_2 == Inf) = 255;
    Frequency_change_4 = fillmissing(Frequency_change_4,'constant',255);
    Frequency_change_4(Frequency_change_4 == Inf) = 255;

    % Creating matrices form the results
    Coeff_matrix = [table.X table.Y Coeff_value P_value];
    table_fitlm = array2table(Coeff_matrix,'VariableNames',...
        {'X' 'Y' 'Coeff_value' 'P_value'});
    Hs_matrix = [table.X table.Y Hs_mean];
    table_Hs_mean = array2table(Hs_matrix,'VariableNames',...
        {'X' 'Y' 'Hs_mean'});
    Mean_matrix = [table.X table.Y Mean];
    table_Mean = array2table(Mean_matrix,'VariableNames',...
        {'X' 'Y' 'Mean'});
    Std_matrix = [table.X table.Y Std];
    table_Std = array2table(Std_matrix,'VariableNames',...
        {'X' 'Y' 'Std'});
    Frequency_2_matrix = [table.X table.Y table.Area Frequency_2];
    table_Frequency_2 = array2table(Frequency_2_matrix,'VariableNames',...
        {'X' 'Y' 'Area' 'Frequency_2'});
    Frequency_4_matrix = [table.X table.Y table.Area Frequency_4];
    table_Frequency_4 = array2table(Frequency_4_matrix,'VariableNames',...
        {'X' 'Y' 'Area' 'Frequency_4'});
    Frequency_change_2_matrix = [table.X table.Y Frequency_change_2];
    table_Frequency_change_2 = array2table(Frequency_change_2_matrix,'VariableNames',...
        {'X' 'Y' 'Frequency_change_2'});
    Frequency_change_4_matrix = [table.X table.Y Frequency_change_4];
    table_Frequency_change_4 = array2table(Frequency_change_4_matrix,'VariableNames',...
        {'X' 'Y' 'Frequency_change_4'});

    % Exporting the required results as tables into different directories
    writetable(table_fitlm(table_fitlm.P_value <= 0.05,1:3),strcat('Output...
    directory\Coeff_05',num2str(i),'.txt'));
    writetable(table_fitlm(:,1:3),strcat('Output directory\Coeff',num2str(i),'.txt'));
    writetable(table_fitlm(:,[1 2 4]),strcat('Output directory\P_value',num2str(i),'.txt'));
    writetable(table_Hs_mean,strcat('Output directory\Hs_mean',num2str(i),'.txt'));
    writetable(table_Mean,strcat('Output directory\Mean',num2str(i),'.txt'));
    writetable(table_Std,strcat('Output directory\Std',num2str(i),'.txt'));
    writetable(table_Frequency_2(:,[1 2 4]),strcat('Output...
    directory\Frequency_2',num2str(i),'.txt'));
    writetable(table_Frequency_2(table_Frequency_2.Frequency_2 >= 0.75,1:3),...
    strcat('Output directory\Frequency_2_area',num2str(i),'.txt'));
    writetable(table_Frequency_4(:,[1 2 4]),strcat('Output...
    directory\Frequency_4',num2str(i),'.txt'));
    writetable(table_Frequency_4(table_Frequency_4.Frequency_4 >= 0.75,1:3),...
    strcat('Output directory\Frequancy_4_area',num2str(i),'.txt'));
    writetable(table_Frequency_change_4,strcat('Output...
    directory\Frequency_change_4',num2str(i),'.txt'));
end
```

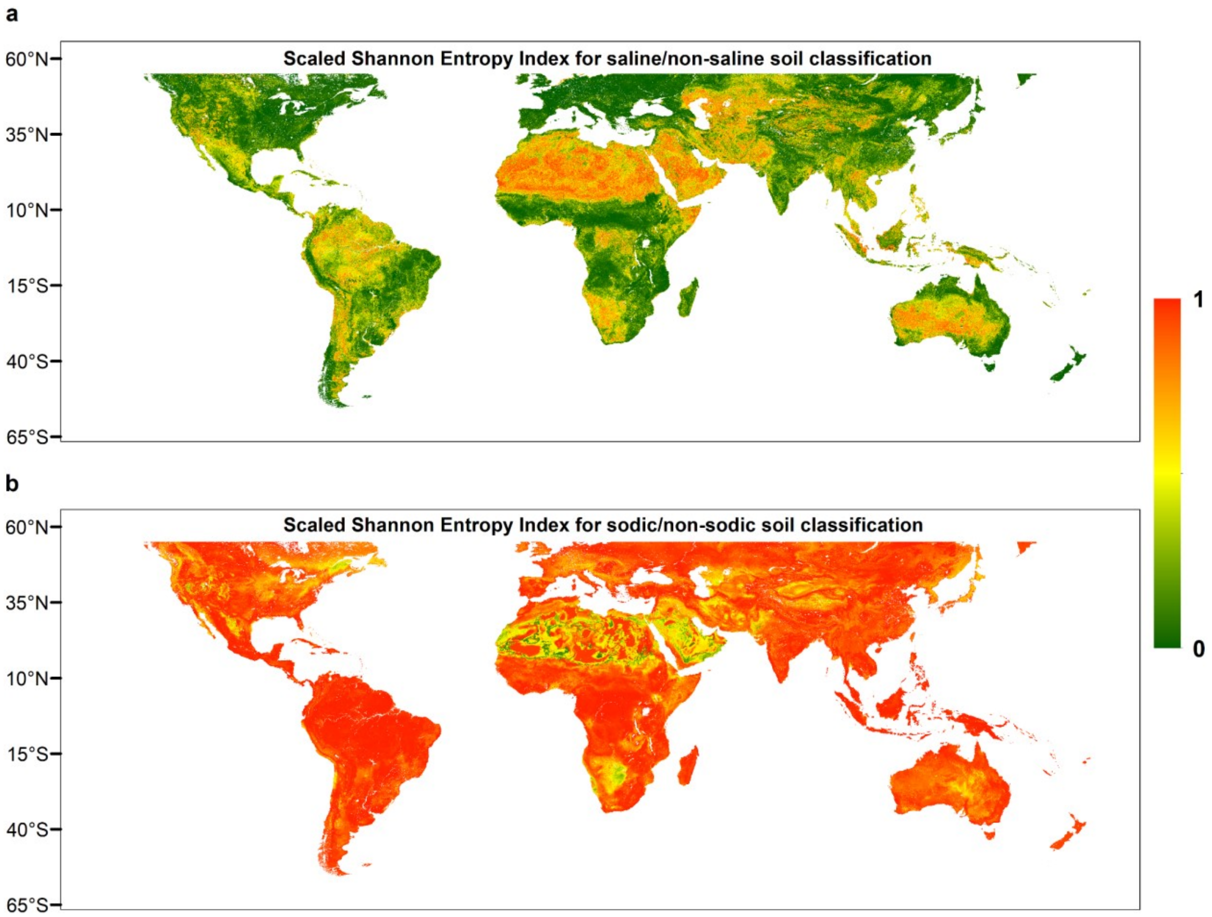


Figure A2-23: Spatial distribution of the scaled Shannon Entropy index (H_s) for the classification part of the two-part models. Values close to zero indicate that the classifier is more certain about the results of binary classification task. Values of H_s should not be confused with the accuracy of classification. Even with a zero H_s, still the classifier can make wrong predictions.

Rasterizing the generated tables

To reduce the required time for rasterising the output tables, first we merged the 50,687 tables and reduced the number of output tables to a number below 500. During this merging process, we also multiplied the predictions for target variables by 100, 10,000, or 100,000 (depending on the needed accuracy) and rounded the results to remove the decimal point from the predictions. Using the signed integer values substantially reduced the required disk space for saving the final raster layers generated from the output tables. The following MATLAB code shows how we merged the tables, converted the float predictions to integer, and defined no data value indicators of the final rasters (generated from the output tables):

```
clc;
clear;
%% Table Merger,
%% This MATLAB code merges a large number of tables in a directory, passes the contents of the
%% merged table to a desirable number of tables, and export those tables into another
%% directory. Also the table variables (exempt X and Y coordinates)
%% will be converted from float to integer during the merging process.

% Reading all tables in a directory (in .txt format) and copying them into the memory
P = strcat('Directory of the original comma delimited input tables');
S = dir(fullfile(P,'*.txt'));
C = cell(1,numel(S));
parfor k = 1:numel(S)
    F = fullfile(P,S(k).name);
    C{k} = readtable(F,'FileType',...
```

```
        'text','Delimiter',',','PreserveVariableNames',true);
end

% Merging tables
Table = vertcat(C{:});

% Multiplying table variables by 100, 10000,or 100000 depending on the needed accuracy and
% rounding the results to remove the decimal point
Table.var = round(Table.var * 10000);
% Setting no data value indicators
Table.var(Table.var == round(no data indicator of input tables * 10000)) = No data indicator;

% Passing the contents of the merged table to a desirable number of tables
[~,~,bin] = histcounts(1:height(Table),500);
T = cell(1,500);
for z  = 1:500
    T{z} = Table(bin == z,:);
end

% Exporting the resulting tables to a desirable directory
parfor z = 1:500
    TT = vertcat(T{z});
    writetable(TT,strcat('Desirable output directory\desirable table name',...
    num2str(z),'.txt'));
end
```

We converted the x- y- values in the output tables to in-memory point feature layers and afterwards converted the generated in-memory point feature layers to rasters using the ArcPy “point to raster” tool. The created rasters were then mosaicked in ArcMap GUI and exported as one single global raster with 0.008333333° spatial resolution. The following Python code shows how we converted the tables to rasters:

```
# Table to raster converter,
# this code gets X, Y, and predicted values for target variables in a .txt table and generates
# the raster datasets.

import multiprocessing # Importing the ArcPy module
from multiprocessing import Process # Importing multi-processing module
import arcpy
import os
import os, fnmatch

# Setting the geo-processing environments
arcpy.env.overwriteOutput = True
arcpy.env.workspace = r"Directory of the tables"

# Reading all .txt tables in a directory
path = r"Directory of the tables"
pattern = "*.txt"
Text_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]

# Defining the function that will be passed to child processes
# Function arguments: ini, end
def cell(ini,end):
 jj = range(ini,end)
 for j in jj:
  string = Text_files[j]
  string = string[0:(len(string)-4)]
  in_table = Text_files[j]
  x_coords = "X" # Longitude in the table
  y_coords = "Y" # Latitude in the table
  # Making a point feature class from the variables in the tables on the memory
  arcpy.MakeXYEventLayer_management(in_table, x_coords , y_coords, "out_Layer")
  # Execution of the ArcPy point to raster tool,
  # 0.008333333 is the resolution of the final raster
  arcpy.PointToRaster_conversion("out_Layer", "Coeff_value", r"Output directory/"+string +
  ".tif","","",0.008333333)

if __name__ == '__main__':
 ii = count = 0
 for i in range(0, number of system cores):
  ini = count
```

```
  end = count + + The number of tables that should be converted to raster by each core
  process = Process(target = cell, args = (ini,end,))
  processes.append(process)
  process.start()
  count = end
```

Zonal statistics

As mentioned earlier in the “Model deployment” section, we classified the predicted values of EC_e and ESP to four classes or bins for each variable (8 classes in total). This classification enabled us to do different analysis on the total area of salt-affected soils between various thresholds. From those four bins for each variable, only the bins with an $EC_e \geq 4$ dS m^{-1} and an ESP $\geq$ 6% were required for area analysis. Instead of saving the values of EC_e or ESP, the calculated area of pixels (based on x- y- coordinates) were saved as the final tables (x-, y-, and Area) into the corresponding folders. The result for each target variable was 39 folders (representing each year between 1980 and 2018) with 3 sub-folders including the generated tables for each bin. A similar script to the one presented in previous section (Rasterizing the generated tables) was used to rasterise all tables within the sub-folders. Each raster was representing the area of pixels labelled with different classes of the soil salinity or sodicity. To make it clearer, we provide an example from the final arrangement of the generated raster files. For EC_e as a target variable in the 1999 main folder, for example, we had three sub-folders with the names of 1, 2, and 3, representing the three bins of salinity: 4 - 8, 8 - 16, and equal or greater than 16, respectively. Within each sub-folder, we had a set of raster files. These raster files included information on the area of pixels, which were estimated to have an EC_e value falling into the corresponding bin of salinity. The following script was used to mosaic the final generated raster files within each sub-folder:

```
# Raster mosaicing,
# This script gets a set of raster datasets in a directory and mosaic all of them into a new
# raster.

# Importing the required modules
import arcpy
from arcpy import env
from arcpy.sa import *
import os
import os, fnmatch

pattern = "*.tif"
j = 1
j = str(j)
for i in range(1980,2019):
 arcpy.env.workspace = r"Directory of the rasters on the disk/ECe/"+str(i)+"/"+j
 path = r" Directory of the rasters on the disk /ECe/"+str(i)+"/"+j
 tif_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]
 arcpy.CreateRasterDataset_management("…/ECe/"+str(i)+"/"+j," ECe_"+str(i)+"_"+j+
  ".tif","0.008333333","32_BIT_FLOAT","","1","","","LZ77","","")
 arcpy.Mosaic_management(tif_files,"…/Mosaiced_rasters/ECe/"+str(i)+"/"+j+"/ECe_"+str(i)+"_"+j+
 ".tif" ,"BLEND","","","No Data Value","","","")
```

Using ArcMap “zonal statistics as table” tool, the area of soils with $EC_e \geq 4$ dS m^{-1} or ESP $\geq$ 6% at the country, continent, biome, land cover, and climate level was calculated from the mosaicked rasters for different years. The following code in Python was used to calculate the sum of salt-affected areas between various thresholds in each zone delineated by input rasters of country, continent, biome, land cover, and climate:

```
# Zonal statistics as table,
# Summarizes the values of a raster within the zones of another dataset and reports the
# results in a table.

import arcpy
```

```
from arcpy import env
from arcpy.sa import *
import os
import os, fnmatch

# Setting the geo-processing environments
arcpy.CheckOutExtension("Spatial")

# Input rasters that delineate the zones
pattern = "*.tif"
inZoneData_1 = "…/Continent_level.tif"
zoneField_1 = "Continent_name"
inZoneData_2 = "…/gadm_country_level.tif"
zoneField_2 = "Country_name"
inZoneData_3 = "…/Biome_level.tif"
zoneField_3 = "Biome_name"
inZoneData_4 = "…/Land_cover_level.tif"
zoneField_4 = "Land_cover_name"
inZoneData_5 = "…/Climate_zone_level.tif"
zoneField_5 = "climate_name"

for i in range(1980,2019):
# i represents the year
 for j in range(1,4): # j represents the bin of salinity or sodicity
  env.workspace = "…/Mosaiced_rasters/ECe/"+str(i)+"/"+str(j)
  path = "…/Mosaiced_rasters/ECe/"+str(i)+"/"+str(j)
  tif_files = [ff for ff in os.listdir(path) if fnmatch.fnmatch(ff, pattern)]
  outZSaT = ZonalStatisticsAsTable(inZoneData_1, zoneField_1, tif_files[0],
  "in_memory/dbf", "DATA", "SUM")
  arcpy.TableToTable_conversion("in_memory/dbf","…/zonal_stat_results/ECe/"+str(i)+"/"+str(j),
  "Continent_level_"+str(i)+".txt")
  arcpy.Delete_management("in_memory/dbf")
  outZSaT = ZonalStatisticsAsTable(inZoneData_2, zoneField_2, tif_files[0],
  "in_memory/dbf", "DATA", "SUM")
  arcpy.TableToTable_conversion("in_memory/dbf","…/zonal_stat_results/ECe/"+str(i)+"/"+str(j),
  "Country_level_"+str(i)+".txt")
  arcpy.Delete_management("in_memory/dbf")
  outZSaT = ZonalStatisticsAsTable(inZoneData_3, zoneField_3, tif_files[0],
  "in_memory/dbf", "DATA", "SUM")
  arcpy.TableToTable_conversion("in_memory/dbf","…/zonal_stat_results/ECe/"+str(i)+"/"+str(j),
  "Biome_level_"+str(i)+".txt")
  arcpy.Delete_management("in_memory/dbf")
  outZSaT = ZonalStatisticsAsTable(inZoneData_4, zoneField_4, tif_files[0],
  "in_memory/dbf", "DATA", "SUM")
  arcpy.TableToTable_conversion("in_memory/dbf","…/zonal_stat_results/ECe/"+str(i)+"/"+str(j),
  "Land_cover_level_"+str(i)+".txt")
  arcpy.Delete_management("in_memory/dbf")
  outZSaT = ZonalStatisticsAsTable(inZoneData_5, zoneField_5, tif_files[0],
  "in_memory/dbf", "DATA", "SUM")
  arcpy.TableToTable_conversion("in_memory/dbf","…/zonal_stat_results/ECe/"+str(i)+"/"+str(j),
  "Climate_level_"+str(i)+".txt")
  arcpy.Delete_management("in_memory/dbf")
```

The generated tables (saved in .txt format) were then imported to MATLAB for area analysis and generation of the figures.

Figures

Plots in Figure 3-2 were initially generated in MATLAB and then copy pasted into ArcMap. The final generated figure was exported by ArcMap. The following code shows the MATLAB part and it calculates the soil areas with an $EC_e \geq 4$ dS m^{-1}/$EC_e \geq 2$ dS m^{-1} and/or ESP $\geq 6\%$/ESP $\geq 15\%$:

```
clc;
clear;

% Reading all tables in a directory containing the output tables of the trend analysis.
% Each table containes X, Y, and pixel area.
% Calculated areas are related to the pixels for which the likelihood of happening ECe or
% ESP equal or greater than specific thresholds is more than 0.75; in other
```

```
% words, in three-fourths of the years between 1980 and 2018, the likelihood of happening ECe
% or ESP for those pixels has been equal or greater than specific thresholds.

P = strcat('…\Tables');% Reading the tables for likelihood of happening ECe >= 2
S = dir(fullfile(P,'*.txt'));
C = cell(1,numel(S));
parfor k = 1:numel(S)
    F = fullfile(P,S(k).name);
    C{k} = readtable(F,'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);
end
Table = vertcat(C{:}); % Merging all tables
Table(Table.Y > 55,:) = [];
longitude = unique(Table.X);
longitude_area = zeros(length(longitude),1);
for i = 1:length(longitude)
    longitude_area(i) = sum(Table.Area(Table.X == longitude(i)));

end
Longitude_figure = [longitude longitude_area];
Longitude_figure = array2table(Longitude_figure,'VariableNames',...
{'longitude' 'longitude_area'});
writetable(Longitude_figure,'…\ECe_2_Longitude_figure.txt');
latitude = unique(Table.Y);
latitude_area = zeros(length(latitude),1);
for i = 1:length(latitude)
    latitude_area(i) = sum(Table.Area(Table.Y == latitude(i)));

end
Latitude_figure = [latitude latitude_area];
Latitude_figure = array2table(Latitude_figure,'VariableNames',{'latitude' 'latitude_area'});
writetable(Latitude_figure,'…\ECe_2_Latitude_figure.txt');


%%%%%%%%%%%
clear;
P = strcat('…\Tables');% Reading the tables for likelihood of happening ECe >= 4
S = dir(fullfile(P,'*.txt'));
C = cell(1,numel(S));
parfor k = 1:numel(S)
    F = fullfile(P,S(k).name);
    C{k} = readtable(F,'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);
end
Table = vertcat(C{:});
Table(Table.Y > 55,:) = [];
longitude = unique(Table.X);
longitude_area = zeros(length(longitude),1);
for i = 1:length(longitude)
    longitude_area(i) = sum(Table.Area(Table.X == longitude(i)));

end
Longitude_figure = [longitude longitude_area];
Longitude_figure = array2table(Longitude_figure,'VariableNames', ...
{'longitude' 'longitude_area'});
writetable(Longitude_figure,'…\ECe_4_Longitude_figure.txt');
latitude = unique(Table.Y);
latitude_area = zeros(length(latitude),1);
for i = 1:length(latitude)
    latitude_area(i) = sum(Table.Area(Table.Y == latitude(i)));

end
Latitude_figure = [latitude latitude_area];
Latitude_figure = array2table(Latitude_figure,'VariableNames',{'latitude' 'latitude_area'});
writetable(Latitude_figure,'…\ECe_4_Latitude_figure.txt');
%%%%%%%%%%%% Generation of subplots
clear;
P = strcat('…\Tables');% Reading the tables for likelihood of happening ESP >= 6
S = dir(fullfile(P,'*.txt'));
C = cell(1,numel(S));
parfor k = 1:numel(S)
    F = fullfile(P,S(k).name);
    C{k} = readtable(F,'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);
end
Table = vertcat(C{:});
Table(Table.Y > 55,:) = [];
```

```
longitude = unique(Table.X);
longitude_area = zeros(length(longitude),1);
for i = 1:length(longitude)
    longitude_area(i) = sum(Table.Area(Table.X == longitude(i)));

end
Longitude_figure = [longitude longitude_area];
Longitude_figure = array2table(Longitude_figure,'VariableNames',{'longitude'
'longitude_area'});
writetable(Longitude_figure,'…\ESP_6_Longitude_figure.txt');
latitude = unique(Table.Y);
latitude_area = zeros(length(latitude),1);
for i = 1:length(latitude)
    latitude_area(i) = sum(Table.Area(Table.Y == latitude(i)));

end
Latitude_figure = [latitude latitude_area];
Latitude_figure = array2table(Latitude_figure,'VariableNames',{'latitude' 'latitude_area'});
writetable(Latitude_figure,'…\ESP_6_Latitude_figure.txt');

P = strcat('\Tables');% Reading the tables for likelihood of happening ESP >= 15
S = dir(fullfile(P,'*.txt'));
C = cell(1,numel(S));
parfor k = 1:numel(S)
    F = fullfile(P,S(k).name);
    C{k} = readtable(F,'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);
end
Table = vertcat(C{:});
Table(Table.Y > 55,:) = [];
longitude = unique(Table.X);
longitude_area = zeros(length(longitude),1);
for i = 1:length(longitude)
    longitude_area(i) = sum(Table.Area(Table.X == longitude(i)));

end
Longitude_figure = [longitude longitude_area];
Longitude_figure = array2table(Longitude_figure,'VariableNames',{'longitude'
'longitude_area'});
writetable(Longitude_figure,'…\ESP_15_Longitude_figure.txt');
latitude = unique(Table.Y);
latitude_area = zeros(length(latitude),1);
for i = 1:length(latitude)
    latitude_area(i) = sum(Table.Area(Table.Y == latitude(i)));

end
Latitude_figure = [latitude latitude_area];
Latitude_figure = array2table(Latitude_figure,'VariableNames',{'latitude' 'latitude_area'});
writetable(Latitude_figure,'…\ESP_15_Latitude_figure.txt');

%% Subplots related to ECe
figure;

subplot(5,6,25:29);
Longitude_figure_2 = readtable('…\ECe_2_Longitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P1 = plot(Longitude_figure_2.longitude,Longitude_figure_2.longitude_area./1000000,...
'Color','r','LineWidth',1.5);
set(gca,'fontname','Arial','FontSize',20)
box on
ax = gca;
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
ax.TickLength = [0.004 0.035];
ylabel('Area (km^{2})','Color','k');
xlabel('Longitude (degree)','Color','k');
hold on
Longitude_figure_4 = readtable('…\ECe_4_Longitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P2 = plot(Longitude_figure_4.longitude,Longitude_figure_4.longitude_area./1000000,...
'Color','b','LineWidth',1.5);
xlim([-180 180]);
ytickformat('%,4.4g')
L = legend([P1 P2],...
{'EC_{e} \geq 2 dS m^{-1}','EC_{e} \geq 4 dS m^{-1}'},'Location','northwest','FontSize',14);
title(L,'Salinity threshold:')
```

```
ax.YAxisLocation = 'left';
ax.XMinorTick = 'on';
ax.YMinorTick = 'on';

subplot(5,6,[12 18 24]);
Latitude_figure_2 = readtable('…\ECe_2_Latitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P1 = plot(Latitude_figure_2.latitude_area./1000000,Latitude_figure_2.latitude,...
'Color','r','LineWidth',1.5);
set(gca,'fontname','Arial','FontSize',20)
box on
ax = gca;
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
ylabel('Latitude (degree)','Color','k');
xlabel('Area (km^{2})','Color','k');
hold on
Latitude_figure_4 = readtable('…\ECe_4_Latitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P2 = plot(Latitude_figure_4.latitude_area./1000000,Latitude_figure_4.latitude,...
'Color','b','LineWidth',1.5);
ylim([-55 55]);
ax.YTick = [-50 -25 0 25 50];
xtickformat('%,4.4g')
L = legend([P1 P2],...
{'EC_{e} \geq 2 dS m^{-1}','EC_{e} \geq 4 dS m^{-1}'},'Location','southeast','FontSize',12);
title(L,'Salinity threshold:')
ax.XMinorTick = 'on';
ax.YMinorTick = 'on';
ax.YAxisLocation = 'right';

subplot(5,6,[13 19]);
Hor_bar_fig_2 = readtable('…\Continent_level.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_2_mid_chi = readtable('…\Middle_east_china.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_2_mid_chi.CONTINENT = Hor_bar_fig_2_mid_chi.NAME_0;
Hor_bar_fig_2_mid_chi.NAME_0 = [];
table_2 = [Hor_bar_fig_2;Hor_bar_fig_2_mid_chi];
Area_2 = sortrows(table_2,'SUM','descend');
Hor_bar_fig_4 = readtable('…\Continent_level.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_4_mid_chi = readtable('…\Middle_east_china.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_4_mid_chi.CONTINENT = Hor_bar_fig_4_mid_chi.NAME_0;
Hor_bar_fig_4_mid_chi.NAME_0 = [];
table_4 = [Hor_bar_fig_4;Hor_bar_fig_4_mid_chi];
Area_4 = sortrows(table_4,'SUM','descend');
Area = [Area_2.SUM Area_4.SUM];
b = barh(Area./1000000000,0.75);
b(1).FaceColor = 'r';
b(2).FaceColor = 'b';
set(gca,'fontname','Arial','FontSize',20)
box on
ax = gca;
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
xlabel('Area (1,000 km^{2})','Color','k');
Area_2.CONTINENT{3} = 'Middle E.';
Area_2.CONTINENT{6} = 'S. America';
Area_2.CONTINENT{7} = 'N. America';
ax.YTickLabel = Area_2.CONTINENT;
ax.YDir = 'reverse';
L = legend({'EC_{e} \geq 2 dS m^{-1}','EC_{e} \geq 4 dS m^{-1}'},...
'Location','southeast','FontSize',12);
title(L,'Salinity threshold:')
ax.Position = [0.1300 0.34 0.1 0.3];
ax.XTick = [0 2500 5000];
xtickformat('%,4.4g')
ax.XMinorTick = 'on';


% %% Subplots related ESP
figure;
```

```
subplot(5,6,25:29);
Longitude_figure_6 = readtable('…\ESP_6_Longitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P1 = plot(Longitude_figure_6.longitude,Longitude_figure_6.longitude_area./1000000,...
'Color','r','LineWidth',1.5);
set(gca,'fontname','Arial','FontSize',20)
box on
ax = gca;
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
ax.TickLength = [0.004 0.035];
ylabel('Area (km^{2})','Color','k');
xlabel('Longitude (degree)','Color','k');
hold on
Longitude_figure_15 = readtable('…\ESP_15_Longitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P2 = plot(Longitude_figure_15.longitude,Longitude_figure_15.longitude_area./1000000,...
'Color','b','LineWidth',1.5);
xlim([-180 180]);
ytickformat('%,4.4g')
L = legend([P1 P2],{'ESP \geq 6%','ESP \geq 15%'},'Location','northwest','FontSize',16);
title(L,'Sodicity threshold:')
ax.YAxisLocation = 'left';
ax.XTick = [-150 -100 -50 0 50 100 150];
ax.XMinorTick = 'on';
ax.YMinorTick = 'on';

subplot(5,6,[12 18 24]);
Latitude_figure_4 = readtable('…\ESP_6_Latitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P1 = plot(Latitude_figure_4.latitude_area./1000000,Latitude_figure_4.latitude,...
'Color','r','LineWidth',1.5);
set(gca,'fontname','Arial','FontSize',20)
box on
ax = gca;
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
ylabel('Latitude (degree)','Color','k');
xlabel('Area (km^{2})','Color','k');
hold on
Latitude_figure_15 = readtable('…\ESP_15_Latitude_figure.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
P2 = plot(Latitude_figure_15.latitude_area./1000000,Latitude_figure_15.latitude,...
'Color','b','LineWidth',1.5);
ylim([-55 55]);
xtickformat('%,4.4g')
L = legend([P1 P2],{'ESP \geq 6%','ESP \geq 15%'},'Location','southeast','FontSize',12);
title(L,'Sodicity threshold:')
ax.YTick = [-50 -25 0 25 50];
ax.XMinorTick = 'on';
ax.YMinorTick = 'on';
ax.YAxisLocation = 'right';

subplot(5,6,[13 19]);
Hor_bar_fig_6 = readtable('…\Continent_level.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_6_mid_chi = readtable('…\Middle_east_china.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_6_mid_chi.CONTINENT = Hor_bar_fig_6_mid_chi.NAME_0;
Hor_bar_fig_6_mid_chi.NAME_0 = [];
Hor_bar_fig_6.OID_ = [];Hor_bar_fig_6_mid_chi.Rowid_ = [];
table_6 = [Hor_bar_fig_6;Hor_bar_fig_6_mid_chi];
table_6(9,:) = table_6(6,:);
table_6(6,:) = [];
Area_6 = sortrows(table_6,'CONTINENT','descend');
Hor_bar_fig_15 = readtable('…\Fre_15_area_continet_level.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_15_mid_chi = readtable('…\Fre_15_area_Middle_east_china.txt','FileType',...
    'text','Delimiter',',','PreserveVariableNames',true);
Hor_bar_fig_15_mid_chi.CONTINENT = Hor_bar_fig_15_mid_chi.NAME_0;
Hor_bar_fig_15_mid_chi.NAME_0 = [];
table_15 = [Hor_bar_fig_15;Hor_bar_fig_15_mid_chi];
table_15(8,2) = {'Europe'};
table_15.OID_ = [];
Area_15 = sortrows(table_15,'CONTINENT','descend');
```

```
Area = [Area_6.SUM Area_15.SUM];
b = barh(Area./1000000000,0.75);
b(1).FaceColor = 'r';
b(2).FaceColor = 'b';
set(gca,'fontname','Arial','FontSize',20)
box on
ax = gca;
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
xlabel('Area (1,000 km^{2})','Color','k');
Area_6.CONTINENT{1} = 'S. America';
Area_6.CONTINENT{2} = 'N. America';
Area_6.CONTINENT{3} = 'Middle E.';
ax.YTickLabel = Area_6.CONTINENT;
ax.XTick = [0 3000 6000];
L = legend({'ESP \geq 6%','ESP \geq 15%'},'Location','southeast','FontSize',12);
title(L,'Sodicity threshold:')
ax.XMinorTick = 'on';
ax.Position = [0.1300 0.34 0.1 0.3];
xlim([0 7000]);
xtickformat('%,4.4g')
```

Figure A2-1 was generated by a combination of MATALB, Microsoft PowerPoint, and ArcMap. The following script was used for creation of the Figure A2-1 in MATLAB and the generated figure was copy-pasted into the ArcMap GUI to become combined with the outputs of the Microsoft PowerPoint. The final version of Figure A2-1 was exported by ArcMap.

```
clc;
clear;
% This code shows how we generated Figure A2-1 in the manuscript.

% subplot (a)
ens = ... load('…\Best binary classification model object for ECe.mat');
Truelabels = ens.ens.Y;
% Validation and Acquiring accuracy metrics
cvens = crossval(ens.ens,'Kfold',10);
Predictedlabels = kfoldPredict(cvens);
Variables = [Truelabels Predictedlabels];
Variables_table = array2table(Variables,'VariableNames',{'Truelabels' 'Predictedlabels'});
writetable(Variables_table, '…\ECe_Classification_cross_validation.txt');
T = readtable('…\ECe_Classification_cross_validation','FileType','text','Delimiter',',',...
       'PreserveVariableNames',true);
Truelabels = T.Truelabels;
Predictedlabels = T.Predictedlabels;
C = confusionmat(Truelabels,Predictedlabels);
cm = confusionchart(C,'RowSummary','row-normalized','ColumnSummary','column-normalized');
% The generated figure was exported to Microsoft PowerPoint and after modifications was sent
to ArcMap
subplot(2,3,1);
set(gca,'fontname','Arial','FontSize',20)
ax = gca;
box on
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
ax.XTick = [];
ax.YTick = [];
ty = ylabel('True class','Color','k');
ty.Position = [-0.08 0.65 0];
tx = xlabel('Predicted class','Color','k');
tx.Position = [0.46 -.087 0];
tit = title('Soil salinity classification','FontSize',18);
text(-0.12,1.1,0,'a','Units','Normalized','fontname','Arial','Color','k','FontSize',24,...
'FontWeight','Bold');

% subplot (b)
ens = load('…\Best binary classification object for ESP.mat');
Truelabels = ens.ens.Y;
% Validation and Acquiring accuracy metrics
cvens = crossval(ens.ens,'Kfold',10);
Predictedlabels = kfoldPredict(cvens);
Variables = [Truelabels Predictedlabels];
```

```
Variables_table = array2table(Variables,'VariableNames',{'Truelabels' 'Predictedlabels'});
writetable(Variables_table,'…\ESP_Classification_cross_validation.txt');
T = readtable('…\ESP_Classification_cross_validation','FileType',...
                  'text','Delimiter',',','PreserveVariableNames',true);
Truelabels = T.Truelabels;
Predictedlabels = T.Predictedlabels;
C = confusionmat(Truelabels,Predictedlabels);
cm = confusionchart(C,'RowSummary','row-normalized','ColumnSummary','column-normalized');
% The generated figure was exported to Microsoft PowerPoint and after modifications was sent
to ArcMap
subplot(2,3,4);
set(gca,'fontname','Arial','FontSize',20)
ax = gca;
box on
ax.XTick = [];
ax.YTick = [];
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
ty = ylabel('True class','Color','k');
ty.Position = [-0.08 0.65 0];
tx = xlabel('Predicted class','Color','k');
tx.Position = [0.46 -.087 0];
title('Soil sodicity classification','FontSize',18);
text(-0.12,1.1,0,'b','Units','Normalized','fontname','Arial','Color','k', ...
'FontSize',24,'FontWeight','Bold');

% Figure 1 subplots (c to f)
% subplot(c)
ens = load('…\Best regression model object fitted to saline class.mat');
ytrue_log = ens.ens.Y;
ytrue = 10.^(ytrue_log);
% Validation and Acquiring accuracy metrics
cvens = crossval(ens.ens,'Kfold',10);
yfit_log = kfoldPredict(cvens);
yfit = 10.^(yfit_log);
Variables = [ytrue yfit];
Variables_table = array2table(Variables,'VariableNames',{'ytrue' 'yfit'});
writetable(Variables_table,'…\ECe_regression_cross_validation.txt');
T = readtable('…\ECe_regression_cross_validation','FileType',...
                  'text','Delimiter',',','PreserveVariableNames',true);
ytrue = T.ytrue;
yfit = T.yfit;
subplot(2,3,2);
binscatter(ytrue,yfit,90)
set(gca,'XScale','log','YScale','log','fontname','Arial','FontSize',20)
title('Actual vs fitted: Saline class','Color','k','FontSize',18);
ylabel('Predicted EC_{e} (dS m^{-1})','Color','k');
xlabel('Observed EC_{e} (dS m^{-1})','Color','k');
xlim([1.9 67]);
ylim([1.9 100]);
xticks([2 5 10 30 65]);
yticks([2 5 10 30 65]);
colormap hot
c = colorbar;
c.Label.String = 'Scatter density in bins';
box on
ax = gca;
ax.XColor = 'k';
ax.YColor = 'k';
ax.LineWidth = 1;
hold on
x = 1.8:0.001:67;y = 1.8:0.001:67;
line(x,y,'Color','r','LineWidth',1);
text(0.025,0.89,'{\it n} = 42,984','Units','Normalized','fontname',...
'Arial','Color','r','FontSize',14)
text(0.025,0.96,'10-fold cross-validation {\it R^{2}} = 0.724', ...
'Units','Normalized','fontname','Arial','Color','r','FontSize',14)
text(-0.2,1.1,0,'c','Units', ...
'Normalized','fontname','Arial','Color','k','FontSize',24,'FontWeight','Bold');

% subplots (d)
ens = load('…\Best model object for regression on sodic class.mat');
ytrue_log = ens.ens.Y;
ytrue = 10.^(ytrue_log);
% Validation and Acquiring accuracy metrics
cvens = crossval(ens.ens,'Kfold',10);
```

```
yfit_log = kfoldPredict(cvens);
yfit = 10.^(yfit_log);
Variables = [ytrue yfit];
Variables_table = array2table(Variables,'VariableNames',{'ytrue' 'yfit'});
writetable(Variables_table,'…\ESP_regression_cross_validation.txt');
T = readtable('…\ESP_regression_cross_validation','FileType',...
                  'text','Delimiter',',','PreserveVariableNames',true);
ytrue = T.ytrue;
yfit = T.yfit;
subplot(2,3,5);
binscatter(ytrue,yfit,100)
set(gca,'XScale','log','YScale','log','fontname','Arial','FontSize',20)
title('Actual vs fitted: Sodic class','Color','k','FontSize',18);
ylabel('Predicted ESP (%)','Color','k');xlabel('Observed ESP (%)','Color','k');
xlim([1 102]);
ylim([1 160]);
xticks([1 5 10 100]);
yticks([1 5 10 100]);
colormap hot
c = colorbar;
c.Label.String = 'Scatter density in bins';
box on
ax = gca;
ax.LineWidth = 1;
ax.XColor = 'k';
ax.YColor = 'k';
hold on
x = 1:0.001:100;y = 1:0.001:100;
line(x,y,'Color','r','LineWidth',1);
text(0.025,0.89,'{\it n} =
197,988','Units','Normalized','fontname','Arial','Color','r','FontSize',14)
text(0.025,0.96,'10-fold cross-validation {\it R^{2}} = 0.726', ...
'Units','Normalized','fontname','Arial','Color','r','FontSize',14)
text(-0.2,1.1,0,'d','Units', ...
'Normalized','fontname','Arial','Color','k','FontSize',24,'FontWeight','Bold');

% subplot (e)
T = readtable('…\predicted_ECe','FileType',...
                 'text','Delimiter',',','PreserveVariableNames',true);
% T is the table including information on the present study predictions, HWSD predictions, and
surface measurements of ECe
ECe = T.ECe;
ECe_predicted = T.ECe_predicted;
t_ece_HWSD = T.ECe_HWSD;
subplot(2,3,3);
hold on
p1 = scatter(ECe,ECe_predicted,7,'filled','o','MarkerFaceColor', ...
[0.9290 0.6940 0.1250],'MarkerEdgeColor',[0.9290 0.6940 0.1250]);
p2 = scatter(ECe,t_ece_HWSD,7,'filled','o','MarkerFaceColor', ...
[0.6350 0.0780 0.1840],'MarkerEdgeColor',[0.6350 0.0780 0.1840]);
set(gca,'XScale','log','YScale','log','fontname','Arial','FontSize',20)
title('Soil salinity predictions','Color','k','FontSize',18)
box on
ax = gca;
ax.XColor = 'k';
ax.YColor = 'k';
ax.LineWidth = 1;
ylabel('Predicted EC_{e} (dS m^{-1})','Color','k');
xlabel('Observed EC_{e} (dS m^{-1})','Color','k');
xlim([0.01 110]);
ylim([0.01 110]);
xticks([0.1 1 10 80]);
yticks([0.1 1 10 80]);
x = 0.01:0.001:105;y = 0.01:0.001:105;
line(x,y,'Color','k','LineWidth',1);
text(0.025,0.93,'{\it n} = 9,293','Units','Normalized', ...
'fontname','Arial','Color','k','FontSize',14)
text(-0.16,1.1,0,'e','Units','Normalized', ...
'fontname','Arial','Color','k','FontSize',24,'FontWeight','Bold');
R_squared_ECe_present = sum((ECe_predicted - mean(ECe)).^2)/sum((ECe - mean(ECe)).^2);
R_squared_ECe_HWSD = sum((t_ece_HWSD - mean(ECe)).^2)/sum((ECe - mean(ECe)).^2);
L = legend([p1 p2],'Present study',...
    'HWSD','Location','southeast','FontSize',16);
L.Title.Color = 'k';
ax.YAxisLocation = 'right';

% subplot (e)
```

```
T1 = readtable('F:\Other_maps_comparison\Gound_points\predicted_ESP_HWSD','FileType',...
                 'text','Delimiter',',','PreserveVariableNames',true);
% T1 is the table including information on the present study predictions, HWSD predictions,
and surface measurements of ESP
subplot(2,3,6);
ESP_1 = T1.ESP;
ESP_predicted = T1.ESP_predicted;
t_esp_HWSD = T1.ESP_HWSD;
hold on
p1 = scatter(ESP_1,ESP_predicted,7,'filled','o','MarkerFaceColor', ...
[0.9290 0.6940 0.1250],'MarkerEdgeColor',[0.9290 0.6940 0.1250]);
p2 = scatter(ESP_1,t_esp_HWSD,7,'filled','o','MarkerFaceColor',[0.6350 0.0780 0.1840], ...
'MarkerEdgeColor',[0.6350 0.0780 0.1840]);
set(gca,'XScale','log','YScale','log','fontname','Arial','FontSize',20)
title('Soil sodicity predictions','Color','k','FontSize',18)
box on
ax = gca;
ax.LineWidth = 1; ax.YAxisLocation = 'right';
ax.XColor = 'k';
ax.YColor = 'k';
ylabel('Predicted ESP (%)','Color','k');
xlabel('Observed ESP (%)','Color','k');
xlim([0.01 600]);
ylim([0.01 1700]);
xticks([0.1 1 10 100]);
xticklabels({'0.1','1','10','100'});
yticklabels({'0.1','1','10','100'});
yticks([0.1 1 10 100]);
x = 0.01:0.001:2000;y = 0.01:0.001:2000;
line(x,y,'Color','k','LineWidth',1);
text(0.025,0.94,'{\it n}_{total} = 30,491','Units','Normalized','fontname', ...
'Arial','Color','k','FontSize',14)
text(-0.16,1.1,0,'f','Units','Normalized','fontname', ...
'Arial','Color','k','FontSize',24,'FontWeight','Bold');
R_squared_ESP_present_HWSD = sum((ESP_predicted - mean(ESP_1)).^2)/ ...
sum((ESP_1 - mean(ESP_1)).^2);
R_squared_ESP_HWSD = sum((t_esp_HWSD - mean(ESP_1)).^2)/ ...
sum((ESP_1 - mean(ESP_1)).^2);
L = legend([p1 p2],'Present study','HWSD','Location','southeast','FontSize',16);
```

References

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research, 16*, 321-357.

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science, 339*(6122), 940-943.

Marthews, T., Dadson, S., Lehner, B., Abele, S., & Gedney, N. (2014). A high-resolution global dataset of topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences Discussions, 11*, 6139-6166.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology, 10*(3), 282-290.

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G. Y., et al. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems, 8*(1), 41-65.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE, 37*(1), 10-21.

Appendix 3

Table A3-1: Predictor importance (percentage) in the final 16 best fitted models.

Model name	Sample upper depth	Sample lower depth	WRB	Clay content	Elevation	Slope	Field capacity	Wilting point	Effective plant rooting depth	dryss[a]	wetss[b]	evspsbl[c]	Pr-fre[d]	Pr-int[e]
CMIP5 models														
GISS-E2-H	3.31	4.03	16.43	8.56	7.43	7.58	4.32	4.24	5.42	3.30	10.77	6.20	11.74	6.66
GISS-E2-R	3.57	3.57	14.65	7.23	5.68	7.18	3.91	3.13	5.00	4.06	5.69	10.35	16.47	9.50
MIROC5-ensemble	4.37	4.56	12.61	10.58	6.48	7.60	3.85	3.70	5.47	3.99	7.00	10.10	13.70	5.99
MIROC-ESM-CHEM	5.36	4.70	14.14	8.72	10.77	6.18	3.45	4.84	5.02	6.39	5.47	8.12	12.43	4.39
MIROC-ESM	5.54	5.86	12.95	7.71	11.26	6.50	3.73	5.00	4.81	5.98	6.10	9.01	10.83	4.72
MRI-CGCM3	4.26	4.05	13.91	8.71	6.10	6.77	3.92	3.83	5.80	5.70	5.14	11.17	14.14	6.49
MRI-ESM1	4.18	4.32	14.19	9.52	6.00	7.11	3.79	3.69	5.61	4.93	5.10	10.23	15.48	5.85
NorESM1-M	5.36	4.96	11.38	7.31	6.18	8.11	3.95	3.53	5.95	4.60	4.69	13.46	14.94	5.57
CMIP6 models														
CESM2-WACCM-ensemble	4.91	4.84	11.05	7.35	6.60	7.24	4.08	3.45	5.57	3.71	10.93	9.74	13.97	6.56
CNRM-ESM2-1	5.39	4.75	12.83	7.95	6.20	7.22	3.70	3.40	6.08	5.52	5.28	9.00	17.40	5.27
GFDL-ESM4	5.51	5.22	11.79	8.80	7.72	6.52	3.71	4.88	3.68	6.46	5.84	8.23	16.77	4.88
INM-CM4-8-	4.82	5.61	11.69	7.37	6.74	7.75	4.40	3.61	5.75	4.51	10.40	8.83	12.80	5.73
INM-CM5-0	3.78	4.61	13.79	7.61	6.74	7.91	4.07	3.57	5.71	3.65	9.02	8.12	15.87	5.54
MIROC-ES2L	5.14	4.85	14.11	9.15	11.01	5.63	3.87	1.51	10.28	4.59	9.11	6.55	7.19	7.00
MRI-ESM2-0	5.54	5.15	10.92	6.47	10.71	6.43	4.11	4.54	4.31	6.06	5.12	8.19	17.08	5.36
NorESM2-LM	4.39	4.29	12.72	7.84	7.66	7.86	5.11	4.12	2.78	7.68	6.31	10.82	13.33	5.11
Average	4.71	4.71	13.07	8.18	7.71	7.10	4.00	3.82	5.45	5.07	7.00	9.26	14.01	5.91

[a] Five-year moving average of daily dry deposition rate of sea salts.
[b] Five-year moving average of daily wet deposition rate of sea salts.
[c] Five-year moving average of daily evapotranspiration.
[d] Five-year moving average of annual precipitation frequency.
[e] Five-year moving average of annual precipitation intensity.

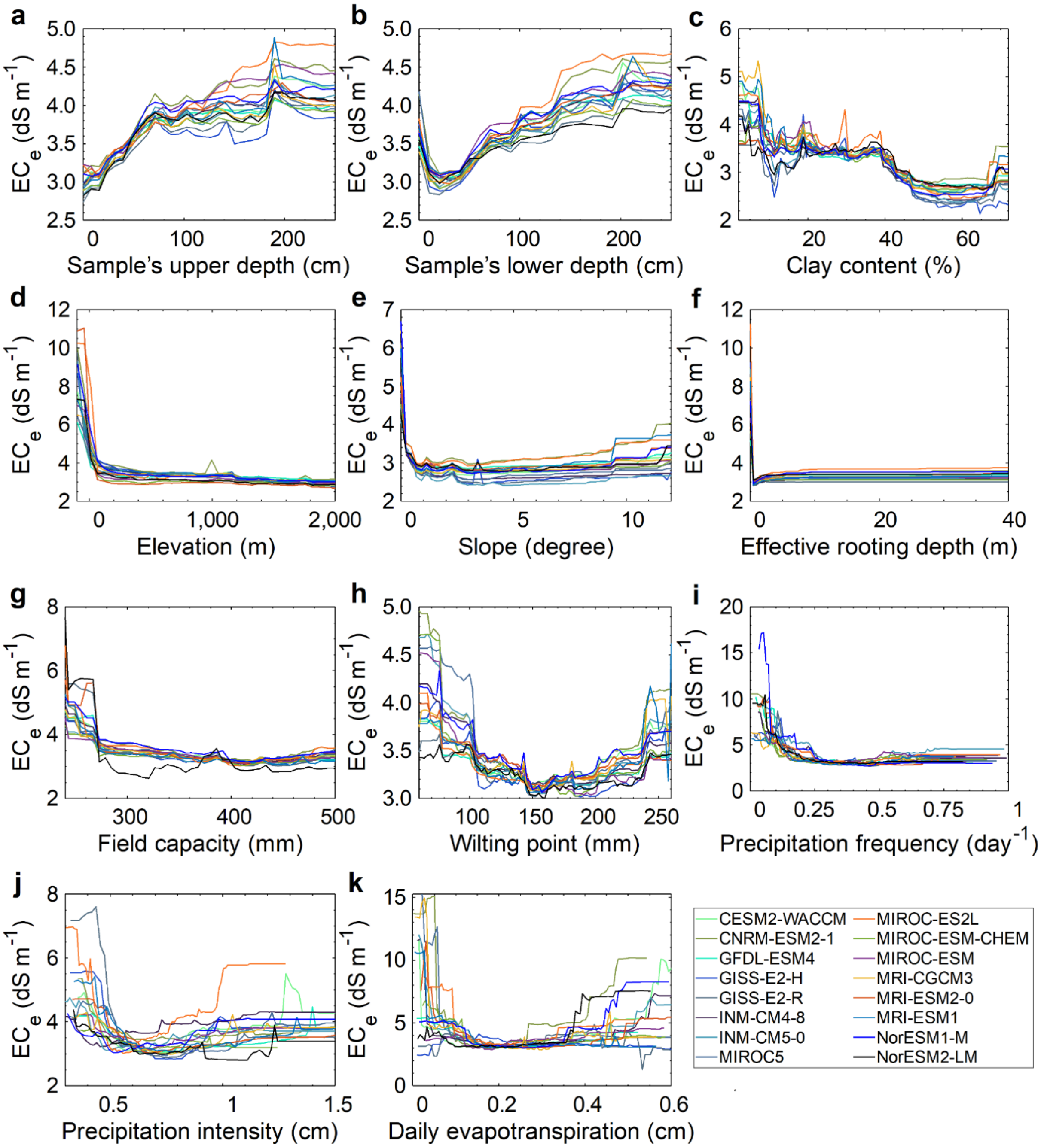


Figure A3-1: The relation between predicted soil salinity level (represented by EC_e) and the predictors used for training each of 16 predictive models of soil salinity. **a** to **h**, purely spatial predictors. **i** to **k**, long-term averages of the spatio-temporal predictors. The legend indicates the name of the corresponding original Global Circulation Models (GCMs) outputs.

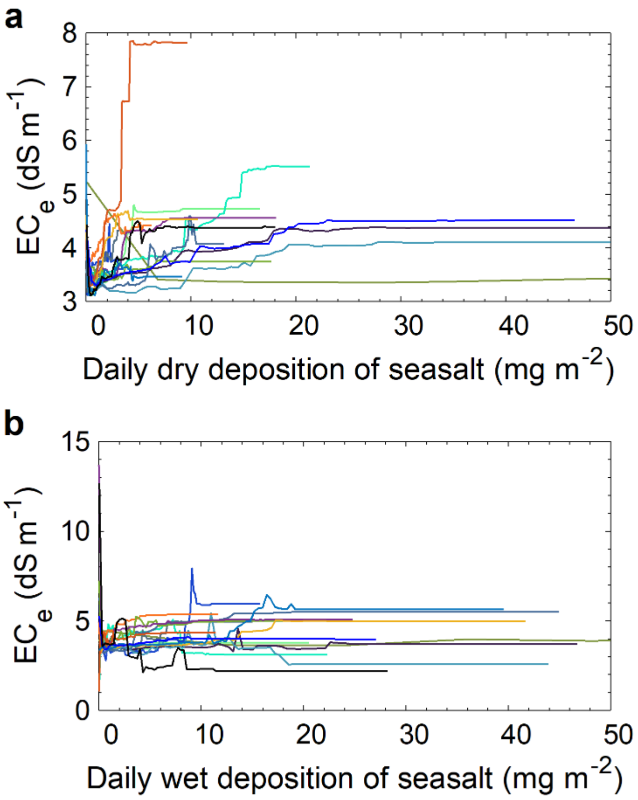


Figure A3-2: The relation between predicted soil salinity level (represented by EC$_e$) and long-term average of daily deposition rates of sea salts used as predictors for training each of 16 predictive models of soil salinity. a, dry deposition rate. **b**, wet deposition rate. Each line and colour is related to a model trained based on the original Global Circulation Models (GCMs) outputs (see Figure A3-1 for the full name of models and corresponding colours).

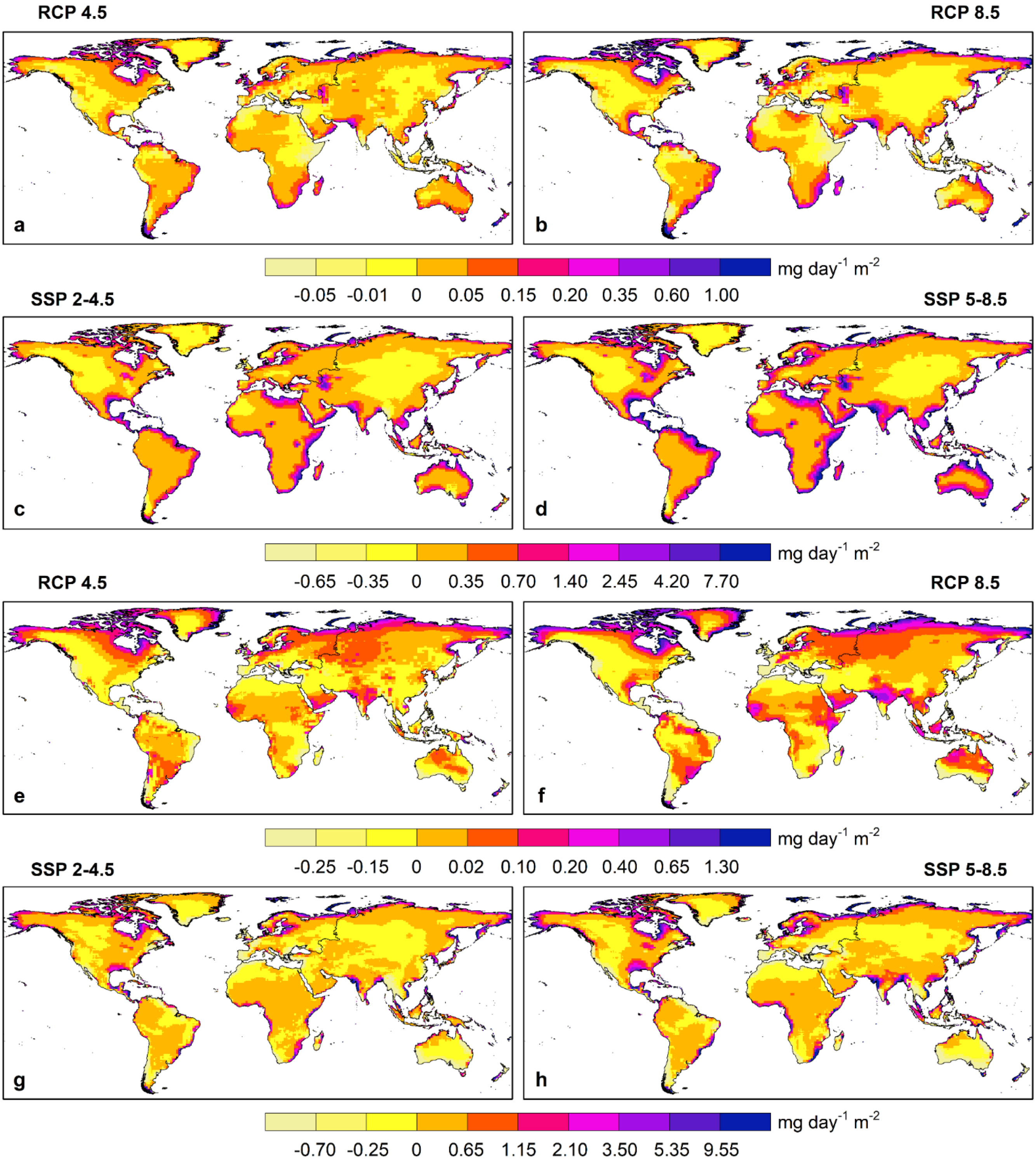


Figure A3-3: Multi-model ensemble mean of the change in deposition rates of sea salts by the end of the century (2071 - 2100), relative to the reference period (1961 - 1990) under different greenhouse gas concentration trajectories. a to **d**, change in the five-year moving average of daily dry deposition rate of sea salts. **e** to **h**, change in the five-year moving average of daily wet deposition rate of sea salts. Positive values indicate an increase in the deposition rate while negative values are indicative of a decrease.

Table A3-2: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the mid-term future (2031 - 2060) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under RCP 4.5 greenhouse gas concentration trajectory.

RCP 4.5 mid-term future (%)								
Country	**Mean**	**Meanci-low**[a]	**Meanci-up**[b]	**Sem**[c]	**Min**	**Max**	**STDEV**[d]	**Variance**
Namibia	6.55	5.60	7.50	0.48	-17.35	26.20	7.92	62.65
Brazil	4.86	3.69	6.04	0.60	-19.76	27.86	10.96	120.02
Mexico	4.22	3.31	5.14	0.46	-22.39	27.98	10.12	102.42
South Africa	4.21	3.34	5.08	0.44	-18.26	28.05	7.76	60.28
Niger	3.67	2.62	4.72	0.53	-25.84	27.98	9.94	98.78
Chad	3.19	2.09	4.28	0.56	-25.22	27.29	10.84	117.59
United States	3.17	2.66	3.67	0.26	-22.90	27.42	10.05	100.99
Mauritania	3.01	1.93	4.10	0.55	-25.78	27.30	9.80	95.97
Mali	2.40	1.18	3.62	0.62	-26.28	28.07	11.94	142.48
Argentina	2.12	1.43	2.82	0.35	-22.33	27.88	8.67	75.13
Pakistan	1.94	1.15	2.73	0.40	-23.64	21.93	6.63	43.96
China	1.60	1.17	2.04	0.22	-25.22	28.10	8.76	76.70
Libya	1.24	0.46	2.03	0.40	-22.54	26.50	9.30	86.52
Ethiopia	0.89	-0.08	1.86	0.49	-18.04	26.62	7.26	52.78
Botswana	0.57	-0.52	1.67	0.56	-20.02	28.01	7.97	63.50
Russia	0.57	0.13	1.01	0.22	-26.51	27.89	10.00	99.96
India	0.43	-0.34	1.21	0.40	-26.43	28.00	9.89	97.89
Australia	0.43	0.15	0.71	0.14	-25.55	28.08	7.00	49.06
Kazakhstan	0.39	-0.14	0.91	0.27	-25.83	28.07	9.31	86.71
Afghanistan	0.17	-0.44	0.79	0.31	-10.56	18.33	4.72	22.24
Canada	-0.19	-0.66	0.28	0.24	-22.69	24.51	6.92	47.93
Iran	-0.19	-0.76	0.38	0.29	-18.59	25.36	7.01	49.17
Bolivia	-0.69	-1.98	0.60	0.65	-24.65	25.54	9.18	84.29
Somalia	-0.80	-2.10	0.51	0.66	-26.90	27.68	9.41	88.57
Sudan	-1.72	-2.52	-0.93	0.40	-25.24	27.52	9.92	98.44
Turkey	-1.96	-3.12	-0.80	0.59	-20.59	27.02	9.24	85.45
Algeria	-3.88	-4.57	-3.18	0.35	-25.40	27.35	9.88	97.54
Egypt	-4.56	-5.22	-3.91	0.33	-18.03	25.50	6.40	41.00
Mongolia	-5.50	-6.61	-4.39	0.56	-26.74	25.34	11.96	143.02
Saudi Arabia	-5.81	-6.36	-5.27	0.28	-24.93	26.26	7.17	51.45

[a] Lower limit of the 95% confidence interval for the mean.

[b] Upper limit of the 95% confidence interval for the mean.

[c] Standard error of the mean.

[d] Standard Deviation.

Table A3-3: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the mid-term future (2031 - 2060) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under RCP 8.5 greenhouse gas concentration trajectory.

RCP 8.5 mid-term future (%)								
Country	**Mean**	**Meanci-low**	**Meanci-up**	**Sem**	**Min**	**Max**	**STDEV**	**Variance**
Namibia	7.15	6.31	7.99	0.43	-8.33	23.47	6.89	47.41
South Africa	5.36	4.37	6.35	0.50	-22.73	24.01	8.52	72.53
Brazil	4.76	3.77	5.76	0.51	-17.06	24.16	9.25	85.54
Botswana	3.38	2.29	4.48	0.56	-20.60	23.91	7.88	62.14
Mauritania	2.83	1.60	4.07	0.63	-20.95	23.95	10.88	118.34
United States	2.44	2.03	2.85	0.21	-24.52	24.01	8.16	66.54
Argentina	2.36	1.78	2.93	0.29	-18.92	23.41	7.21	52.00
Australia	2.19	1.96	2.42	0.12	-15.52	24.08	5.66	32.05
Niger	1.72	0.63	2.81	0.55	-24.22	23.50	10.02	100.41
Bolivia	1.56	0.55	2.57	0.51	-18.60	23.48	7.48	56.02
Mali	0.58	-0.47	1.63	0.54	-23.40	23.93	10.22	104.49
Mexico	0.51	-0.34	1.35	0.43	-22.47	22.93	9.37	87.71
Chad	0.33	-0.85	1.50	0.60	-24.92	23.89	11.33	128.32
Pakistan	-0.33	-1.02	0.35	0.35	-12.94	22.13	5.78	33.36
Libya	-0.63	-1.28	0.02	0.33	-18.39	23.93	7.77	60.37
Somalia	-0.94	-1.99	0.11	0.53	-18.35	23.68	7.64	58.41
China	-0.96	-1.44	-0.47	0.25	-24.84	23.81	9.47	89.69
Canada	-1.50	-1.94	-1.05	0.23	-21.11	23.34	6.61	43.67
India	-1.68	-2.23	-1.14	0.28	-23.26	23.84	7.11	50.48
Russia	-2.31	-2.65	-1.98	0.17	-23.66	23.24	7.66	58.65
Afghanistan	-2.39	-3.00	-1.77	0.31	-14.12	10.06	4.78	22.89
Kazakhstan	-2.55	-2.92	-2.18	0.19	-22.87	19.98	6.69	44.71
Sudan	-2.56	-3.21	-1.90	0.33	-21.44	24.12	8.30	68.92
Ethiopia	-2.59	-3.58	-1.60	0.50	-22.11	19.80	7.46	55.69
Turkey	-3.74	-4.72	-2.76	0.50	-23.70	22.09	7.89	62.29
Egypt	-4.59	-5.23	-3.94	0.33	-15.56	22.20	6.27	39.31
Algeria	-4.69	-5.33	-4.04	0.33	-20.08	24.11	8.99	80.79
Iran	-4.89	-5.44	-4.33	0.28	-23.86	23.11	6.84	46.84
Mongolia	-5.99	-7.00	-4.97	0.52	-24.58	23.89	10.93	119.39
Saudi Arabia	-6.92	-7.30	-6.54	0.19	-24.84	12.93	5.04	25.40

Table A3-4: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the mid-term future (2031 - 2060) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under SSP 2-4.5 greenhouse gas concentration trajectory.

SSP 2-4.5 mid-term future (%)								
Country	**Mean**	**Meanci-low**	**Meanci-up**	**Sem**	**Min**	**Max**	**STDEV**	**Variance**
Botswana	8.97	8.09	9.85	0.45	-3.03	22.05	6.11	37.29
Mongolia	7.48	6.37	8.58	0.56	-10.61	21.78	8.17	66.79
South Africa	6.24	5.46	7.02	0.40	-16.35	22.09	6.93	48.08
Brazil	5.93	5.15	6.70	0.39	-18.27	22.02	7.41	54.91
Australia	5.77	5.52	6.01	0.12	-14.28	22.05	5.96	35.47
Namibia	5.44	4.73	6.16	0.36	-11.54	21.98	5.96	35.55
Mauritania	2.32	1.83	2.80	0.25	-12.72	18.28	4.75	22.58
Bolivia	1.19	0.30	2.07	0.45	-19.41	19.69	6.58	43.31
China	0.75	0.33	1.18	0.22	-21.58	22.08	8.08	65.36
Mexico	0.70	0.13	1.26	0.29	-17.72	20.89	6.38	40.70
Algeria	0.30	-0.03	0.64	0.17	-10.70	19.37	4.89	23.90
Libya	0.14	-0.31	0.59	0.23	-22.20	20.91	5.50	30.23
Mali	0.03	-0.50	0.56	0.27	-21.86	17.93	5.46	29.84
Egypt	-0.24	-0.66	0.18	0.21	-9.57	15.81	4.07	16.59
Sudan	-0.59	-1.10	-0.07	0.26	-22.18	20.73	6.59	43.38
Saudi Arabia	-0.72	-1.24	-0.19	0.27	-18.86	18.83	7.00	48.99
Argentina	-0.92	-1.38	-0.47	0.23	-21.72	22.09	5.70	32.45
Afghanistan	-1.16	-1.85	-0.47	0.35	-16.47	19.73	5.36	28.77
Turkey	-1.61	-2.36	-0.86	0.38	-17.12	17.25	6.03	36.40
United States	-2.24	-2.62	-1.86	0.19	-21.30	22.01	7.83	61.23
Kazakhstan	-2.29	-2.65	-1.94	0.18	-21.53	22.04	6.43	41.32
Pakistan	-2.42	-3.28	-1.56	0.44	-19.23	13.76	7.29	53.10
Canada	-2.53	-2.86	-2.19	0.17	-20.11	12.00	4.98	24.77
Chad	-3.21	-3.90	-2.52	0.35	-22.31	17.93	7.02	49.21
Iran	-3.31	-3.80	-2.82	0.25	-18.82	22.06	6.08	36.97
Ethiopia	-3.78	-4.50	-3.05	0.37	-17.31	10.74	5.48	30.06
India	-4.22	-4.78	-3.67	0.28	-22.18	21.48	7.20	51.86
Niger	-4.58	-5.32	-3.84	0.38	-21.82	19.33	7.49	56.12
Russia	-5.50	-5.79	-5.20	0.15	-22.19	21.28	6.58	43.33
Somalia	-6.82	-7.66	-5.99	0.42	-20.25	19.53	6.13	37.55

Table A3-5: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the mid-term future (2031 - 2060) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under SSP 5-8.5 greenhouse gas concentration trajectory.

SSP 5-8.5 mid-term future (%)								
Country	**Mean**	**Meanci-low**	**Meanci-up**	**Sem**	**Min**	**Max**	**STDEV**	**Variance**
Botswana	9.53	8.56	10.50	0.49	-20.40	24.33	6.96	48.40
South Africa	8.54	7.70	9.39	0.43	-12.86	24.36	7.40	54.75
Brazil	8.00	7.12	8.87	0.45	-15.79	24.79	8.11	65.80
Namibia	6.62	5.77	7.47	0.43	-9.47	24.65	7.06	49.90
Mongolia	5.81	4.36	7.27	0.74	-19.68	24.29	10.91	118.96
Australia	5.75	5.51	5.99	0.12	-16.52	24.76	5.98	35.70
Mexico	3.55	2.86	4.25	0.35	-14.70	24.38	7.73	59.81
Mauritania	3.23	2.67	3.78	0.28	-10.70	24.69	5.40	29.14
Mali	1.23	0.58	1.88	0.33	-20.94	23.08	6.67	44.47
Bolivia	1.13	0.10	2.16	0.52	-21.25	22.92	7.64	58.31
Libya	0.80	0.32	1.27	0.24	-16.78	22.69	5.75	33.05
Algeria	0.69	0.24	1.14	0.23	-18.33	24.17	6.58	43.25
Afghanistan	0.21	-0.56	0.99	0.39	-13.76	21.02	6.01	36.16
China	0.09	-0.33	0.51	0.21	-22.80	24.78	8.20	67.26
Egypt	-0.05	-0.56	0.46	0.26	-11.99	22.86	4.98	24.85
Argentina	-0.60	-1.06	-0.14	0.24	-18.37	23.22	5.82	33.82
Sudan	-0.97	-1.63	-0.30	0.34	-22.45	24.63	8.47	71.79
United States	-1.08	-1.56	-0.61	0.24	-24.89	24.72	9.71	94.19
Turkey	-1.89	-2.89	-0.88	0.51	-21.19	23.61	8.02	64.39
Pakistan	-2.34	-3.17	-1.51	0.42	-20.40	24.14	7.01	49.08
Iran	-2.70	-3.18	-2.22	0.25	-16.43	24.31	5.94	35.29
Canada	-2.90	-3.26	-2.54	0.18	-21.20	16.10	5.39	29.06
Kazakhstan	-3.06	-3.41	-2.71	0.18	-21.43	24.55	6.27	39.31
India	-3.18	-3.79	-2.56	0.31	-24.23	24.80	7.96	63.30
Ethiopia	-3.26	-4.34	-2.19	0.55	-23.83	24.41	8.09	65.38
Saudi Arabia	-3.70	-4.22	-3.18	0.26	-22.36	23.92	6.85	46.93
Chad	-4.00	-4.83	-3.17	0.42	-24.21	24.47	8.42	70.87
Niger	-4.31	-5.13	-3.49	0.42	-21.62	24.00	8.22	67.56
Russia	-5.43	-5.73	-5.14	0.15	-24.71	23.85	6.71	44.99
Somalia	-7.28	-8.32	-6.25	0.52	-24.21	20.27	7.54	56.87

Table A3-6: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the long-term future (2071 - 2100) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under RCP 4.5 greenhouse gas concentration trajectory.

RCP 4.5 long-term future (%)								
Country	**Mean**	**Meanci-low**	**Meanci-up**	**Sem**	**Min**	**Max**	**STDEV**	**Variance**
Namibia	12.36	11.36	13.36	0.51	-9.85	33.00	8.14	66.32
South Africa	9.50	8.32	10.68	0.60	-13.59	32.63	10.17	103.47
Brazil	6.55	5.14	7.96	0.72	-31.68	33.33	13.05	170.34
Mexico	5.68	4.70	6.67	0.50	-25.42	32.06	10.91	119.02
Mauritania	4.08	2.60	5.56	0.75	-21.73	32.97	13.15	172.81
Mali	3.04	1.71	4.38	0.68	-29.36	33.42	13.01	169.31
Botswana	2.77	1.11	4.42	0.84	-18.75	33.07	11.86	140.73
United States	2.73	2.14	3.33	0.30	-34.46	33.47	11.97	143.35
Niger	2.66	1.35	3.97	0.66	-26.86	33.31	12.20	148.89
Libya	2.54	1.62	3.46	0.47	-23.31	33.43	10.99	120.83
Chad	2.54	1.13	3.94	0.72	-31.16	33.43	13.80	190.35
Australia	2.26	1.93	2.59	0.17	-25.25	33.41	8.30	68.91
Argentina	2.01	1.27	2.74	0.37	-22.38	30.27	9.23	85.20
Pakistan	0.94	0.09	1.79	0.43	-28.16	29.71	7.16	51.29
China	0.34	-0.23	0.90	0.29	-34.53	33.48	11.37	129.19
Ethiopia	0.21	-0.92	1.34	0.57	-25.69	27.68	8.48	71.89
Bolivia	-0.24	-1.46	0.97	0.62	-21.33	26.61	8.96	80.26
Turkey	-0.24	-1.77	1.28	0.77	-23.34	31.18	12.01	144.31
Afghanistan	-0.33	-1.14	0.47	0.41	-12.54	27.95	6.27	39.29
Iran	-0.53	-1.17	0.11	0.33	-21.01	27.60	7.92	62.72
Somalia	-0.54	-2.09	1.00	0.78	-32.74	30.66	11.19	125.12
India	-0.76	-1.60	0.08	0.43	-31.17	33.07	10.69	114.37
Sudan	-1.81	-2.65	-0.97	0.43	-29.66	33.03	10.75	115.46
Canada	-2.87	-3.61	-2.13	0.38	-32.82	31.59	11.04	121.85
Algeria	-3.29	-4.17	-2.41	0.45	-32.13	33.55	12.02	144.42
Kazakhstan	-4.59	-5.16	-4.02	0.29	-34.60	31.66	10.28	105.77
Egypt	-5.18	-6.14	-4.21	0.49	-30.94	33.02	9.37	87.82
Russia	-6.25	-6.79	-5.70	0.28	-34.46	33.59	12.53	157.03
Mongolia	-7.46	-8.54	-6.39	0.55	-33.54	27.81	11.67	136.13
Saudi Arabia	-9.42	-10.02	-8.83	0.30	-32.65	33.41	7.86	61.75

Table A3-7: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the long-term future (2071 - 2100) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under RCP 8.5 greenhouse gas concentration trajectory.

	RCP 8.5 long-term future (%)							
Country	**Mean**	**Meanci-low**	**Meanci-up**	**Sem**	**Min**	**Max**	**STDEV**	**Variance**
Brazil	15.10	13.25	16.95	0.94	-27.92	39.92	15.09	227.64
Namibia	13.57	12.10	15.04	0.75	-8.74	39.86	11.80	139.24
South Africa	11.20	9.41	13.00	0.91	-21.17	39.96	14.48	209.64
Mexico	6.38	4.96	7.80	0.72	-40.77	39.84	15.17	230.00
Botswana	6.06	4.28	7.83	0.90	-19.94	39.66	12.52	156.81
Mauritania	5.90	3.99	7.81	0.97	-22.07	39.73	16.50	272.32
Argentina	4.76	3.83	5.69	0.47	-26.49	39.38	11.68	136.47
Bolivia	4.34	2.53	6.15	0.92	-21.73	39.11	13.31	177.22
United States	3.70	3.02	4.39	0.35	-34.41	39.72	13.66	186.59
Australia	3.31	2.88	3.73	0.22	-21.55	39.96	10.57	111.67
Libya	2.80	1.38	4.22	0.72	-21.55	39.95	15.91	253.12
Somalia	2.77	0.98	4.56	0.91	-30.59	29.54	12.92	167.00
Pakistan	2.69	1.83	3.54	0.43	-16.62	28.51	7.23	52.24
Chad	2.38	0.72	4.03	0.84	-31.89	40.02	16.31	265.92
Niger	2.19	0.88	3.51	0.67	-25.42	37.94	12.50	156.19
Ethiopia	1.66	0.08	3.24	0.80	-29.61	38.14	11.76	138.39
Mali	1.05	-0.41	2.52	0.75	-27.74	39.35	14.61	213.55
Turkey	-0.39	-2.55	1.77	1.10	-29.74	39.12	16.72	279.59
China	-0.80	-1.60	-0.01	0.41	-41.24	40.02	15.51	240.69
Sudan	-1.28	-2.42	-0.14	0.58	-33.08	39.15	14.51	210.49
Canada	-2.09	-2.91	-1.28	0.41	-26.59	35.76	12.10	146.53
Afghanistan	-2.12	-2.85	-1.39	0.37	-14.46	19.16	5.66	32.01
India	-2.20	-2.89	-1.50	0.35	-23.77	36.78	9.07	82.33
Kazakhstan	-5.01	-5.52	-4.49	0.26	-33.99	39.59	9.25	85.47
Russia	-6.13	-6.61	-5.65	0.25	-32.31	40.05	11.04	121.79
Egypt	-6.47	-7.65	-5.29	0.60	-21.77	34.51	11.51	132.47
Iran	-6.99	-7.70	-6.28	0.36	-33.32	32.16	8.82	77.71
Algeria	-7.33	-8.35	-6.31	0.52	-29.26	39.74	13.87	192.47
Mongolia	-8.78	-10.22	-7.34	0.73	-37.41	37.14	15.58	242.64
Saudi Arabia	-12.96	-13.45	-12.46	0.25	-34.86	18.45	6.56	43.08

Table A3-8: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the long-term future (2071 - 2100) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under SSP 2-4.5 greenhouse gas concentration trajectory.

	SSP 2-4.5 long-term future (%)							
Country	**Mean**	**Meanci-low**	**Meanci-up**	**Sem**	**Min**	**Max**	**STDEV**	**Variance**
South Africa	10.49	9.51	11.47	0.50	-12.10	29.16	8.48	71.88
Botswana	9.73	8.84	10.62	0.45	-4.79	25.48	6.40	40.93
Australia	9.31	8.97	9.65	0.17	-12.48	29.42	8.22	67.60
Namibia	8.57	7.70	9.44	0.44	-7.92	29.00	7.31	53.46
Brazil	7.78	6.73	8.83	0.53	-20.67	29.35	10.05	101.01
Mongolia	7.71	6.07	9.36	0.83	-13.39	28.69	11.42	130.38
Mexico	5.40	4.57	6.23	0.42	-17.17	29.15	9.27	86.01
Mauritania	4.05	3.38	4.72	0.34	-18.78	25.51	6.53	42.70
Mali	2.41	1.82	3.00	0.30	-15.53	22.84	6.12	37.44
Bolivia	1.57	0.50	2.64	0.54	-17.18	24.38	7.97	63.49
Algeria	1.04	0.58	1.50	0.24	-13.30	28.53	6.74	45.44
Sudan	1.01	0.37	1.65	0.33	-19.91	29.35	8.19	67.11
Libya	0.89	0.34	1.45	0.28	-26.50	29.13	6.67	44.46
Afghanistan	0.60	-0.20	1.40	0.41	-14.23	22.22	6.20	38.47
Egypt	0.35	-0.15	0.86	0.26	-12.05	15.36	4.89	23.93
China	-0.15	-0.67	0.37	0.26	-26.51	29.37	9.82	96.47
Argentina	-1.03	-1.58	-0.48	0.28	-28.14	27.12	6.95	48.32
United States	-1.32	-1.85	-0.80	0.27	-28.23	29.30	10.78	116.18
Saudi Arabia	-1.80	-2.30	-1.31	0.25	-19.13	16.06	6.62	43.79
Chad	-1.84	-2.60	-1.09	0.38	-26.42	23.57	7.73	59.71
Pakistan	-1.93	-2.73	-1.12	0.41	-16.99	25.26	6.82	46.48
Iran	-2.97	-3.47	-2.48	0.25	-18.90	18.71	6.11	37.35
Turkey	-3.09	-4.21	-1.98	0.57	-22.71	27.78	8.90	79.15
Niger	-3.23	-3.98	-2.48	0.38	-22.85	26.55	7.69	59.06
India	-4.23	-4.93	-3.52	0.36	-28.68	25.30	9.14	83.59
Kazakhstan	-4.30	-4.78	-3.81	0.25	-25.59	29.33	8.72	76.04
Canada	-4.76	-5.19	-4.33	0.22	-27.63	20.49	6.45	41.54
Ethiopia	-4.92	-6.20	-3.64	0.65	-26.72	26.95	9.58	91.73
Russia	-6.95	-7.30	-6.60	0.18	-28.39	28.68	7.81	60.96
Somalia	-9.91	-11.19	-8.63	0.65	-28.19	24.05	9.28	86.16

Table A3-9: Computed statistics based on the multi-model ensemble mean of the predicted gird-cell level relative change in soil salinity (EC_e) in the long-term future (2071 - 2100) relative to the reference period (1961 - 1990) for the 30 countries with the highest number of dryland grid-cells in our analysis under SSP 5-8.5 greenhouse gas concentration trajectory.

	SSP 5-8.5 long-term future (%)							
Country	**Mean**	**Meanci-low**	**Meanci-up**	**Sem**	**Min**	**Max**	**STDEV**	**Variance**
Botswana	24.94	22.71	27.16	1.12	-5.89	44.10	13.10	171.74
South Africa	21.35	19.84	22.85	0.76	-15.09	43.32	12.51	156.41
Namibia	17.69	16.14	19.24	0.79	-7.07	43.52	12.54	157.13
Brazil	16.21	14.77	17.66	0.73	-16.80	44.09	13.07	170.80
Australia	14.60	14.17	15.03	0.22	-41.51	44.12	10.39	107.91
Mexico	12.66	11.38	13.94	0.65	-19.32	43.38	13.87	192.43
Mongolia	6.18	3.67	8.69	1.27	-34.40	44.21	20.04	401.77
Mali	5.76	4.78	6.73	0.50	-14.98	42.29	10.10	102.05
Mauritania	4.12	3.25	4.99	0.44	-14.54	31.63	8.45	71.46
Afghanistan	3.15	2.17	4.14	0.50	-17.56	25.87	7.64	58.37
Libya	2.87	2.20	3.55	0.35	-19.15	41.13	8.14	66.26
Bolivia	2.37	0.82	3.92	0.79	-19.19	42.96	11.55	133.39
Argentina	1.57	0.82	2.32	0.38	-21.11	43.57	9.40	88.40
Egypt	0.44	-0.14	1.01	0.29	-13.14	34.62	5.62	31.57
Sudan	0.43	-0.47	1.32	0.46	-24.87	43.55	11.46	131.34
Turkey	-0.04	-1.82	1.74	0.90	-26.92	42.28	14.18	201.19
United States	-0.11	-0.89	0.67	0.40	-39.52	42.90	15.91	253.14
China	-0.72	-1.39	-0.06	0.34	-38.33	43.81	13.09	171.32
Algeria	-1.14	-1.84	-0.44	0.36	-22.39	42.55	10.10	102.10
Pakistan	-1.15	-2.35	0.05	0.61	-20.74	37.28	10.14	102.91
India	-1.92	-2.76	-1.07	0.43	-27.94	42.21	10.96	120.06
Saudi Arabia	-2.64	-3.33	-1.96	0.35	-24.21	21.90	9.13	83.42
Iran	-2.79	-3.49	-2.09	0.35	-23.79	40.07	8.60	73.92
Ethiopia	-2.95	-4.88	-1.02	0.98	-30.30	41.58	14.40	207.24
Chad	-4.89	-5.91	-3.87	0.52	-28.99	39.68	10.48	109.82
Niger	-5.68	-6.88	-4.49	0.61	-40.41	41.24	12.13	147.14
Kazakhstan	-7.00	-7.61	-6.39	0.31	-32.99	43.51	11.05	121.99
Canada	-7.10	-7.60	-6.59	0.26	-31.75	24.19	7.54	56.80
Russia	-7.95	-8.44	-7.46	0.25	-40.19	43.63	11.21	125.58
Somalia	-12.03	-13.84	-10.22	0.92	-37.56	36.36	13.30	176.98

Table A3-10: Countries with the highest number of dryland grid-cells and the estimates of their total dryland area used in the present study.

Country	Number of grid-cells	Total dryland area (km^2)
Australia	2,421	6,717,230.56
United States	1,647	3,907,880.47
China	1,595	3,697,849.00
Russia	2,056	3,660,984.24
Kazakhstan	1,271	2,619,580.23
Algeria	825	2,243,524.51
Saudi Arabia	684	1,921,997.39
Sudan	633	1,870,833.54
India	656	1,859,113.43
Libya	582	1,597,000.93
Canada	862	1,555,847.54
Iran	590	1,537,911.25
Argentina	620	1,490,518.78
Mexico	511	1,414,650.16
Brazil	429	1,294,077.03
Mali	418	1,226,669.13
Chad	409	1,211,794.68
Niger	404	1,186,006.78
Mauritania	369	1,065,585.74
Egypt	366	1,009,647.31
Mongolia	463	999,368.89
South Africa	317	853,035.45
Namibia	281	801,629.98
Pakistan	278	749,601.24
Ethiopia	222	675,178.49
Somalia	210	641,504.11
Bolivia	217	634,507.78
Turkey	253	608,965.38
Afghanistan	234	601,733.75
Botswana	206	588,102.91

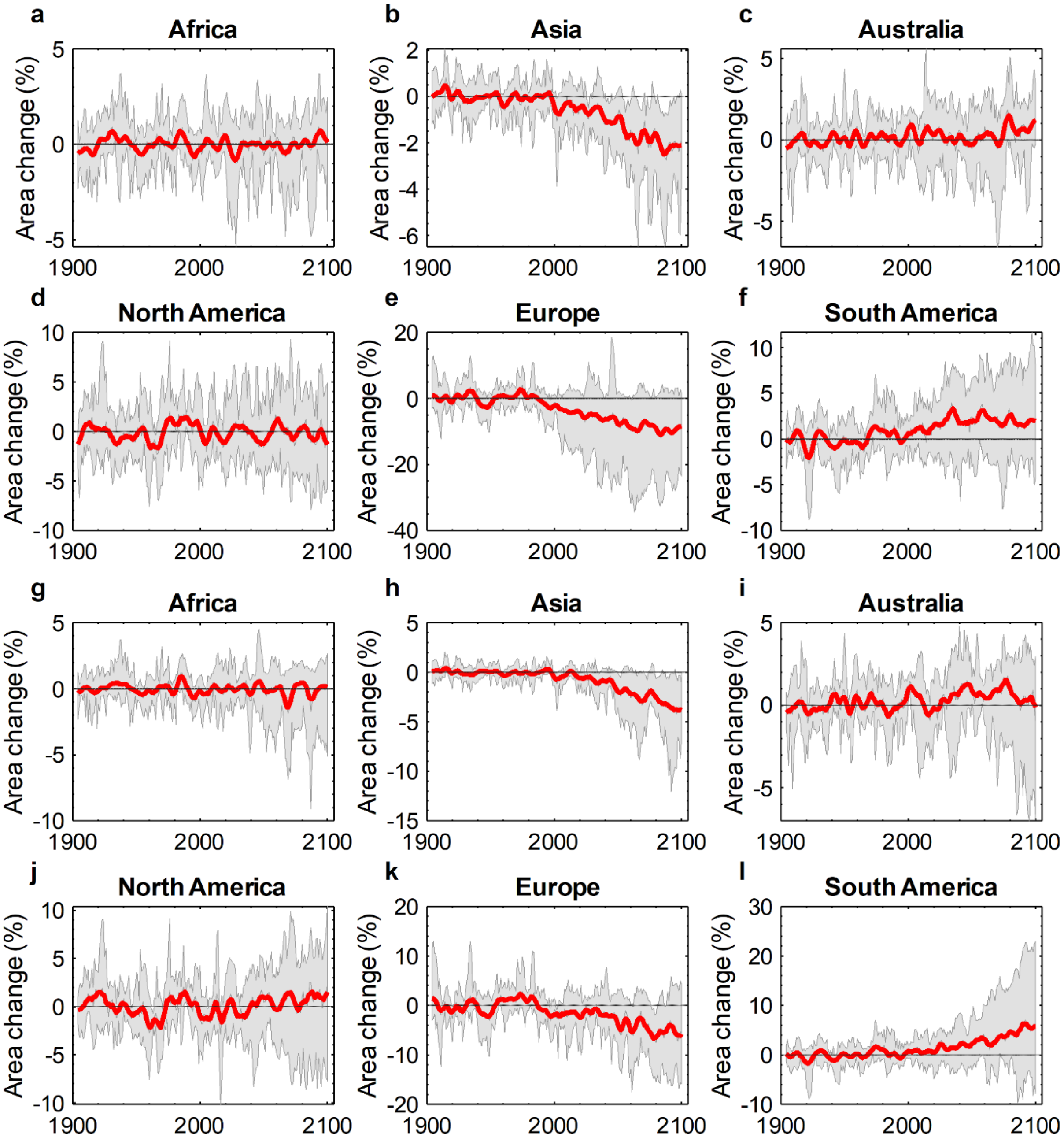


Figure A3-4: Continental-level predicted annual change in the total area of soils with an $EC_e \geq 2$ dS m^{-1} relative to the 20th century average (1904 - 1999) for the models obtained from CMIP5 data project. a to **f**: Relative change under RCP 4.5 greenhouse gas concentration trajectory. **g** to **l**: Relative change under RCP 8.5 greenhouse gas concentration trajectory. Shaded areas show the minimum and maximum range of the relative changes predicted by multi-model ensemble members. Red lines show the low-pass filtered (5-year running window) of the multi-model ensemble mean of the predicted variations; since all spatio-temporal predictors are five year moving averages, 1904 is the beginning of the period.

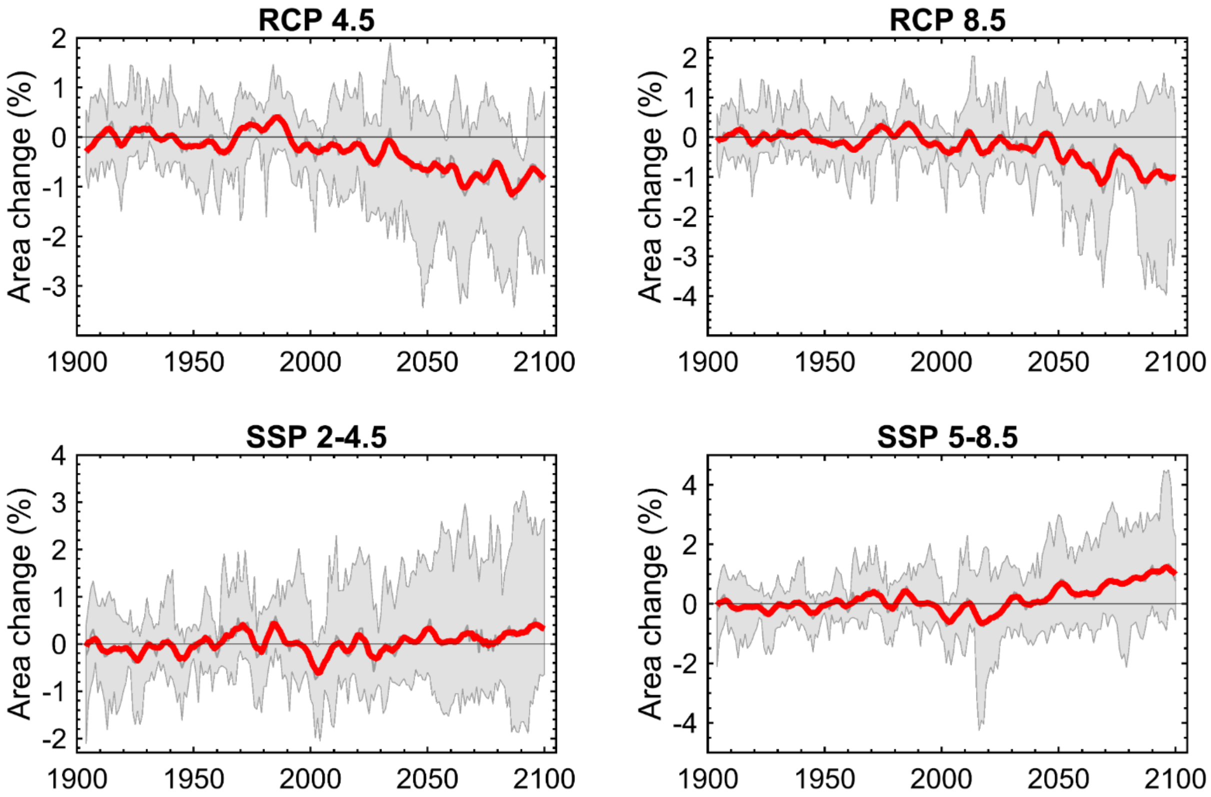


Figure A3-5: Predicted annual change in the total area of soils with an $EC_e \geq 2$ dS m^{-1} relative to the average of 1904 - 1999 period under different greenhouse gas concentration trajectories on the global scale. Shaded areas show the minimum and maximum range of the relative changes predicted by multi-model ensemble members. Red lines show the low-pass filtered (5-year running window) of the multi-model ensemble mean of the predicted variations; since all spatio-temporal predictors are five year moving averages, 1904 is the beginning of the period.

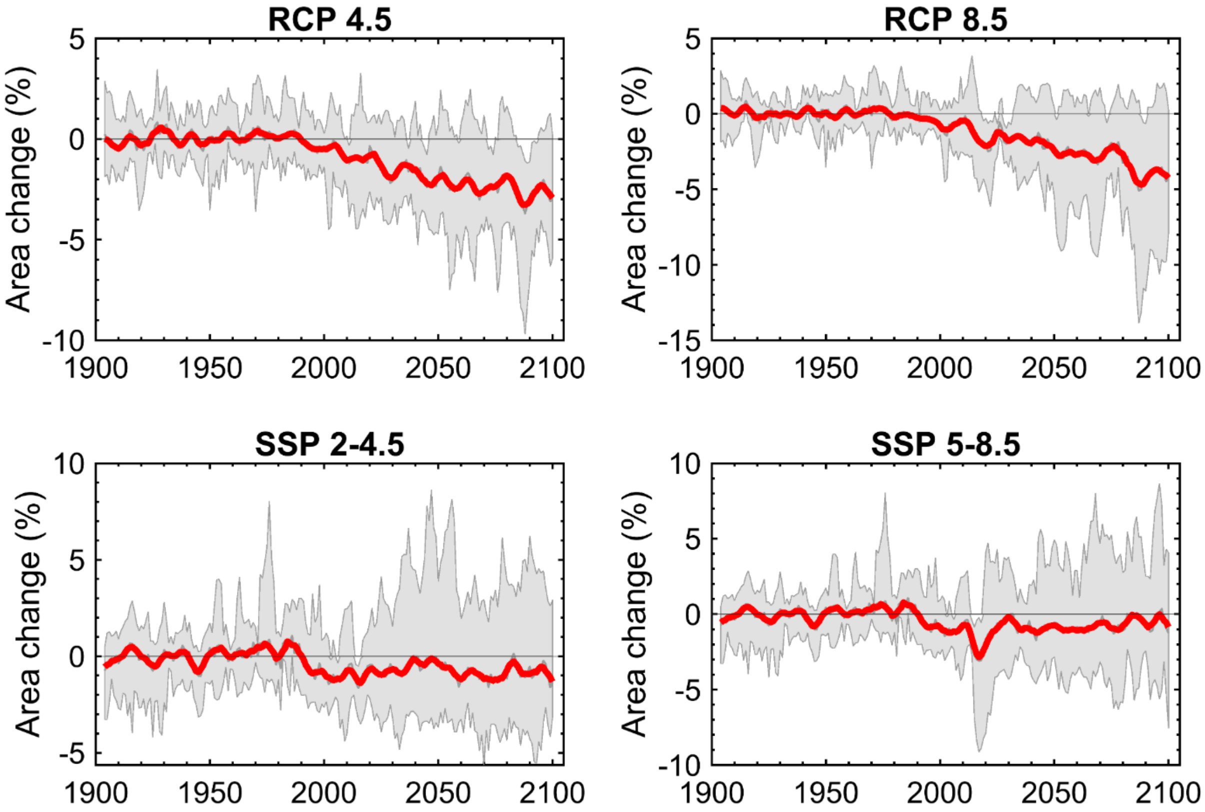


Figure A3-6: Predicted annual change in the total area of soils with an EC$_e$ ≥ 4 dS m^{-1} relative to the average of 1904 - 1999 period under different greenhouse gas concentration trajectories on the global scale. Shaded areas show the minimum and maximum range of the relative changes predicted by multi-model ensemble members. Red lines show the low-pass filtered (five-year running window) of the multi-model ensemble mean of the predicted variations; since all spatio-temporal predictors are five-year moving averages, 1904 is the beginning of the period.

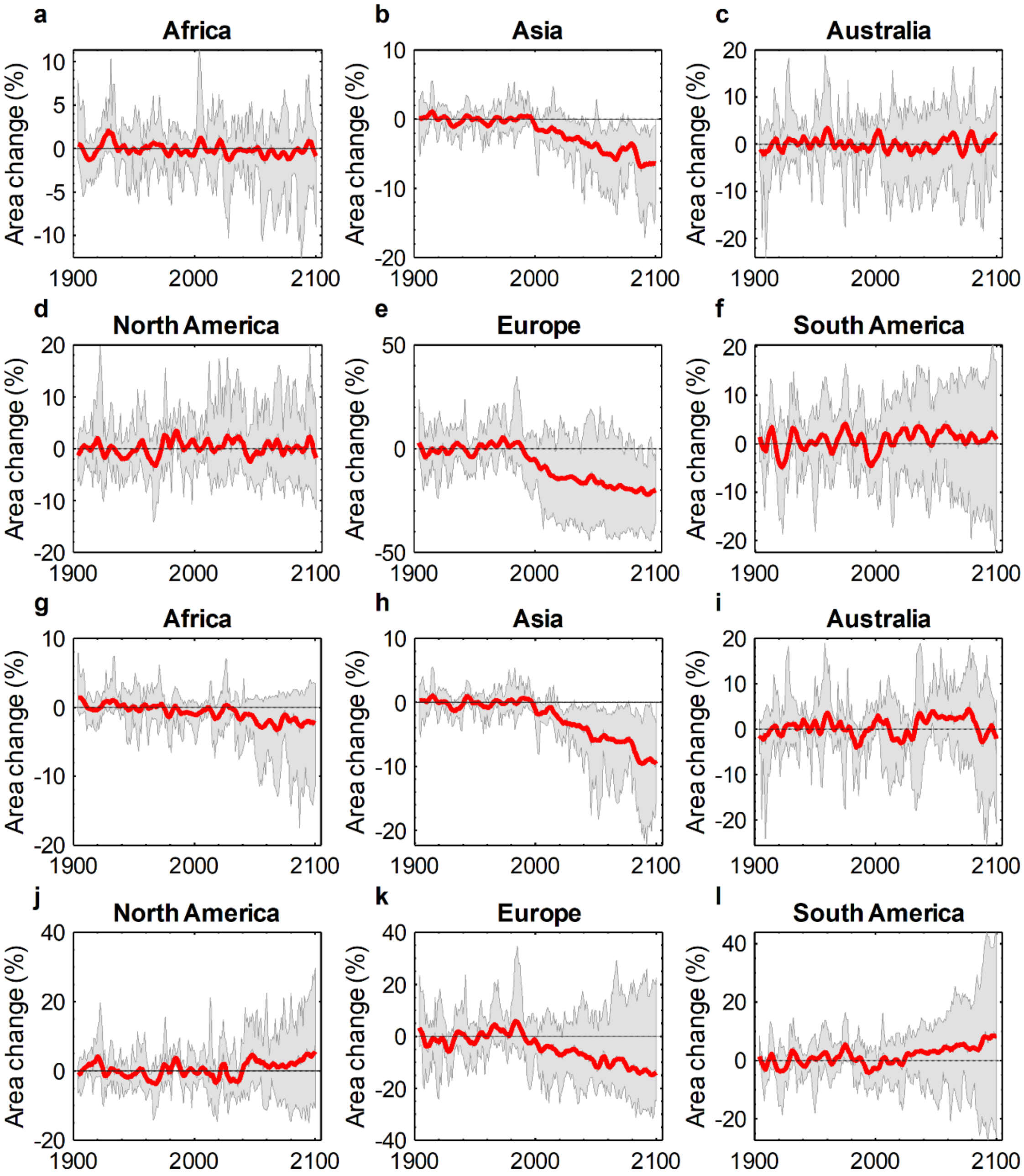


Figure A3-7: Continental-level predicted annual change in the total area of soils with an $EC_e \geq 4$ dS m^{-1} relative to the 20th century average (1904 - 1999) for the models obtained from the CMIP5 data. a to **f**: Relative change under RCP 4.5 greenhouse gas concentration trajectory. **g** to **l**: Relative change under RCP 8.5 greenhouse gas concentration trajectory. Shaded areas show the minimum and maximum range of the relative changes predicted by multi-model ensemble members. Red lines show the low-pass filtered (five-year running window) of the multi-model ensemble mean of the predicted variations; since all spatio-temporal predictors are five-year moving averages, 1904 is the beginning of the period.

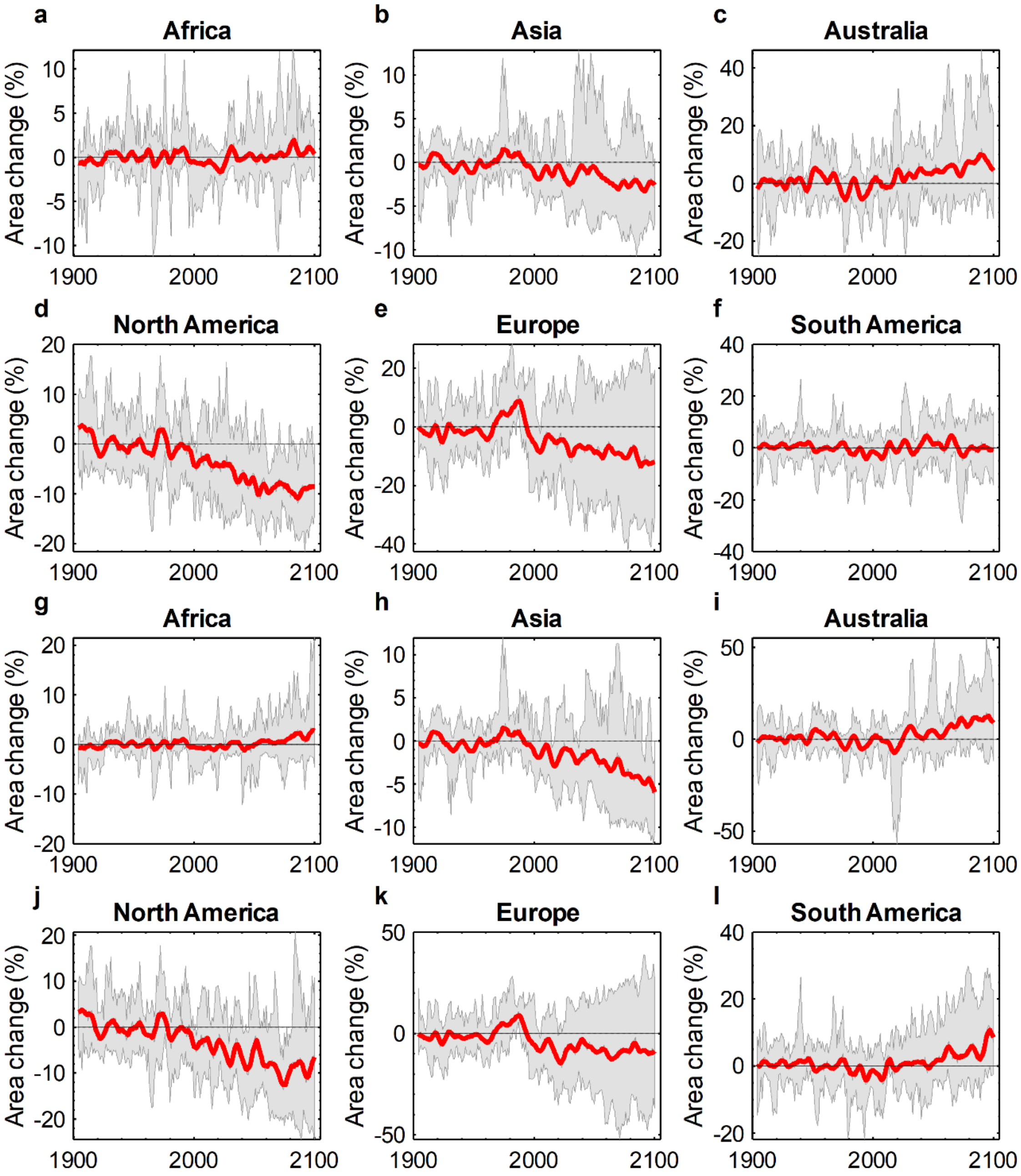


Figure A3-8: Continental-level predicted annual change in the total area of soils with an $EC_e \geq 4$ dS m^{-1} relative to the 20th century average (1904 - 1999) for the models obtained from CMIP6 data project. a to **f**: Relative change under SSP 2-4.5 greenhouse gas concentration trajectory. **g** to **l**: Relative change under SSP 5-8.5 greenhouse gas concentration trajectory. Shaded areas show the minimum and maximum range of the relative changes predicted by multi-model ensemble members. Red lines show the low-pass filtered (five-year running window) of the multi-model ensemble mean of the predicted variations; since all spatio-temporal predictors are five-year moving averages, 1904 is the beginning of the period.

Table A3-11: Country-level predicted change in the total area of soils with an $EC_e \geq 2$ dS m^{-1} at the long-term future (2071 - 2100), compared to the 20th century average (1904 - 1999) under different greenhouse gas concentration trajectories.

Country	Scenarios							
	RCP 4.5 mid-term future (%)	RCP 4.5 long-term future (%)	RCP 8.5 mid-term future (%)	RCP 8.5 long-term future (%)	SSP 2-4.5 mid-term future (%)	SSP 2-4.5 long-term future (%)	SSP 5-8.5 mid-term future (%)	SSP 5-8.5 long-term future (%)
Afghanistan	-1.69	-2.55	-3.05	-4.29	-0.60	-0.33	-0.45	-0.28
Algeria	0.38	0.98	0.57	-0.03	0.36	0.48	0.55	0.68
Argentina	0.02	-1.68	-0.12	-0.36	-0.95	-1.03	-0.53	-0.11
Australia	0.02	0.70	0.79	0.60	1.59	2.40	1.36	3.38
Bolivia	-1.33	-2.79	-0.03	-1.52	0.47	0.30	-0.06	-0.54
Botswana	1.03	3.63	3.74	4.82	5.59	5.11	5.13	7.52
Brazil	12.19	14.32	8.57	29.50	18.69	23.92	24.81	43.08
Canada	-4.22	-5.86	-3.19	-5.85	-5.01	-6.96	-5.38	-9.98
Chad	-2.47	-2.23	-2.02	-3.20	1.19	1.17	0.50	0.81
China	-1.00	-2.18	-1.65	-2.37	0.73	1.03	0.88	1.97
Egypt	-2.15	-2.25	-2.18	-5.74	0.21	0.23	0.15	0.21
Ethiopia	0.39	1.31	-0.08	3.83	-3.11	-6.23	-3.89	-4.99
India	-1.30	-1.60	-1.63	-0.92	-3.20	-3.24	-2.48	-2.94
Iran	0.20	0.49	-0.49	-1.88	-0.64	-1.45	-0.98	-1.47
Kazakhstan	-1.02	-2.44	-0.27	-2.58	-0.14	-1.09	-0.15	-2.52
Libya	0.26	-0.07	-1.34	1.13	0.55	0.69	0.58	0.47
Mali	0.17	0.37	-0.68	-1.97	-0.76	-0.23	-0.77	0.33
Mauritania	1.43	1.88	2.47	4.85	-0.02	0.61	0.55	0.60
Mexico	4.32	6.54	0.29	12.57	0.69	6.75	5.13	14.51
Mongolia	-5.71	-9.22	-10.49	-13.30	4.59	7.29	5.76	8.05
Namibia	2.30	3.63	2.56	4.40	1.25	1.80	1.75	2.72
Niger	1.55	1.14	1.34	0.28	0.09	0.34	0.89	0.53
Pakistan	0.03	0.05	0.09	0.29	0.26	0.27	0.30	0.33
Russia	-4.45	-7.93	-2.50	-7.82	-1.71	-2.88	-2.01	-4.16
Saudi Arabia	-0.46	-1.48	-0.68	-2.63	0.28	0.20	0.04	0.56
Somalia	-1.75	-1.90	-1.82	0.88	-3.42	-6.22	-4.19	-8.53
South Africa	1.02	3.88	2.92	5.65	1.30	3.45	2.94	7.18
Sudan	-1.51	-2.15	-1.92	-4.16	-0.25	0.43	0.25	1.22
Turkey	-2.36	-1.41	-5.21	-5.35	-0.35	-2.83	-1.03	-1.00
United States	1.22	1.54	2.12	2.30	-3.59	-3.10	-2.57	-2.63

Table A3-12: Continental-level predicted change in the total area of soils with an $EC_e \geq 4$ dS m^{-1} in the mid and long-term futures relative to the average of the 1904 - 1999 period under different greenhouse gas concentration trajectories.

Scenarios	Continent					
	Africa	Asia	Australia	North America	Europe	South America
RCP 4.5; mid-term (%)	-0.37	-3.73	-0.50	0.04	-15.72	2.27
RCP 4.5; long-term (%)	-0.32	-5.35	0.26	0.00	-20.07	1.03
RCP 8.5; mid-term (%)	-1.60	-4.80	2.38	1.21	-8.26	3.26
RCP 8.5; long-term (%)	-1.99	-7.99	0.65	3.04	-12.77	5.97
SSP 2-4.5; mid-term (%)	0.08	-1.17	3.76	-7.50	-8.14	1.00
SSP 2-4.5; long-term (%)	0.76	-2.56	6.78	-9.10	-10.79	-0.75
SSP 5-8.5; mid-term (%)	0.07	-1.89	2.94	-6.34	-6.96	1.09
SSP 5-8.5; long-term (%)	1.60	-4.26	10.40	-9.54	-8.25	5.19

Table A3-13: Country-level predicted change in the total area of soils with an $EC_e \geq 4$ dS m^{-1} at the long-term future (2071 - 2100), compared to the 20th century average (1904 - 1999) under different greenhouse gas concentration trajectories.

Country	**Scenarios**							
	RCP 4.5 mid-term future (%)	**RCP 4.5 long-term future (%)**	**RCP 8.5 mid-term future (%)**	**RCP 8.5 long-term future (%)**	**SSP 2-4.5 mid-term future (%)**	**SSP 2-4.5 long-term future (%)**	**SSP 5-8.5 mid-term future (%)**	**SSP 5-8.5 long-term future (%)**
Afghanistan	-3.20	-4.57	-6.90	-10.31	-1.23	0.71	0.07	3.36
Algeria	-4.30	-3.74	-5.58	-12.48	-0.62	-1.90	-3.67	-2.91
Argentina	3.32	-0.24	3.06	2.70	-3.20	-7.19	-3.93	-4.92
Australia	-0.50	0.26	2.38	0.65	3.76	6.78	2.94	10.40
Bolivia	-2.78	-3.61	-0.96	-3.77	-3.65	-6.55	-5.53	-9.02
Botswana	-1.50	-0.34	2.93	4.28	14.37	11.59	15.79	25.44
Brazil	2.10	1.45	4.48	27.61	14.73	20.31	19.85	62.34
Canada	-3.82	-4.44	-7.06	-7.68	-9.45	-14.06	-10.30	-19.17
Chad	1.18	0.86	-0.84	3.28	-7.50	-4.54	-8.43	-9.06
China	-0.10	-0.59	-2.73	0.28	-1.32	-4.08	-1.93	-9.77
Egypt	-4.85	-7.45	-5.74	-11.77	0.13	-0.44	0.18	0.03
Ethiopia	-8.68	-9.84	-9.24	8.80	-7.51	-10.80	-9.37	-12.79
India	-4.81	-5.91	-4.98	-11.19	-5.97	-7.23	-5.30	-6.08
Iran	-3.54	-3.66	-6.96	-10.73	-2.12	-2.88	-2.16	-2.47
Kazakhstan	-7.01	-9.63	-6.27	-9.49	-3.49	-7.01	-4.09	-10.21
Libya	0.61	0.23	-0.79	1.18	-1.73	-0.91	-1.57	-0.97
Mali	-1.30	1.75	1.20	-3.33	5.96	8.94	11.28	22.00
Mauritania	2.13	3.12	4.61	7.55	5.30	7.30	6.87	1.48
Mexico	8.31	12.51	7.27	24.73	8.53	18.28	16.02	38.15
Mongolia	-14.66	-16.69	-16.47	-18.85	5.80	8.07	4.72	3.64
Namibia	10.63	14.75	8.44	15.37	4.85	8.54	6.29	14.99
Niger	7.52	10.29	9.22	10.63	-5.51	-3.71	-4.55	-4.70
Pakistan	1.32	0.28	-0.06	2.56	-1.25	-0.69	-0.83	-1.08
Russia	-15.57	-21.50	-13.97	-23.93	-11.35	-14.64	-11.05	-20.87
Saudi Arabia	-3.22	-5.81	-4.10	-10.96	4.48	3.13	0.92	3.79
Somalia	-3.81	0.64	-0.98	10.04	-6.69	-16.02	-9.46	-20.77
South Africa	8.88	18.62	10.71	24.53	6.73	14.10	10.08	24.38
Sudan	-4.46	-5.77	-6.13	-8.18	0.01	1.80	-1.63	0.85
Turkey	-10.34	-7.68	-17.97	0.21	-3.98	-10.04	-6.91	-8.31
United States	1.09	0.73	5.03	6.38	-9.74	-11.17	-8.39	-12.67

Table A3-14: The optimal values of hyperparameters and goodness-of-fit quantified by accuracy metrics for the final 16 best fitted models.

Model name	Number of learning cycles	Learn rate	Minimum leaf size	Maximum number of splits	Number of variables to sample	*MSE*[a]	*RMSE*[b]	*MAE*[c]	*NSE*[d]	R^2	Minimum Objective function	Ensemble aggregation method
CMIP5 models												
GISS-E2-H	210	0.090	14	487	2	13.30	3.65	1.60	0.72	0.72	2.659	LSBoost
GISS-E2-R	79	0.101	20	21,996	3	13.47	3.67	1.57	0.72	0.72	2.712	LSBoost
MIROC5-ensemble	69	0.137	8	35,913	2	13.11	3.62	1.51	0.72	0.72	2.737	LSBoost
MIROC-ESM-CHEM	337	0.039	6	1,824	1	12.68	3.56	1.45	0.73	0.73	2.643	LSBoost
MIROC-ESM	38	0.146	16	640	6	12.61	3.55	1.41	0.73	0.73	2.766	LSBoost
MRI-CGCM3	73	0.107	9	8,266	3	12.83	3.58	1.47	0.73	0.73	2.652	LSBoost
MRI-ESM1	43	0.135	3	2,290	7	13.39	3.66	1.54	0.72	0.72	2.736	LSBoost
NorESM1-M	66	0.110	3	3,375	3	13.05	3.61	1.45	0.73	0.73	2.729	LSBoost
CMIP6 models												
CESM2-WACCM-ensemble	210	0.090	14	487	2	13.30	3.65	1.60	0.72	0.72	2.659	LSBoost
CNRM-ESM2-1	79	0.101	20	21,996	3	13.47	3.67	1.57	0.72	0.72	2.712	LSBoost
GFDL-ESM4	69	0.137	8	35,913	2	13.11	3.62	1.51	0.72	0.72	2.737	LSBoost
INM-CM4-8-	337	0.039	6	1,824	1	12.68	3.56	1.45	0.73	0.73	2.643	LSBoost
INM-CM5-0	38	0.146	16	640	6	12.61	3.55	1.41	0.73	0.73	2.766	LSBoost
MIROC-ES2L	73	0.107	9	8,266	3	12.83	3.58	1.47	0.73	0.73	2.652	LSBoost
MRI-ESM2-0	43	0.135	3	2,290	7	13.39	3.66	1.54	0.72	0.72	2.736	LSBoost
NorESM2-LM	66	0.110	3	3,375	3	13.05	3.61	1.45	0.73	0.73	2.729	LSBoost

[a] Mean Squared Error.

[b] Root Mean Squared Error.

[c] Mean Absolute Error.

[d] Nash–Sutcliffe Model Efficiency Coefficient.

Table A3-15: Lower and upper limits of 95% confidence intervals of the mean for 10-fold cross-validation accuracy metrics calculated for trained models. We repeated the model trainings and hyperparameter tuning jobs for each of 16 datasets for 30 times and calculated confidence intervals of the mean based on 1,000 bootstrapped samples (with replacement) derived from the 30 validation metrics.

Model name	***MSE***		***RMSE***		***MAE***		***NSE***		R^2		**Minimum Objective function**	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
CMIP5 models												
GISS-E2-H	14.331	15.420	3.784	3.925	1.612	0.698	0.675	0.698	0.676	0.699	2.766	2.820
GISS-E2-R	14.331	15.420	3.784	3.925	1.612	0.698	0.675	0.698	0.676	0.699	2.766	2.820
MIROC5-ensemble	13.874	14.595	3.724	3.819	1.572	0.708	0.693	0.708	0.694	0.709	2.728	2.795
MIROC-ESM-CHEM	13.776	15.018	3.708	3.871	1.573	0.710	0.684	0.710	0.685	0.710	2.708	2.781
MIROC-ESM	13.300	14.194	3.647	3.766	1.518	0.720	0.701	0.720	0.702	0.720	2.667	2.755
MRI-CGCM3	13.968	15.233	3.735	3.900	1.583	0.706	0.679	0.706	0.681	0.707	2.719	2.792
MRI-ESM1	14.221	15.307	3.770	3.910	1.552	0.701	0.678	0.701	0.681	0.702	2.728	2.794
NorESM1-M	14.363	15.592	3.786	3.944	1.576	0.698	0.672	0.698	0.674	0.699	2.768	2.826
CMIP6 models												
CESM2-WACCM-ensemble	13.402	14.584	3.658	3.816	1.544	0.718	0.693	0.718	0.694	0.718	2.687	2.776
CNRM-ESM2-1	14.174	15.675	3.758	3.955	1.591	0.702	0.670	0.702	0.673	0.703	2.738	2.805
GFDL-ESM4	14.127	15.361	3.757	3.918	1.556	0.703	0.677	0.703	0.681	0.704	2.741	2.818
INM-CM4-8-	14.179	15.327	3.763	3.912	1.574	0.701	0.677	0.701	0.679	0.703	2.738	2.817
INM-CM5-0	13.753	14.842	3.706	3.849	1.586	0.710	0.687	0.710	0.688	0.711	2.745	2.781
MIROC-ES2L	14.264	15.267	3.774	3.904	1.572	0.700	0.679	0.700	0.681	0.701	2.743	2.809
MRI-ESM2-0	14.300	15.263	3.778	3.905	1.590	0.699	0.679	0.699	0.680	0.700	2.735	2.802
NorESM2-LM	14.262	15.147	3.775	3.891	1.623	0.700	0.681	0.700	0.685	0.701	2.709	2.810
Average	14.020	15.122	3.742	3.886	1.575	0.705	0.682	0.705	0.684	0.706	2.728	2.799

Table A3-16: Coefficients of determination (R^2) between measured and predicted values of soil salinity (EC_e) by 16 best fitted models at different soil depth intervals.

Model name	Depth below the surface (cm)					
	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100	100 - 200
CMIP5 models						
GISS-E2-H	0.62	0.73	0.76	0.76	0.78	0.71
GISS-E2-R	0.62	0.73	0.76	0.76	0.78	0.71
MIROC5-ensemble	0.64	0.73	0.76	0.78	0.80	0.72
MIROC-ESM-CHEM	0.65	0.74	0.79	0.77	0.81	0.73
MIROC-ESM	0.64	0.74	0.80	0.79	0.82	0.73
MRI-CGCM3	0.64	0.74	0.78	0.77	0.79	0.73
MRI-ESM1	0.63	0.72	0.77	0.77	0.80	0.72
NorESM1-M	0.62	0.72	0.75	0.77	0.80	0.73
CMIP6 models						
CESM2-WACCM-ensemble	0.64	0.73	0.77	0.78	0.80	0.73
CNRM-ESM2-1	0.66	0.74	0.78	0.78	0.80	0.73
GFDL-ESM4	0.63	0.73	0.79	0.77	0.79	0.73
INM-CM4-8-	0.64	0.72	0.77	0.78	0.81	0.74
INM-CM5-0	0.65	0.74	0.77	0.76	0.79	0.72
MIROC-ES2L	0.62	0.72	0.76	0.77	0.77	0.71
MRI-ESM2-0	0.64	0.72	0.78	0.78	0.80	0.73
NorESM2-LM	0.63	0.74	0.78	0.78	0.79	0.72
Average	0.64	0.73	0.77	0.77	0.80	0.73

Computer codes

This section makes the scripts and codes, required to regenerate the results, available to readers. Please note ArcGIS Desktop 10.x license is needed to run ArcPy module. Also, MATLAB Parallel Computing plus Statistics and Machine Learning toolboxes are required for running the MATLAB codes provided here.

Purely spatial predictors were directly pre-processed (including projections and resampling) in ArcGIS for Desktop environment (herein we refer to its central application: ArcMap). Soil texture raster datasets of clay content at different depths were averaged using ArcMap “raster calculator” tool. We extracted the values of predictors at the training input profiles locations by ArcMap “Extract Multi Values to Points” tool and saved the results as a table in .text format. The netCDF files of spatio-temporal predictors were processed by Climate Data Operators software and converted to multi-band rasters by ArcMap “Make NetCDF Raster Layer” tool. We extracted the values of these predictors at the locations of input profiles data too and saved them as tables in .text format. At each location there were 201 values, representing the years between 1900 and 2100. These data were finally merged, five-year moving averages of the spatio-temporal predictors were calculated and we attributed the moving averages of values of the spatio-temporal predictors at the locations to the input profiles data based on the year of acquisition of the sample profile using the following code:

```
clc;
clear;

%% This code merges the purely spatial and spatio-temporal predictors for each of the 16 input
datasets and prepare them for model training

% Importing the static predictors input data table
Soil_pro = readtable('D:\Projection_ML\Training_data_spatial\...
Training_data_static_predictors.txt','FileType','text','Delimiter','comma');

Model_name = {'Model_name_1'...
    'Model_name_2'...
    'Model_name_n'};


for ii = 1:length(Model_name) % Repeating the process for all 16 models

    % Importing input spatio-temporal data tables computed based on the
    % output of each of the 29 GCMs
    dryss_table = readtable(strcat('D:\Projection_ML\Training_data_spatio-...
    temporal\',Model_name{ii},...
    '\dryss_',Model_name{ii},'.txt'),'FileType','text','Delimiter','comma');
    wetss_table = readtable(strcat('D:\Projection_ML\Training_data_spatio-...
    temporal\',Model_name{ii},'\wetss_',Model_name{ii},'.txt'), ...
    'FileType','text','Delimiter','comma');
    evspsbl_table = readtable(strcat('D:\Projection_ML\Training_data_spatio-  ...
    temporal\',Model_name{ii},'\evspsbl_',Model_name{ii},'.txt'), ...
   'FileType','text','Delimiter','comma');
    pr_mean_table = readtable(strcat('D:\Projection_ML\Training_data_spatio-...
    temporal\',Model_name{ii},'\pr_mean_',Model_name{ii},'.txt'), ...
    'FileType','text','Delimiter','comma');
    pr_fre_table = readtable(strcat('D:\Projection_ML\Training_data_spatio- ...
    temporal\',Model_name{ii},'\pr_fre_',Model_name{ii},'.txt'), ...
    'FileType','text','Delimiter','comma');


    % Joining the tables
    T_dryss = ...
    join(Soil_pro,dryss_table,'LeftKeys','profile_id','RightKeys','Base_points_summarised');
    T_wetss = ...
    join(Soil_pro,wetss_table,'LeftKeys','profile_id','RightKeys','Base_points_summarised');
    T_evspsbl = ...
    join(Soil_pro,evspsbl_table,'LeftKeys','profile_id','RightKeys','Base_points_summarised');
    T_pr_mean = ...
```

```
    join(Soil_pro,pr_mean_table,'LeftKeys','profile_id','RightKeys','Base_points_summarised');
    T_pr_fre =  ...
    join(Soil_pro,pr_fre_table,'LeftKeys','profile_id','RightKeys','Base_points_summarised');

    % Calculation of the five-year moving averages
    Year_matrix = table2array(Soil_pro(:,7));
    dryss_matrix = table2array(T_dryss(:,22:222));
    dryss_matrix = movmean(dryss_matrix,[4 0],2,'omitnan','Endpoints','discard');
    wetss_matrix = table2array(T_wetss(:,22:222));
    wetss_matrix = movmean(wetss_matrix,[4 0],2,'omitnan','Endpoints','discard');
    evspsbl_matrix = table2array(T_evspsbl(:,22:222));
    evspsbl_matrix = movmean(evspsbl_matrix,[4 0],2,'omitnan','Endpoints','discard');
    pr_mean_matrix = table2array(T_pr_mean(:,22:222));
    pr_mean_matrix = movmean(pr_mean_matrix,[4 0],2,'omitnan','Endpoints','discard');
    pr_fre_matrix = table2array(T_pr_fre(:,22:222));
    pr_fre_matrix = movmean(pr_fre_matrix,[4 0],2,'omitnan','Endpoints','discard');

    dryss = zeros(size(Year_matrix,1),1);
    wetss = zeros(size(Year_matrix,1),1);
    evspsbl = zeros(size(Year_matrix,1),1);
    pr_mean = zeros(size(Year_matrix,1),1);
    pr_fre = zeros(size(Year_matrix,1),1);


    % Attributing spatio-temporal data to location of the points based on
    % the year of sampling the profile
    for i = 1:size(Year_matrix,1)
        dryss(i,1) = dryss_matrix(i,Year_matrix(i,1)-1903);
        wetss(i,1) = wetss_matrix(i,Year_matrix(i,1)-1903);
        evspsbl(i,1) = evspsbl_matrix(i,Year_matrix(i,1)-1903);
        pr_mean(i,1) = pr_mean_matrix(i,Year_matrix(i,1)-1903);
        pr_fre(i,1) = pr_fre_matrix(i,Year_matrix(i,1)-1903);
    end

    input_matrix = [Soil_pro.profile_id Soil_pro.profile_la Soil_pro.POINT_X Soil_pro.POINT_Y
...
    Soil_pro.Year...
        Soil_pro.upper_dept Soil_pro.lower_dept Soil_pro.WRB Soil_pro.Clay...
        Soil_pro.Elev Soil_pro.Slope Soil_pro.Field_capa Soil_pro.Wilt_point
        Soil_pro.Root_depth...
        dryss wetss evspsbl pr_fre pr_mean Soil_pro.elcosp_v_1];
    Table = array2table(input_matrix,'VariableNames',{'profile_id' 'profile_la' 'X' 'Y'
'Year'...
        'upper_dept' 'lower_dept' 'WRB' 'Clay'...
        'Elevation' 'Slope' 'Field_capa' 'Wilt_point' 'Root_depth' 'dryss'...
        'wetss' 'evspsbl' 'pr_fre' 'pr_mean' 'ECe'});

    % Saving output in a table format on disk
    writetable(Table,strcat('D:\Projection_ML\Training_input\',Model_name{ii},'.txt'));

end
```

16 input tables were generated for model training. These tables were later imported to MATAB to train an ensemble of regression tree learners using the following code:

```
clc;
clear;

%% Fitting an ensemble of regression trees for the 16 input datasets
% This script returns the tuned fitrensemble hyperparameters for 30 iterations using
fitrensemble % function in order to calculate the confidence intervals using bootstrapping
technique.
% This is regression using ensemble of trees (fitrensemble) on ECe as
% a target variable


Model_name = {'Model_name_1'...
    'Model_name_2'...
    'Model_name_n'};

for ii = 1:length(Model_name) % Repeating the process for all 16 models
    % Importing the original datasets prepared for training
    % Preprocessing the original dataset
```

```
    ECe = ...
    readtable(strcat('D:\Projection_ML\Training_input\',Model_name{ii},'.txt'),'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);

    table = standardizeMissing(ECe,-9999); % Replacing the missing values with -9999
    table(sum(ismissing(table),2) > 0,:) = [];% dropping the rows with missing values
    X = table(:,6:20);
    X.WRB = categorical(X.WRB);% Categorizing the categorical variables in the training set

    % Training and hyperparameter tuning job for 30 times
    % We used holdout method (25% held out) with the maximum 100 objective
    % function evaluations to optimize the ensemble hyperparameters
    % "ens" is the object of the final trained model
    for i = 1:30
        ens = fitrensemble(X,'ECe',...
            'OptimizeHyperparameters','all',...
            'HyperparameterOptimizationOptions',struct('Holdout',0.25,'UseParallel',true,...
            'MaxObjectiveEvaluations',100,'Repartition',true,'ShowPlots',false,'Verbose',1));
        % Acquiring and saving hyperparameter tuning job results on disk


save(strcat('D:\Projection_ML\Trained_ensemble\',Model_name{ii},'\ens_',num2str(i)),'ens');
    end
end
```

The training process on each of the 16 input datasets was repeated for 30 times. Using the following code, we cross-validated the models, calculated the accuracy metrics, and built their 95% confidence intervals of the mean:

```
clc;
clear;


%% Calculation of the confidence intervals of the mean for 10-fold cross-validation accuracy
% metrics
% Accuracy metrics including mean squared error (mse), mean absolute error (mae),
% NSE, and coefficient of determination (R^2) are computed

Model_name = {'Model_name_1'...
    'Model_name_2'...
    'Model_name_n'};

for ii = 1:length(Model_name) % Repeating the process for all 16 models

    ECe = ...
    readtable(strcat('D:\Projection_ML\Training_input\',Model_name{ii},'.txt'),'FileType',...
        'text','Delimiter',',','PreserveVariableNames',true);

    table = standardizeMissing(ECe,-9999);
    table(sum(ismissing(table),2) > 0,:) = [];

    % % Pre-allocating memory to variables with increasing size in each iteration
    Num_learning_cycles = zeros(30,1); Learn_rate = zeros(30,1); Min_leaf_size = zeros(30,1);
    Max_num_splits = zeros(30,1); Num_variables_to_sample = zeros(30,1);
    mse = zeros(30,1); mae = zeros(30,1);NSE = zeros(30,1); MinObjective = zeros(30,1);
    R2= zeros(30,1); Method = cell(30,1);

    % % Validation and Acquiring accuracy metrics
        ytrue = table.ECe;
    % This loop cross-validates the fitted models using 10-fold cross-validation
    for i = 1:30
        % Loading the saved model objects
        ens = ...
        load(strcat('D:\Projection_ML\Trained_ensemble\',Model_name{ii},'\ens_',num2str(i)));
        % Saving tunned hyperparameters
        MinObjective(i,1) = ens.ens.HyperparameterOptimizationResults.MinObjective;
        Method{i,1} = ...
        able2array(ens.ens.HyperparameterOptimizationResults.XAtMinObjective(1,1));
        Num_learning_cycles(i,1) = ...
        able2array(ens.ens.HyperparameterOptimizationResults.XAtMinObjective(1,2));
        Learn_rate(i,1) = ...
        able2array(ens.ens.HyperparameterOptimizationResults.XAtMinObjective(1,3));
        Min_leaf_size(i,1) =  ...
        able2array(ens.ens.HyperparameterOptimizationResults.XAtMinObjective(1,4));
```

```
        Max_num_splits(i,1) = ...
        able2array(ens.ens.HyperparameterOptimizationResults.XAtMinObjective(1,5));
        Num_variables_to_sample(i,1) = ...
        able2array(ens.ens.HyperparameterOptimizationResults.XAtMinObjective(1,6));
        % Validation of the trained ensemble using 10-fold cross-validation
        cvens = crossval(ens.ens,'Kfold',10);
        yfit = kfoldPredict(cvens);
        mse(i,1) = mean((ytrue - yfit).^2);
        mae(i,1) = mean(abs(yfit - ytrue));
        NSE(i,1) = 1 - sum((ytrue - yfit).^2)/sum((ytrue - mean(ytrue)).^2);
        % Coefficent of determination
        R2(i,1) = (sum((ytrue-mean(ytrue)).*(yfit-mean(yfit)))...
            /(sqrt(sum((ytrue-mean(ytrue)).^2))*sqrt(sum((yfit-mean(yfit)).^2))))^2;
    end
    rmse = sqrt(mse);% Root mean square error


    % Exporting the output into a table
    Statistics = [Num_learning_cycles Learn_rate Min_leaf_size Max_num_splits ...
    Num_variables_to_sample...
        mse rmse mae NSE R_2 MinObjective];
    ststistics_bootci = [mse rmse mae NSE R2 MinObjective];
    Statistics_table = array2table(Statistics,'VariableNames',{'Num_learning_cycles’
'Learn_rate'  'Min_leaf_size'...
    'Max_num_splits' 'Num_variables_to_sample'...
    'mse' 'rmse' 'mae' 'NSE' 'R2' 'MinObjective'});
    % Saving the table on disk
    Statistics_table.Method = Method;
    writetable(Statistics_table,strcat('D:\Projection_ML\Trained_ensemble\',Model_name{ii},...
    '\Bootstrapping_input.txt'));

    % % Bootstrapping
    % Computing the 95% confidence intervals of the mean for the statistics calculated
    % in the  above loop
    % using 1,000 bootstrap iterations. “bootci” creates each bootstrap sample
    % by sampling with replacement from the rows of the data arguments and
    % computes the confidence interval by bias corrected and accelerated percentile method.

    opt = statset('UseParallel',false);
    ci = bootci(1000,{@nanmean,ststistics_bootci},'type','bca','Options',opt);
    ci = array2table(ci,'VariableNames',{'mse' 'rmse' 'mae' 'NSE' 'R_2' 'MinObjective'});
    % Exporting the output into a table
    writetable(ci,strcat('D:\Projection_ML\Trained_ensemble\',Model_name{ii},'\Reg_CI.txt'));
end
```

Similar to the procedure used for input training profiles data, we extracted the values of pure spatial and spatio-temporal predictors to the locations of the drylands that we needed to make predictions for. The best fitted models were chosen from the 30 models trained for each of the 16 input datasets and used for prediction of the soil salinity (EC_e) at different depths. Different predictions were then averaged using trapezoidal rule to the depth of one meter from the surface and saved in .text tables for further analysis and calculation of the soils with salinity of $EC_e \geq 2$ dS m^{-1}. For each location, x- y- coordinates and 197 prediction from 1904 to 2100 were made. The following script shows how we merged the tables of purely spatial and spatio-temporal predictors and made predictions for soil salinity:

```
clc;
clear;

% This code merges the pure spatial and spatio-temporal predictors'
% values and use them to predict the soil salinity (ECe) at new locations

% Importing the purely spatial predictors table
Soil_pro =
readtable('D:\Projection_ML\new_data\Base_points\Spatial_preditors.txt','FileType','text','Del
imiter','comma');
Soil_pro = standardizeMissing(Soil_pro,-9999);
Soil_pro = fillmissing(Soil_pro,'nearest');
Model_name = {'Model_name_1'...
    'Model_name_2'...
    'Model_name_n'};
```

```
for ii = 1:length(Model_name) % Repeating the process for all 16 models
    % Finding the trained ensemble with the lowest RMSE
    Table = readtable(strcat('D:\Projection_ML\Trained_ensemble\', ...
    Model_name{ii},'\Bootstrapping_input.txt'),'FileType', ...
   'text','Delimiter',',','PreserveVariableNames',true);
    Table.Num = (1:30)';
    Table = sortrows(Table,'rmse','ascend');

    % Loading the best model
    Regression =  load(strcat('D:\Projection_ML\Trained_ensemble\', ...
    Model_name{ii},'\ens_',num2str(table2array(Table(1,13)))));

    % Importing the tables of spatio-temporal predictors
    dryss_table = readtable(strcat('D:\Projection_ML\new_data\',Model_name{ii},...
    '\dryss_',Model_name{ii},'.txt'),'FileType','text','Delimiter','comma');
    wetss_table = readtable(strcat('D:\Projection_ML\new_data\',Model_name{ii},...
    '\wetss_',Model_name{ii},'.txt'),'FileType','text','Delimiter','comma');
    evspsbl_table = readtable(strcat('D:\Projection_ML\new_data\',Model_name{ii},...
    '\evspsbl_',Model_name{ii},'.txt'),'FileType','text','Delimiter','comma');
    pr_mean_table = readtable(strcat('D:\Projection_ML\new_data\',Model_name{ii},...
    '\pr_mean_',Model_name{ii},'.txt'),'FileType','text','Delimiter','comma');
    pr_fre_table = readtable(strcat('D:\Projection_ML\new_data\',Model_name{ii},...
    '\pr_fre_',Model_name{ii},'.txt'),'FileType','text','Delimiter','comma');

    % Merging the tables of purely spatial and spatio-temporal predictors
    T_dryss =
join(Soil_pro,dryss_table,'LeftKeys','pointid','RightKeys','Base_points_drylands');
    T_wetss =
join(Soil_pro,wetss_table,'LeftKeys','pointid','RightKeys','Base_points_drylands');
    T_evspsbl = ...
    join(Soil_pro,evspsbl_table,'LeftKeys','pointid','RightKeys','Base_points_drylands');
    T_pr_mean = ...
    join(Soil_pro,pr_mean_table,'LeftKeys','pointid','RightKeys','Base_points_drylands');
    T_pr_fre = ...
    join(Soil_pro,pr_fre_table,'LeftKeys','pointid','RightKeys','Base_points_drylands');



    WRB = table2array(Soil_pro(:,12)); % Categorizing the categorical variables in the
    % training  set
    upper_dept = repmat([0 10 30 60 100],length(WRB),1);
    Clay = table2array(Soil_pro(:,15));
    Elevation = table2array(Soil_pro(:,14));
    Slope = table2array(Soil_pro(:,13));
    Root_depth = table2array(Soil_pro(:,5));
    Wilt_point = table2array(Soil_pro(:,8));
    Field_capa = table2array(Soil_pro(:,7));

    % Attributing spatio-temporal data to location of the points based on
    % the year of sampling the profile
    dryss = fillmissing(table2array(T_dryss(:,19:219)),'nearest');
    dryss = movmean(dryss,[4 0],2,'omitnan','Endpoints','discard');
    wetss = fillmissing(table2array(T_wetss(:,19:219)),'nearest');
    wetss = movmean(wetss,[4 0],2,'omitnan','Endpoints','discard');
    evspsbl = fillmissing(table2array(T_evspsbl(:,19:219)),'nearest');
    evspsbl = movmean(evspsbl,[4 0],2,'omitnan','Endpoints','discard');
    pr_mean = fillmissing(table2array(T_pr_mean(:,19:219)),'nearest');
    pr_mean = movmean(pr_mean,[4 0],2,'omitnan','Endpoints','discard');
    pr_fre = fillmissing(table2array(T_pr_fre(:,19:219)),'nearest');
    pr_fre = movmean(pr_fre,[4 0],2,'omitnan','Endpoints','discard');

    % Predicting the soil salinity at the five depths: 0, 10, 30, 60, and 100 cm
    % and averaging the output by the trapezoidal rule
    Salinity = zeros(length(WRB),size(dryss,2));
    Y = zeros(length(WRB),5);
    for i = 1:size(dryss,2)
        for j = 1:5
            X = [upper_dept(:,j) upper_dept(:,j) WRB Clay Elevation...
                Slope Field_capa Wilt_point Root_depth...
                dryss(:,i) wetss(:,i) evspsbl(:,i) pr_fre(:,i)...
                pr_mean(:,i)];
            X = array2table(X,'VariableNames',{'upper_dept' 'lower_dept'...
                'WRB' 'Clay' 'Elevation' 'Slope' 'Field_capa'...
                'Wilt_point' 'Root_depth' 'dryss' 'wetss' 'evspsbl' 'pr_fre' 'pr_mean'});
            X.WRB = categorical(X.WRB);
```

```
            Y(:,j) = predict(Regression.ens,X);
        end
        Salinity(:,i) = trapz([0 10 30 60 100],Y,2)/100;
    end

    % Replacing the missing values with flag of the missing values (-9999)
    Salinity = fillmissing(Salinity,'constant',-9999);

    % Preparing a tables for saving the results in
    Years = cell(1,201); count = 1904;
    for i = 5:201
        Years{i} = num2str(count);
        count = count + 1;
    end

    % Saving the table of results on disk
    Years{1} = 'pointid'; Years{2} = 'grid_code'; Years{3} = 'X'; Years{4} = 'Y';
    grid = [Soil_pro.pointid Soil_pro.grid_code Soil_pro.POINT_X Soil_pro.POINT_Y Salinity];
    T_result = array2table(grid,'VariableNames',Years);
 writetable(T_result,strcat('D:\Projection_ML\Results\',Model_name{ii},'_future{ii}','.txt'));
end
```

The following code was used for calculation of the cell area in WGS 1984 spatial coordinates system.

```
Calculating m^2 area of a WGS 1984 square pixel
% Adapted from: https://gis.stackexchange.com/a/127327/2397
% Parameters:
% cell_size (float): Pixel size in the Geographic coordinates (WGS 1984)which is 0.5 here
% Returns: Area of square pixel of side length cell_size in m^2

f_up = deg2rad(T.Y + cell_size/2);
f_down = deg2rad(T.Y - cell_size/2);
zm_up = (1 - e*sin(f_up));
zp_up = (1 + e*sin(f_up));
area_up = pi * b^2 * (log(zp_up./zm_up)/(2*e) + sin(f_up)./(zp_up.*zm_up));
zm_down = (1 - e*sin(f_down));
zp_down = (1 + e*sin(f_down));
area_down = pi * b^2 * (log(zp_down./zm_down)/(2*e) + ...
sin(f_down)./(zp_down.*zm_down));
cell_area = cell_size/360.*(area_up - area_down);
```