Mapping land suitability for informal, small-scale irrigation development using spatial modelling and machine learning in the Upper East Region, Ghana

DOI:
10.1016/j.scitotenv.2021.149959

Document Version
Accepted author manuscript

Title Page

Mapping land suitability for informal, small-scale irrigation development using spatial modelling and machine learning in the Upper East Region, Ghana

Komlavi Akpoti[1, *], Thomas P. Higginbottom[2], Timothy Foster[2], Roshan Adhikari[3], Sander J. Zwart[4]

[1] Regional Centre for Energy and Environmental Sustainability (RCEES), University of Energy and Natural Resources (UENR), Sunyani, Ghana

[2] Department of Mechanical, Aerospace and Civil Engineering, University of Manchester, Manchester, UK.

[3] Global Development Institute, University of Manchester, Manchester, UK

[4] International Water Management Institute (IWMI), Accra, Ghana

* akpotikomlavi@yahoo.fr, +233551790614

Highlights (for review : 3 to 5 bullet points (maximum 85 characters including spaces per bullet point)

- Framework on the potential for farmer-led irrigation was developed and implemented
- Water availability and other biophysical indicators only partially predict potential
- Market accessibility and water infrastructure were found key predictors
- 179,584 ± 49,853 hectares are potentially suitable for small-scale irrigation in Ghana’s Upper East Region

ABSTRACT

Small-scale irrigation has gained momentum in recent years as one of the development priorities in Sub-Saharan Africa. However, farmer-led irrigation is often informal with little support from extension services and a paucity of data on land suitability for irrigation. To map the spatial explicit suitability for dry season small-scale irrigation, we developed a method using an ensemble of boosted regression trees, random forest, and maximum entropy machine learning models for the Upper East Region of Ghana. Both biophysical predictors including surface and groundwater availability, climate, topography and soil properties, and socio-economic predictors which represent demography and infrastructure development such as accessibility to cities and proximity to roads were considered. We assessed that 179,584 ± 49,853 ha is suitable for dry-season small-scale irrigation development when only biophysical variables are considered, and 158,470 ± 27,222 ha when socio-economic variables are included alongside the biophysical predictors, representing 77-89% of the current rainfed-croplands. Travel time to cities, accessibility to small reservoirs, exchangeable sodium percentage, surface runoff that can be potentially stored in reservoirs, population density, proximity to roads, and elevation percentile were the top predictors of small-scale irrigation suitability. These results suggested that the availability of water alone is not a sufficient indicator for area suitability for small-scale irrigation. This calls for strategic road infrastructure development and an improvement in the support to farmers for market accessibility. The suitability for small-scale irrigation should be put in the local context of market availability, demographic indicators, and infrastructure development.

KEYWORDS

Food security, small-scale farmers, farmer led-irrigation, semi-arid region, land suitability, ecological niche modelling

1 INTRODUCTION

Agriculture represents a key economic sector in Sub-Saharan Africa (SSA), with smallholders playing an important role in national and regional food security (see Lowder et al., 2016 for a comprehensive discussion on farmland size and smallholders). In Ghana, for example, small-scale farms – with average farm sizes between 0.5 and 2 hectares – produce 95% of the country's food crops (Mendes et al., 2014). While the exact definition of small-scale irrigation is disputed in the literature (Turner, 1994), it is considered in this paper as irrigation where individual farmers or small groups of farmers have more control over the source of water they use for irrigation and the type of technology they use for small-scale, market-oriented agriculture. In this context, other terminologies for small-scale irrigation are distributed irrigation, small private irrigation, smallholder irrigation, or farmer-led irrigation (Xie et al., 2021). As opposed to large-scale irrigation, small-scale irrigation requires lower investment costs, is easier to operate and maintain, requires very little in terms of enterprise and management capability, and has a potentially less negative environmental impact (Tafesse, 2003). Small-scale irrigation is also considered a key tool to transform agriculture and food systems in SSA (Ringler et al., 2020). Intensification and expansion of irrigation offer a pathway for improving agricultural productivity, helping to address challenges of rural poverty (Burney et al., 2013; de Bont et al., 2019), food insecurity and malnutrition (Balana et al., 2020), and poor health outcomes (Domènech, 2015) across SSA. Irrigation is a key pathway for smallholder farmers to build resilience towards climate change (Alemayehu & Bewket, 2017). In addition, the adoption of small-scale irrigation is also being driven by growing demand for food, including vegetables and fruits, as a consequence of increases in income and changing diets of the growing middle-income consumers in urban areas in SSA (Balana et al., 2020).

While irrigation expansion is considered as a tool for poverty alleviation and key policy priority for donors and governments in SSA, only 4 to 6% of agricultural land in SSA is equipped with irrigation infrastructure, compared to 37% in Asia (Wiggins & Lankford, 2019; Burney et al., 2013). However, there are considerable renewable freshwater resources available in the region, although there is an uneven distribution of these resources in SSA. For example, West Africa has an estimated total renewable water resource of $1,315 \times 10^9$ m^3/year (the majority of these resources are located in countries such as Nigeria, Liberia, Guinea, Sierra Leone, and Mali), of which only 2% is withdrawn for human purposes (Namara and Sally, 2014). Having acknowledged this, the potential for future expansion and intensification of irrigation in SSA appears high, especially in areas where the resources are available (Wiggins & Lankford, 2019). Thus, small-scale irrigation is recognized as a mechanism for increasing productivity and income in the rural areas of developing countries. However, this potential is often not realized, as many publicly funded systems are underperforming, run-down, and in serious need of maintenance and refurbishment (Namara and Sally, 2014). Many irrigation schemes' performance has been sub-optimal with generally disappointing returns to investments, particularly in the case of large public irrigation schemes (Namara and Sally, 2014). While there were limited donors investment in irrigation infrastructure in SSA during the 1980s and 1990s, a revitalization and increased interest was shown towards the mid-2000s (Wiggins & Lankford, 2019) as governments and donors seek to improve food security, enhance resilience against climatic shocks (Misra, 2014) and decouple agriculture from rainfall variability (Cooper & Coe, 2011). In June of 2002, the Comprehensive Africa Agriculture Development Programme (CAADP) of the New Partnership for Africa's Development (NEPAD) was endorsed by the African Ministers assembled at the FAO Regional Conference for Africa. The NEPAD and Food and Agriculture Organisation (FAO) of the United Nations initiated the CAADP, in which African countries pledge 10% of their national budget towards agriculture to spark an annual agricultural growth rate of 6% (NEPAD, 2003). The CAADP provides a framework for restoring agricultural growth, rural development, and food security in the African region. It has four key focus areas, so-called Pillars, for agricultural improvement and investment. Pillar 1 deals with land and water management, aiming at extending the area under sustainable land management and reliable water control systems (NEPAD, 2003). Building on CAADP goals and countries' commitments and as a response to declining levels of investment in agricultural water, the AfDB, FAO, IFAD, IWMI, and the World Bank came together in 2007 and jointly prepared a collaborative agricultural water strategy known as "Investment in agricultural water for poverty reduction and economic growth in sub-Saharan Africa" (FAO, 2008). In March 2008, during the first Africa Water Week, the African Minister's Council on Water (AMCOW) called on NEPAD to inaugurate a new partnership – Agricultural Water for Africa (AgWA) – that would re-engage African countries, donors, as well as regional and international organizations in the development of water resources for food production, economic growth and poverty reduction (FAO, 2008). AgWa's major roles are advocacy; highlighting messages such as water for food, water for wealth, and water for life, and mobilizing resources (Namara and Sally, 2014). It also shares knowledge to improve the availability of information and knowledge at regional and national levels among agricultural water management professionals among others (Namara and Sally, 2014).

Small-scale irrigation development is back into the development agenda for SSA. Due to past and current underperformance of large-scale irrigation projects, informal farmer-led irrigation is considered a sustainable pathway for irrigation expansion in SSA (Higginbottom et al., 2021). Research has shown that the economic return of profitable small-scale irrigation expansion is more than twice the return of large-scale, dam-based centralized schemes while at the same time, the suitability for small-scale irrigation is five times higher than large-scale irrigation in SSA (You et al., 2011). Small-scale irrigation is managed by individual farmers, households, or small groups of farmers that self-supply irrigation from different sources using a variety of technologies to produce high-value crops such as vegetables alongside traditional staple crops such as rice and maize. Indeed, small-scale, farmer-led irrigation development is strongly determined by “market” but not all dynamics are related to the most commons high-value crops also termed as horticulture crops. In the remote areas with difficult access, it might be highly profitable to produce staple food to sell locally during the hungry gap period (or just to avoid having to buy staple food whose value may triple at this time). This kind of niche market (a local variation of price due to (in)accessibility) might not be well captured by national statistics and can face significant annual variation (as it is related to rainy season production). In other word, there are also food security mechanisms involved. At last, what will matter will be both the access to market demand but also the offer which depends on the existence of a non-rainfed cropping system that may impact local production (e.g., flood recession cropping systems). Much of the growth in smallholder irrigation is informal and not constructed or operated through the intervention of a government or donor agency (Drechsel et al., 2006). As such, these farmer-led irrigation systems are often poorly represented in official government statistics on the irrigated agricultural area as well as policies that target an increase in agricultural production and irrigation development. The lack of data on where irrigation suitability currently exists and related data on trends, opportunities, and constraints in informal irrigations is a major limitation for donors' and policymakers' interventions (Namara et al., 2011). As such, mapping the suitability for farmer-led irrigation is important to help identify the right mix of interventions for the planning of future irrigation development initiatives (Namara et al., 2010).

Many recent studies have sought to map the suitability for smallholder irrigation development in SSA. For example, You et al., (2011) combined biophysical and socio-economic assessment to estimate the suitability for irrigation while Xie et al., (2014) developed a framework for estimating the suitability for small-scale irrigation for the whole of SSA. These studies used a combination of environmental suitability and rural demographic analysis, hydrologic and crop simulation, crop prices, and cost-benefit analysis to account for the suitability. A similar method was applied to estimate the suitability for the expansion of small-scale irrigation in Nigeria (Xie et al., 2017). Schmitter et al., (2018) developed a framework for mapping the geo-spatial suitability of solar-based PV pumping for irrigation by combining solar radiation and availability of water resources and linkage to markets using weighted overlay multi-criteria evaluation in Ethiopia. Gumma et al., (2011) used fuzzy methods and irrigation statistics to map Ghana’s irrigated areas while a GIS Multi-Criteria Evaluation (MCE) was used to evaluate the suitability of land for surface and shallow groundwater irrigation in Ghana (Worqlul et al., 2019). The majority of these methods are based primarily on multicriteria evaluations, which identify suitable areas for irrigation expansion based on expert opinion about the biophysical and socio-economic suitability of different areas for irrigation development. Since the development of the framework for land evaluation (land suitability/land capability) by the FAO in 1976, many methods have since been used to assess agricultural land suitability including land suitability for irrigation. One of these methods is called expert knowledge or judgment. In this context, an experienced rice farmer for example can be considered an expert. In land suitability assessment/modelling, many factors are used, generally refers to multi-criteria evaluation (MCE). While the choice of factors/predictors in the MCE by experts are mostly accurate on a field scale, the same cannot be said for landscape-scale where complexities, interactions among factors are often overlooked by expert judgment. This often introduces bias on which factors should be prioritized in the decision process, especially when large numbers of predictors are considered (Akpoti et al., 2019) and where complex systems such as the heterogenous landscape of small-scale cropland are modeled. However, in the recent improvement in computation and algorithms including machine learning methods, these complexities and variations are taken into account, providing added value in the analysis[1]. To better understand the suitability for small-scale irrigation, we propose a methodology that combines advanced statistics through machine learning and spatial modelling. The approach was developed to accurately map the suitability for small-scale irrigation by setting models that yield complex response surfaces/predictions (Zurell et al., 2020) based on a large enough sample size of cropland for calibration. The approach also produces statistically ranked predictors based on survey data and their level of influence in the system. Besides, the approach produces response shapes that summarise the relationship between estimated suitable small-scale cropland and the biophysical and socio-economic environments which are then subjected to plausibility checks against available knowledge (Zurell et al., 2020) on cropland suitability for small-scale irrigation. Our approach, which is based on ecological niche modelling (Elith & Leathwick, 2009; Phillips et al, 2008), has been used to model agricultural land suitability (Akpoti et al., 2020; Akpoti et al., 2021; Heumann et al., 2011). Ecological niche modelling

[1] Although complexities and interactions among predictors are taken into account by the machine learning models, feedback loops which are characteristics of complex systems such small-scale irrigations systems are not clearly represented. This is because the machine learning models as adopted in this paper are not set up to interactively and dynamically take into account farmers behavioural change (due to endogenous and or exogenous factors) towards small-scale irrigations systems.

also referred to as correlative approach, associates the current and known geographical occurrences of species with environmental geographical data to generate a suitability gradient that is projected in geographical space (Peterson, 2006). Thus, the method links spatial predictors to estimate agricultural landscape suitability using known crops/croplands' geographical locations (Akpoti et al., 2021).

Irrigated area statistic data is not readily available in Ghana. There is an estimated 33,800 ha of irrigated land against 6.9 million hectares of cultivable land in Ghana (Namara et al., 2011), with irrigated agriculture representing less than 0.5% of the total cultivable area. Gumma et al., (2011) reported an irrigated area derived from remote sensing data of 32,421 ha with a conclusion that the estimated area was 20–57% higher than irrigated areas reported by Ghana's Irrigation Development Authority (GIDA). Also, of the gross estimated 1.9 million ha potentially irrigable area in the country based on FAO AQUASTAT, less than 2% has been developed (Mendes et al., 2014). Recent research by Worqlul et al., (2019) reported that approximately 9% of the area of Ghana was suitable for surface irrigation. About 186,000 ha is irrigated with water lifting technology (Namara et al., 2013) of an estimated 1 million ha of suitable land for small-scale irrigation (Xie et al., 2014). In this paper, we develop a spatial modelling framework for mapping the suitability for informal irrigation development to support goals of sustainable and efficient expansion of smallholder irrigation in SSA as a tool for poverty alleviation and economic development (Burney et al., 2013). We apply the framework to a case study in the semi-arid Upper East Region of northern Ghana, where water insecurity is a key driver of low agricultural productivity and food insecurity (Al-hassan, 2015; Dittoh et al., 2013). One of Ghana's irrigation policy directions is to enhance irrigated agriculture productivity with recognition of small-scale irrigation as one of the principal categories of irrigation in the country (Ghana Irrigation Development Authority, 2011). This policy also recognizes the lack of data on the overall extent of informal irrigation, which limits the support to small-scale farmers by the extension services. For successful programs and interventions, quantified data is key for making informed decisions about where to target investments to support the expansion and intensification of irrigation. This study was designed to identify some of the key factors driving the suitability of land for irrigation and their interactions at the landscape level as opposed to a single farm or field-scale intervention. Therefore, our analysis addresses a critical data gap to inform policy in the region around the location of existing irrigated agriculture and feasible areas for future development, drawing on an extensive set of geo-located ground truth data collected through surveys in 2020 in the region. The scope of our study is limited to small-scale, farmer-led dry season irrigation as opposed to rainy season irrigation. In contrast to supplementary rainy season irrigation which is mostly used to grow staple crops, dry season irrigation is mainly applied to cash crops in a profit oriented agriculture in limited rainfall environments where production would otherwise not be feasible (Xie et al., 2021). Dry season irrigation access is a precondition for cultivation in many semi-arid and arid regions such as West Africa, and has been shown to result in significant improvements in farmer incomes and food security – hence identifying potential areas for dry season irrigation expansion is a key need for policy alongside efforts to increase use of irrigation as a supplemental buffer during the rainy season.

The remainder of the paper is structured as follows: First, we assessed the predictors that define land suitability for farmer-led small-scale irrigation. These parameters are related to climatic variables, agricultural water productivity and soil water content, soil chemical properties, soil physical properties, vegetation cover, and socio-economic variables. Secondly, we assessed the partial response surface of the predictors to the suitability level for small-scale irrigation; then we tested a series of model specifications for predicting small-scale, farmer-led irrigation suitability. Finally, we developed an ensemble of spatial probabilistic and binary predictions of suitability for small-scale irrigation. The results from this study are to support data-driven solutions for the realization of the promise of small-scale irrigation, especially in semi-arid regions.

2 METHODS AND MATERIALS

2.1 Study area

The study was conducted in the Upper East Region of Ghana located in the northeast of Ghana in West Africa between longitudes 0° and 1° 60'W and latitudes 10° 33'N and 11° 17' (Figure 1) with a geographic area of 8,842 km^2. The region is in the Sudan Savannah zone with annual rainfall between 645 mm and 1250 mm. The region has a unimodal rainfall pattern, with a summer (May to September) monsoon rainy season followed by a long dry season between October and April, and experiences frequent droughts (Dietz et al., 2004). The soil in the region has low fertility, low content of organic matter, and mostly coarse-textured, with high susceptibility to soil erosion due to shallow surface soil profiles (Amegashie et al., 2012).

The majority of the population in the area is rural, with agricultural production as the major source of livelihood. Crops such as millet, sorghum, groundnut are grown in the rainy season, while leafy vegetables, okra, onions, peppers, and tomatoes are prioritized in the dry season where irrigation is available (See Figure 1). Irrigation systems and water sources in the region are heterogenous, including shallow groundwater pumping, small surface water reservoirs, and larger government-managed irrigation dams (Annor et al., 2009). The latter comprises two major irrigation schemes: the 850 ha Vea scheme near Bolgatanga and the 2,490 ha Tono scheme near Navrongo. Note that there is little cropping in the dry season in areas without irrigation due to low rainfall and high crop water requirements during this period of the year in the Upper East Region.

2.2 Methodology

In the following sections 2.2.1 and 2.2.2, we described the plot level data and candidate predictors while in section 2.2.3, we presented an overview of the small-scale irrigation suitability mapping. We discussed the methodology for the machine learning models' parameterizations for small-scale irrigation mapping, as well as the statistical evaluations of the predictions under section 2.2.4. We reported the methodology used to compare our predictions with existing land use data and the model restriction approach in sections 2.2.4 and 2.2.5 respectively.

2.2.1 Survey protocol of small-scale, informal irrigation plots

To train predictive machine learning models of suitable areas for small-scale informal irrigation mapping, baseline data on existing informal irrigated areas in the study region were required. As part of the FutureDAMS (Design and Assessment of water-energy-food-environment Mega-Systems), the International Water Management Institute (IWMI) and the University of Manchester (UoM) organized a survey between May and July 2020 to collect data on informal irrigation activities to serve as data inputs for the development of decision support system that could foster improvement in informal irrigation activities. The survey was carried by enumerators drawn from the various districts and with adequate knowledge of the agricultural and socio-demographic context of the communities. The data on small-scale irrigation were collected through interviews and GPS tracking of irrigation plots. The World Bank's Survey Solution software was used to design the questionnaires, and the survey was implemented using GSM-enabled mobile tablets. The data was collected through surveys of approximately 1,200 households in the 15 districts across the Upper East Region. For each household, information was collected on whether the farmer cultivated the land during the prior (2019-2020) dry season (a total of 707 out of 1,200 surveyed households reported cultivation). For each of these 707 farmers, information was collected on the location of their main agricultural plot, the position of this plot in the landscape (uplands, inland valleys, fringes, floodplains), physical properties of the soil (clay, sand, etc.), source of water for irrigation (groundwater, small reservoirs, rivers), irrigation equipment, water lifting methods, crop types cultivated, seasonal irrigation practices, irrigation constraints, and available investment opportunities. For each of the 707 plots, the boundary of the plot was recorded using a GPS device based on which plot centroids were computed (see Figure 1).

The categories of plot owners were either the household heads or the spouses, children, grandchild, parents' in-law, sons/daughters' in-law, or another relative of the household head. In terms of positions within the topography, the surveyed plots were distributed among fringes (5%), inland valleys (16%), lowlands/floodplains (30%), and uplands (49%), and within the sand, clay, and silt soil textures. Sources for irrigation were groundwater (60%), reservoirs (21%), and rivers/streams (19%). Irrigation equipment in the various farms was bucket/watering can, diversion/gate, electric pump, hand pump, petrol/kerosene/diesel pump, solar-powered pump, plastic/lay flat pipe/water hose pipe, canal/ditch, drip, furrows, sprinkler/rain hose. In terms of investments in irrigation, most farmers invest in irrigation infrastructure for their usage (93% of the surveyed plots). Others participate in a collaborative investment in irrigation infrastructure for joint usage (4.8%) whiles the remaining 2.2% benefited from a project investment in irrigation development (but not from the government).

Spatial location data collection often results in spatial clustering with a bias toward easily accessible locations such as roads and towns (Reddy & Dávalos, 2003). These biases enhance location spatial auto-correlation (Jane Elith & Leathwick, 2009) which can result in model overfitting to the predictors (Phillips et al., 2009). To reduce sampling bias and increase independence among field observations in the geographic and environmental spaces due to clustering points (Boria et al., 2014), we applied spatial filtering of 2 km between irrigated plots, which randomly sampled the irrigated plots to be retained for the modelling. We used the SDMtoolbox for the spatial filtering process (Brown et al., 2017), resulting in a final record of records of 426 irrigated plots.

2.2.2 Predictors selection

To evaluate potential drivers of the location of small-scale irrigation development, we reviewed previous literature (You et al., 2011; Xie et al., 2017; Nakawuka et al., 2018; Schmitter et al., 2018; de Bont et al., 2019) to define 30 candidate predictors (see Table 1 and Figure 2 for the list of predictors). These predictors were categorized under socio-economic variables, water availability, climate and energy, topographical indicators, soil physical and chemical properties. Spatial auto-correlation among predictors was previously shown as a major source of uncertainties in spatial predictions of species distribution (Braunisch et al., 2013). Pearson correlation was therefore used to access the collinearity between any two pairs predictors. The correlation coefficients were analyzed using the 'corrplot' package in the R programming language (Wei et al., 2017) (see Figure 3). We used a threshold of $|r| < 0.75$ to exclude one of the paired spatially dependent predictors. To decide on which of the spatially dependent predictors to drop, the simulation was run to evaluate the relative influence of these predictors alternatively. Eventually, the predictor with the highest influence was retained for the subsequent modelling. Additionally, we used a stepwise elimination of the least contributing predictors. We did this by ranking the predictors' importance produced by the modelling algorithms and eliminate the lowest-ranked predictors one at a time. We rerun models several times until no important drop in the model's accuracies such as the area under the curves (AU ≥ 0.8) and percentage correctly classified (PCC ≥ 75%) were observed. Eventually, a total of 23 predictors were maintained in the final modelling (see Table 1).

2.2.2.1 Water availability predictors

Geographic distribution of irrigation is often linked to physical access to enough water, whether surface or groundwater (Wiggins & Lankford, 2019). The sources of water for irrigation in the study area are primarily dams and small reservoirs, groundwater, rivers, and streams (see Figure S1, Supplementary material). In the semi-arid regions at the margins of the Sahel, large numbers of small reservoirs capture surface runoff during the rainy season, storing water for use during the dry season. For the local population, small reservoirs are important water sources that help them cope with droughts (Annor et al., 2009). In this study, Euclidean Distance to small reservoirs was computed as a proxy for accessibility to reservoirs based on 384 small reservoirs'[2] geolocation data. The reservoir data was developed by the International Water Management Institute (IWMI) which used a binary Random Forest classification on Sentinel-2 images for the dry season (Ghansah & Zwart, 2020).

The ability of small and large reservoirs to effectively support small-scale irrigation is dependent on these storage systems being able to capture sufficient surface runoff availability after accounting for

[2] Small reservoirs are water storage facilities of hectares of water coverage and with generally shallow depth. In the case of the Upper East region, the surface areas of the reservoirs data used in this study varied between 0.09 and 37 ha.

losses of rainfall to evapotranspiration and groundwater recharge (Xie et al., 2014). This water storage is likely to occur during the rainy season, 3 to 4 months before the dry season irrigation takes place. The amount of runoff and its spatial and temporal variation is influenced by climate, vegetation, soil, and topology (You et al., 2011). To estimate the potential surface runoff availability that can be captured, we used the Soil and Water Assessment Tool (SWAT). The hydrological cycle simulated by the SWAT model is based on Equation (1) (Neitsch et al., 2011)

$$SW_t = SW_0 + \sum_{i=1}^{t} \left(R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw} \right) \quad (1)$$

where SW_t is the final soil water content (mm), SW_0 is the initial soil water content on a day *i* (mm), *t* is the time (days), R_{day} is the amount of precipitation on a day *i* (mm), Q_{surf} is the amount of surface runoff on a day *i* (mm), E_a is the amount of evapotranspiration on a day *i* (mm), W_{seep} is the amount of water entering the vadose zone from the soil profile on a day *i* (mm), and Q_{gw} is the amount of return flow on a day *i* (mm). We used the Soil Conservation Service (SCS) curve number method in SWAT to estimate the surface runoff volume. We used manual calibration to adjust for the surface runoff. We considered daily climate data on rainfall, minimum, and maximum temperature, wind speed, solar radiation from 1979-2014 of the Climate Forecast System Reanalysis (CFSR) global weather data for SWAT (https://globalweather.tamu.edu/, Fuka et al., 2014). We used land-use data from the European Space Agency (ESA) (http://2016africalandcover20m.esrin.esa.int/) and FAO digital soil map of the world (Nachtergaele et al., 2010) to set up the SWAT model. We considered a cumulative 5-years average surface runoff for July, August, September, and October.

Groundwater availability was represented in this study by two indicators: depth to groundwater table and aquifer productivity. The cost of extraction depends mainly on the depth of the water table (Amjath-Babu et al., 2016), while groundwater yield is a major determinant of suitability for groundwater-based irrigation (Foster et al., 2015). Geospatial estimates of depth to groundwater were obtained for the Upper East region from Fan et al., (2013). The aquifer productivity was computed using an inverse distance weighted (IDW) interpolation technique of the Geostatistical Analyst tool in ArcGIS on 2081 borehole yield data points extracted from the Ghana national borehole database.

Finally, in addition to previously mentioned predictors, we also considered one further hydrologic determinant of irrigation suitability – proximity to river networks. Proximity to rivers is another important predictor for irrigable land assessment from direct river abstractions (Assefa et al., 2018), and was computed using the Euclidean Distance tool in ArcGIS based on derived stream network data. The stream network data includes small streams, perennial and ephemeral river derived using the Digital Elevation Model (DEM) data. DEM data was obtained from Shuttle Radar Topography Mission (SRTM) 30m product (https://earthexplorer.usgs.gov/).

2.2.2.2 Climate and energy predictors

The amount of precipitation during the rainy season will influence surface runoff to be stored for dry season irrigation. We considered the precipitation of the wettest quarter, a quarterly index that approximates total precipitation that prevails during the wettest quarter from the WorldClim Version 2 database (Fick & Hijmans, 2017). Solar radiation is also important for both the solar water pumping and the water balance in the study area. We obtained solar radiation data from the WorldClim Version 2 database. WorldClim data is an average over the period 1970-2000. See Table 1 for the units and spatial resolution of climate and energy predictors.

2.2.2.3 Topographical predictors

The survey of the small-scale irrigation plots showed that the edges of the valleys or fringes, inland valleys, lowlands/flood plains, and uplands are the ecologies exploited in the Upper East Region (see Figure S3 in the supplementary material). To discriminate between these positions in the landscape and accurately represent them in the predictors, two morphometric indices that combine different derivatives of DEM were computed, namely the elevation percentile and the topographic wetness index (TWI). DEM was resampled from 30m to 60m and 90m resolution to capture the various sizes of inland valleys in the landscape. The three DEMs -30m, 60m, and 90m were then further smoothed using the *raster.gaussian.smooth* function of the Spatial Analysis and Modelling Utilities (*spatialEco*) package in R (Evans et al., 2020).

The elevation percentile was measured by a ranking of elevation concerning a circular surrounding area (Gallant & Dowling, 2003). Thus, the elevation percentile is robust in defining the local topography in the defined surrounding radius. The elevation percentile was computed based on a size 11 x 11 filter kernel in the x and y directions using the function *wbt_elev_percentile* of the *whitebox* package in R (Lindsay, 2016). The elevation percentile was computed for the 3 DEM sizes. The percentile products were further normalized using the fuzzy small membership function in ArcGIS. The elevation percentiles were finally combined using the fuzzy sum membership function.

The topographic wetness index is one of the key determinants of soil moisture spatial variability based on the assumption that in sloped terrain, topography controls the movement of water (Schmidt & Persson, 2003). We used the DEMs 30, 60, and 90m from which we derived the TWI. The TWI of each pixel in the study area is a function of the upslope area (A) per unit contour length and the local slope (tanB) as:

$$TWI = \ln\left(\frac{A}{\tan(B)}\right) \tag{2}$$

TWI products were computed using the function *wbt_wetness_index* of the whitebox package in R. The TWI were further normalized using fuzzy linear membership function in ArcGIS, then combined using fuzzy sum. Also, the slope was considered as irrigation is more likely to occur in an area with a gentle slope as it affects land preparation and irrigation efficiency (Xie et al., 2014).

2.2.2.4 Soil physical properties

Three soil physical properties obtained from Hengl et al., (2017) were used in the modelling, namely bulk density, available soil water capacity, and coarse fragment. Bulk density is related to the physical and chemical properties of the soil and it plays an important role in soil water retention (Al-shammary et al., 2018). Also, irrigation treatment has been linked to available soil water capacity (Panda et al., 2004) while Coarse fragment affects soil workability and crop growth (Obour et al., 2017). See Table 1 for the units and spatial resolution of the soil physical properties predictors.

2.2.2.5 Soil chemical properties

We considered cation exchange capacity, electrical conductivity, base saturation percentage, exchangeable sodium percentage, organic carbon, total nitrogen and phosphorus, soil pH, and exchangeable potassium as the chemical properties in the models. Cation exchange capacity influences the soil's capacity to hold onto essential nutrients (Juhos et al., 2019). Electrical conductivity is a measure of soil salinity and is one of the soil properties that influences crop productivity (Corwin & Lesch, 2003). The availability of soil nutrients increases with base saturation percentage (Havlin, 2005) while an excess of exchangeable sodium percentage harms the physical and nutritional properties of the soil, with the consequent reduction in crop growth (Yadav et al., 1988). The primary nutrients for crops in the models are represented by total nitrogen, exchangeable potassium, and total phosphorus which have an impact on crop productivity (Dadhich et al., 2017). Also, organic carbon maintains soil structure and forms the basis for the successful use of mineral fertilizers while soil pH is an important factor in soil productivity.

2.2.2.6 Socio-economic predictors

Proximity to paved roads is a proxy for accessibility to markets and agricultural inputs (Schmitter et al., 2018). This variable was computed using the Euclidean Distance tool in ArcGIS based on OpenStreetMap data. Also, the travel time required to access nearby populated areas indicates access to markets because the adoption of irrigation relies on market access for agricultural inputs and other equipment (Xie et al., 2014). The travel time data obtained from the accessibility database (Weiss et al., 2018). Accessibility is the travel time required to reach the nearest urban center via surface transport. Urban centers are defined as a contiguous area with 1,500 or more inhabitants per square kilometer or a majority of built-up land cover coincident with a population center of at least 50,000 inhabitants (Weiss et al., 2018). The data was developed with gridded surfaces that quantify the geographical positions and salient attributes of roads, railways, rivers, water bodies, land cover types, topographical conditions (slope angle and elevation), and national borders and were combined to create a global ‘friction surface’, effectively enumerating the generalized rates at which humans can move through each pixel of the world’s surface (Weiss et al., 2018). The adoption of small-scale irrigation is influenced by the high population density for agricultural inputs and the market for selling agricultural products (Worqlul et al., 2017). We also considered the Gridded Population of the World (GPW), v4 data (CIESIN, 2016). Proximity to town is also a proxy for market accessibility (Schmitter et al., 2018) and was computed using the Euclidean Distance tool in ArcGIS based on town location data.

2.2.3 Overview of the small-scale irrigation suitability mapping

We applied the Ecological Niche Models (ENMs) approach on cultivated crop/croplands to estimate the suitability for farmer-led irrigation. This approach was previously used to assess irrigated rice expansion in Burkina Faso (Akpoti et al, 2021). ENMs, also known as species distribution models (SDMs), are correlative models that link suitable areas of a species by inferring environmental conditions among predictors and predicting suitability over the study area (Hochman et al., 2013; Peterson, 2006). In this study, we implemented three machine learning algorithms —Boosted Regression Trees (BRT), Maximum Entropy (MaxEnt), and Random Forest (RF) to estimate the suitability for small-scale, informal irrigation. These algorithms were previously shown to accurately model both cropland extent and suitability (Akpoti et al., 2020; Müller et al., 2013; Singh et al., 2017). Informal farmer-led irrigation is considered a distributed irrigation system, meaning that access to water, distribution, and use occurs or near the same location (Burney et al., 2013). Therefore, the suitability of small-scale irrigation can be estimated at the plot level (Xie et al., 2021) and infer the suitability for on a spatial scale to other non-sampled areas. The application of the ENMs, which are trained on the biophysical and socio-economic characteristics of plots at the micro-level is therefore appropriate.

We considered two cases in the simulation of the suitability for small-scale irrigation. In the first case, hereafter referred to as case 1, only biophysical predictors of irrigation suitability were included. Biophysical predictors included the availability of water for irrigation, the optimal land scale conditions, the climatic conditions, soil physical and chemical properties (see Table 1 for more metadata on the predictors and Section 2.2.2 for predictors description). In the second case (case 2), we estimate the suitability for small-scale irrigation by considering additional socio-economic predictors alongside biophysical variables included in case 1. Variables included as proxies of socio-economic conditions that may influence suitability for irrigation development were population density, travel time to major cities, and accessibility to paved roads. Considering both biophysical and socio-economic determinants of irrigation is important as previous studies have shown that these variables are key determinates of agricultural development in SSA (Ghana Irrigation Development Authority, 2011; Kong et al., 2016; Rebelo et al., 2010), and thus should be considered when trying to identify suitability scope for targeted interventions to expand and intensify irrigation access for smallholders.

The modelling procedure draws on the earlier work by Akpoti et al., (2020, 2021) and is implemented in Software for Assisted Habitat Modelling (SAHM) (Morisette et al., 2013). The overall modelling framework is organized into five main steps (See Figure 4). Step 1 consisted of input data preparation. The input data included candidate predictors, field observation data of small-scale farmer-led irrigation plots for the various dry season crop species, and reference layer data. Step 2 consisted of data pre-processing within and outside the SAHM workflow. The pre-processing consisted of a random sampling of the field observation data with a minimum distance of 2km (spatial filtering) to avoid the model's overfitting to predictors data (See Section 2.2.1) and the projection, aggregation using mean method—, resampling — using the nearest neighbor method, and clipping to match the 30 m by 30 m grid cells of the template layer. Step 3) consisted of preliminary model analysis and decision. To discriminate between suitable and unsuitable areas for irrigation relative to the environmental and socio-economic predictors, the second level of information is required by the machine learning models. This information is obtained from a sample of points from the study region, also called background or pseudo-absence data. A total of 12,000 background samples was randomly generated using the Kernel density estimate method (Duong, 2015) and the module for merge dataset was used to extract the values of each predictor layer to the point locations included in the field data. The models were calibrated using 70% of the data called training data and validated using the remaining 30% of the data called testing data. Multi-collinearity is a redundancy assessment between predictors to access variables with similar distributions that might result in a biased or overfitted models due to lack of linear independence among the predictors. Multicollinearity makes it challenging to determine the effect of the individual predictors on the target variable and to identify the variables that should be included in the final model (Ohana-Levi et al., 2019; Sahour et al., 2020). The predictors' multicollinearity was assessed based on the cut-off threshold of 0.75. Step 4) consisted of correlative models (BRT, MaxEnt, RF) parametrizations, and step 5 focused on the output routines for models results post-processing and evaluation when final results are satisfactory (the “Yes” arrow). Otherwise, models’ adjustments are made under the “No” arrows.

2.2.4 Models’ parameterizations for small-scale irrigation mapping

Machine learning algorithms including Boosted regression trees (BRT), Random Forest (RF), Maximum Entropy (MAXENT) algorithms were used for mapping the suitability for small scale, informal irrigation. The specifications for each model are described below.

2.2.4.1 Boosted regression trees prediction

BRT algorithm, also known as stochastic gradient boosting (Elith et al., 2006) or boosted additive trees (Araújo & New, 2007) is a combination of decision trees and boosting. BRT uses decision trees to link predictors to the response by recursive binary splits using a logit link function while improving the prediction by iteratively taking into account the weak learners (Elith et al., 2008). The model implementation was based on 1,000 trees such that the model accounted for interactions and nonlinear relations among predictors (Morisette et al., 2013). We considered the 10-folds internal cross-validation technique for model simplification— model pruning. We used a bag fraction of 0.7 to control the proportion of the data that is used to fit the model at each step. A weak learning rate of 0.005 was used to control the amount each tree contributes to the model to avoid model overfitting (Elith et al. 2008) with a tree complexity of 3. Similar parameter specifications were previously used by Müller et al., (2013) in a determinant of cropland modelling.

2.2.4.2 Maximum entropy prediction

MaxEnt is a general-purpose machine learning method for prediction from incomplete information or presence-only data by finding the probability distribution of maximum entropy using the jackknife maximum likelihood estimator (Phillips et al., 2006). The model sets a constraint using the density estimation approach, such that the expected value of each predictor closely matches its empirical average. The default setting of the MaxEnt model tends to perform well due to the testing of the model on a large dataset (Elith & Graham, 2009). We considered the default settings in the exception of the background sample which was set to 12,000 and the maximum iteration which was set to 10,000. The feature selection and regularization are in relation with the observed data (Phillips et al., 2006), which in our case were 443 small-scale irrigation locations. Previous studies on irrigated and rainfed rice suitability estimation have used similar parameter setting for the MaxEnt model (Akpoti et al., 2020; Akpoti et al., 2021).

2.2.4.3 Random forest prediction

RF is an ensemble of decision trees algorithm where each tree is grown with a randomized subset of predictors (Breiman, 2011). The model, as implemented in the SAHM package is based on the R package ‘randomForest’ (Breiman et al., 2011). The training parameters specified in the modelling are the number of trees to grow in the forest which was set to 1000, the number of randomly selected predictor variables at each node was set to 7, and the minimal number of observations at the terminal nodes of the trees was put at 5.

2.2.4.4 Ensemble modelling

Multi-models’ applications in spatial prediction differ in their modelling performance and predictions, and some consensus methods are needed to produce the central tendency and reduce the uncertainty of predictions (Crimmins et al., 2013). We combined the three algorithms in an ensemble using a simple average as previously recommended by Marmion et al., (2009) based on the conditions that testing

Area Under the Curve (AUC) of each model is at least 0.8. We also considered a consensus approach to defining the number of algorithms that classified any given pixel as suitable.

2.2.4.5 Thresholds definition

The prediction of the 3 machine learning models produced a continuous probability map of suitability for small-scale irrigation with pixel values ranging from 0 to 1. Thresholds are needed for models' probability surfaces transformation into binary outputs and for evaluation metrics estimation. To convert these raster continuous surfaces into binary (suitable, unsuitable), we selected a threshold based on sensitivity equals specificity, i.e. equal chances of true positive and true negative rates. Sensitivity is the probability of actual small-scale irrigation areas predicted while specificity is the probability of actual background points predicted. This approach is considered as objective and preferred measure of discrimination power compared to many other methods (Jiménez-Valverde, 2014) because it gives an equal chance of error in classifying suitable and unsuitable areas.

2.2.4.6 Evaluation of model performance

For the evaluation of the model's performances, we considered four metrics including threshold-independent, graphical assessment methods such as Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) and threshold-dependent dependent metrics such as sensitivity, specificity, Percent Correctly Classified (PCC), and True Skill Statistics (PCC). The mathematical description of these metrics can be seen in Tharwat, (2020); and is widely used for spatial modelling evaluation (Jarnevich et al., 2017; Akpoti et al., 2020). According to Peterson et al., (2011), AUC values less than 0.5 shows that models perform worse than random; AUC values of 0.5 are models that are not better than random; 0.5–0.7 is an indication of models' poor performance; 0.7–0.9 as reasonable/moderate performance; and greater than 0.9 as high performance.

2.2.5 Comparison between predictions and existing land use land cover data

We validated the predictions of the current analysis by comparing the spatial distribution of the suitability for small-scale irrigation with two recent land cover products, namely the 20m CCI Land Cover product of 2016 (http://2016africalandcover20m.esrin.esa.int/) and the 100m WaPOR land cover data (https://wapor.apps.fao.org/catalog/WAPOR_2/2/L2_LCC_A) of 2019. We created a mask of the suitability estimate with the land cover products and reported the land area percentage of each land type.

2.2.6 Model restriction

We excluded forest reserves and protected areas from the analysis obtained from the World Database on Protected Areas (WDPA) (https://www.protectedplanet.net/country/GH). Other constraints such as settlements obtained from the Global Human Settlement Layer (GHSL) (Florczyk et al., 2019), tree cover areas, and open water bodies were also excluded with the data derived from ESA CCI Land Cover - Sentinel-2A for Africa (http://2016africalandcover20m.esrin.esa.int/).

3 RESULTS

3.1 Model performance measures

The evaluation of the models' performances in the prediction of areas suitable for small-scale irrigation varies from moderate to high performance (see Table 1 for summary statistics). When trained on 70% of the irrigation data and background samples, the threshold independent metric AUC varies from 0.88 to 0.95 with the highest score shown by BRT, followed by MaxEnt and RF. The threshold-dependent metrics such as models' accuracies (PPC) vary from 79% to 88% following the same models' performance arrangement as in the case of AUC. The discrimination between suitable and unsuitable areas for small-scale irrigation was performed based on the equal probability (Sensitivity = Specificity), with values varying from 0.81 to 0.88. The test evaluation of the models based on 30% of the data follow the same trend as in the case of training with slightly lower values; except in the case of RF where testing AUC is marginally higher than the training. For training and testing, the True Skill Statistics (TSS) values vary from 0.58 to 0.62. Although only minor differences are shown across the models, overall, BRT has the highest values in the evaluation metrics, followed by RF and MaxEnt.

3.2 Suitability distribution of small-scale irrigation

3.2.1 Case 1: Suitability distribution using biophysical predictors only

The binary (suitable vs unsuitable) surfaces for the suitability of small-scale irrigation as estimated by the three algorithms are shown in Figure 5a-c, where biophysical predictors only are considered in the modelling. All the models show similar prediction patterns across the study area, with the overall suitability (Figure 5d) depicts three main clusters that match the binary predictions as reported in Figure A1 in Appendix 1. The BRT, MaxEnt, and RF suitable area predictions are obtained for corresponding thresholds of 0.63, 0.29, and 0.52 on training data, respectively, where thresholds are obtained based on the equal probability of sensitivity and specificity. The first cluster of suitable areas extends from longitude 1°12'W to the western part of the study area. This cluster mainly corresponds to leafy vegetables, pepper, okra, and tomato crops in small-scale irrigation areas (see Figure 1). The second cluster extends from longitude 1°12'W to 0.5°W with pepper as the dominant irrigated crop followed by tomatoes, onions, and leafy vegetables. The third cluster extends from longitude 0.5°W eastward with onions as the dominant irrigated crop followed by pepper and okra. In the case where only biophysical predictors were considered to model suitability, the RF model shows the highest predicted area of 227,805 ha, followed by MaxEnt with an estimated area of 182,701 ha and BRT with a predicted area of 128,245 ha. The predictions of RF, MaxEnt, and BRT represent 26%, 21%, and 15% of the total area of the Upper East Region respectively.

3.2.2 Case 2: Suitability distribution using biophysical and socio-economic predictors

Similar patterns are shown in the spatial predictions when socio-economic predictors are included in the predictions. However, in this case, suitability for farmer-led irrigation is more pronounced to areas with easy access to roads, shorter travel time to cities, and higher population density areas (see Figure 6 and Figure A2, Appendix 1). Similarly, the RF model shows the highest predicted area (194,231 ha), followed by MaxEnt (152,934 ha) and BRT (128245 ha); representing 22%, 17%, and 15% of the study area respectively. The socio-economic factors constrain irrigation expansion suitability of 4% (RF), 4% (MAXENT), and 0% (BRT) compared to case 1 of biophysical predictors only. All three algorithms captured 50-52% of the estimated suitable area while any two models captured 22-23% of the suitable area. The remaining 26-27% is predicted by any one model of the three algorithms (see Figure 6d for the consensus mapping). These results suggest that the application of a multi-model approach provides higher confidence in the estimation of the suitable areas for farmer-led, small-scale irrigation.

3.3 Predictors importance

3.3.1 Case 1: Predictors importance under biophysical suitability modelling

When considering only biophysical predictors (case 1), accessibility to small reservoirs (D_RESRV) with a relative influence of 26% and exchangeable sodium percentage (ESP) with a relative influence of 20% are the top predictors of irrigation suitability for all three algorithms (See Figure 7). D_RESRV shows a decreasing exponential response curve to suitability (Figure 8), showing that a shorter distance to the reservoirs is a key indicator for small-scale irrigation expansion. This is consistent with the importance of surface water availability for dry season irrigation in semi-arid regions. ESP unexpectedly shows an increasing response curve with suitability. This may be explained by the low fertility and sodic soil in the region.

Aside from these two main predictors, the 3 algorithms diverged in terms of the ranking of other predictors. Overall, soil organic carbon content (ORC) — 6%, available soil water capacity (WWP) — 6%, bulk density (BLD) —5%, elevation percentile (ELVPERC) — 5%, base saturation percentage (BSP) — 4%, proximity to stream/rivers (D_RIVERS) — 4%, Surface runoff (SURQ) — 3%, exchangeable potassium (EXKX) — 3%, soil texture fraction (SAND) — 3% along with the two main predictors (D_RESRV and ESP) are the top 10 predictors (Figure 7d). ORC response curve increases with suitability up to 5 g/kg then decreases afterward. ORC is also a measure of soil nutrients for crops and water availability in the soil. WWP response curve decreases with suitability, which may be explained by the low soil moisture content of the semi-arid regions. ELEVPERC, which is the measure of the lowness of pixels relative to the surrounding upland, shows an increased, almost linear relationship with suitability. This suggests that small-scale irrigation tends to be located in the valley bottoms, lowlands, fringes where water availability and soil humidity are high compared to smaller irrigation suitability in the drier upland. This response is confirmed by the trend in the topographic wetness index (TWI), a proxy for soil moisture, and a compound index, although not part of the top 10 predictors, which has relative importance in the ranking. Surface runoff (SURQ), dynamically computed as the cumulative sum of the 4 months (July to October), is the fourth most important predictor. SURQ, which values range from 64 mm (which is mostly related to October as the onset of the dry season in the region) to 252 mm, is a proxy for water availability which can be captured in small reservoirs for dry season irrigation. The response curve of SURQ shows a complex relationship with suitability with maximum suitability at 100 mm, and a sharp decrease to 125 mm followed by a constant shape.

All models seldom included groundwater productivity (YGW), depth to groundwater table (WTD), slope, and pH as important variables (Figure 7d). However, the analysis of partial response curves is more revealing (see Figure 8). The response curve of the YGW shows a trend of increasing suitability with higher levels of aquifer productivity, which supports the general acceptance that high yield is important for groundwater-based irrigation although with relatively low groundwater potential across the region compared to surface water irrigation capacity. On the other hand, irrigation suitability shows a decreasing trend with increasing WTD, which is also consistent with the notion that deeper groundwater tables will be less accessible for small-scale irrigation that mostly depends on manual or low-powered lift technologies. EXKX, which is one of the indicators of soil fertility, response curve increases with suitability, while gentle slopes (less than 10%) represent the optimal terrain for small-scale irrigation. On the other hand, pH reveals a more complex response with a decreasing response curve of suitability between 5.5-6.0 and an increasing trend from pH of 6.0.

3.3.2 Case 2: Predictors importance under biophysical and socio-economic suitability modelling

When considering both biophysical and socio-predictors of irrigation suitability, the variable importance as ranked by the three algorithms shows some discrepancies in the position of the predictors, except for the accessibility (ACCESS) variable – measured as travel time to cities – that is consistently the most important predictor of irrigation suitability with a relative influence of 21% (Figure 7d). This shows the importance of shorter travel time for access to cities and urban centers in the suitability for small-scale, farmer-led irrigation. This is supported by the partial response curve of exponential decrease with suitability. The partial response of D_ROADs shows an exponential decrease with suitability. Road accessibility is a key factor for agricultural development for market access and input. As such D_ROADS is a key variable for farmer-led irrigation. The major crops irrigated in the dry season in the region are leafy vegetables, onions, pepper, okra, and tomatoes, which are mainly destined for urban centers and road access is a key determinant for the distribution of the harvested crops. Population density shows two main distributions: a decrease in suitability for small-scale irrigation between 50 to 250 persons/km^2 to a sharp increase in suitability for irrigation from 250 people/km^2. The jackknife test of AUC in the MaxEnt model confirms the importance of these variables with AUCs greater than 0.65 when individual predictors are considered. The same test shows that when these predictors are omitted, the drop in the overall AUC seems to be relatively high (Figure A3, Appendix 1), especially for POP_D and D_ROADS. Comparatively to case 1, Accessibility to the reservoirs (D_RESER)—10%, Exchangeable sodium percentage (ESP)—9%, Surface runoff (SURQ) —8%, distance to roads (D_ROADS) —6% elevation percentile (ELVEPRC) — 6% remain top predictors, showing the consistency in the model predictions.

3.4 Comparison between predictions and existing land use land cover data

We compared the estimated suitability for small-scale irrigation expansions with two recent land cover products: The 20 m resolution CCI land cover product of 2016 and the FAO Water Productivity Open-access portal (WaPOR) land use data (Figure 10 and Figure 11). Results show that the suitability areas suitable for small-scale irrigation correspond to shrubland, grassland, cropland, regularly flooded areas, and spare vegetation areas for the CCI land cover product. For the WaPOR land cover data, the suitability for small-scale irrigation corresponds to shrubland, grassland, rainfed cropland, irrigated cropland, and sparse vegetation areas. In both cases, the suitability is largely under rainfed cropland, corresponding to 77-86% and 82-89% for CCI land cover and WaPOR land use data respectively followed by shrubland (9-16%) and grassland (0.6-6.3%). Other land cover types such as regularly flooded areas and spare vegetation areas represent between less than 0.1% and 0.5%. The irrigated cropland in the WaPOR land cover product represents 0.1 % (194.2 ha) which are largely captured by the two major irrigation schemes in the study area, Vea and Tono.

4 DISCUSSION

4.1 Comparison of the predictions with existing irrigation suitability data

The current study estimated the suitability for small-scale irrigation in the Upper East at 179,584 ± 49,853[3] ha when biophysical variables only are considered in the modelling and 158,470 ± 27,222 ha when socio-economic variables are included along with the biophysical predictors. A previous study in the Upper East Region on irrigation potential based on four watersheds drained by the White Volta, Red Volta, Sissili, and Kulpawn rivers and information on gross irrigation water requirements (of onions, peppers, and tomatoes), area of soil suitable for irrigation and available water resources reported a value of 23,450 ha (Akomeah et al., 2009). The difference between our predictions and the aforementioned estimates is likely to result mainly in the study area and methodology used. We used machine learning to infer suitable conditions for small-scale irrigation of the entire Upper East Region based on a comprehensive survey of 707 irrigation plots while the former used limiting factors analysis on soil variables of only four catchments. In general, 86-89% of the suitability for farmer-led irrigation corresponds to the current rainfed-cropland of existing land cover data.

[3] These figures represent the ensemble (average) of the predicted areas from the 3 algorithms (BRT, RF, MaxEnt) and the standard deviation.

4.2 Drivers/predictors of small-scale irrigation suitability

Our findings show that the suitability for small-scale irrigation expansion in the study was strongly influenced by the accessibility or travel time to cities, indicative of the degree of connectivity between rural farming areas and surrounding urban markets and centers (Weiss et al., 2018). This finding is consistent with previous research elsewhere in West Africa that showed travel time to be an important factor influencing the expansion of rice cultivation in West Africa (Akpoti et al., 2020). In the context of the dry season irrigated agriculture in the Upper East Region of Ghana, where leafy vegetables, peppers, onions, and tomatoes are the main crops grown, timely transportation of the agricultural products to market is essential to ensure that these products are kept in good condition. In line with this assumption, accessibility to roads and population density were also among the top predictors of irrigation suitability further reinforcing the important role of market access for informal farmer-led irrigation developments. The importance of market access has also be recognized in other areas of SSA (Dorosh et al., 2012). In Mozambique, for example, the availability of market and population density for sufficient labor availability were among the major determinants of farmer-led irrigation expansion (Beekman et al., 2014). Similarly, it was shown that reforms in Mali concerning urbanization and better road access from the irrigated area to the market have created new incentives to irrigate (Wiggins & Lankford, 2019). This suggests that enhanced support for strategic road infrastructure development and improvements in market accessibility may be key for realizing irrigation suitability in the Upper East Region of Ghana, a result that is also likely to be broadly transferable to other parts of SSA.

Beyond the apparent role of these factors, the causal relationships among these factors are complex. For example, there is more competition over plots located close to a city or in a highly densified area, meaning they have a higher chance to be in the hand of better-off farmers, who have a better capacity to invest in irrigation and production costs. Thus, the links between “close to a city” and “irrigation” can be related to market proximity or to a higher ability to financially take charge of irrigation. An irrigation system is a complex social-ecological system where demographic variability influences the area of irrigation systems which may reduce or expand depending on the change in population and water availability (Puy et al., 2017). It is shown that the complex paths leading to the expansion of irrigated areas can be linked to a power function of population size (Puy, 2018). There is evidence of a positive relationship between population density and measures of land intensification (Muyanga & Jayne, 2014). In this study, the partial response of population with suitability for small-scale irrigation showed a complex pattern with lower population density and a positive linear trend for above 250 persons/km^2. However, the link between changes in population size and irrigated areas does not necessarily reflect a simple cause-effect relationship but a more nuanced process in which both variables influence each other through different feedback loops (Puy et al., 2017), reflecting underlying endogeneity (Muyanga & Jayne, 2014). This has implications for achieving a more sustainable balance between human and environmental welfare (Puy, 2018), especially considering that the population is expected to increase along with water usage for irrigation in SSA. Also, areas with lower population density are likely to be an area with limited economic potential/economic attractivity. It is therefore not surprising that such areas are characterized by fewer farmers with the financial capacity to irrigate than in other areas. Besides, it has been shown that households in areas with low potential and/or declining land access may choose to migrate to areas which they perceive to be of relatively higher potential and/or better land access (Muyanga & Jayne, 2014).

Many other constraints to the uptake of small-scale irrigation exits and have been discussed in the literature (Bjornlund et al., 2017; Lefore et al., 2019; Glitse et al., 2018) including but not limited to access to credit and finance for motorized pumps, access to appropriate technologies, gender among others. Land tenure and culture may also play a role in small-scale irrigation expansion (Xie et al., 2014). For most of these factors, the data is simply lacking or difficult to spatialize.

Alongside market access, accessibility to water for irrigation was one of the other key determinants of the suitability for small-scale irrigation expansion. Access to the surface water reservoirs, in particular, was overall the second most important predictor of irrigation suitability in the Upper East Region of Ghana, while the potential surface runoff that can be captured in a reservoir and other surface water storage for further use in the dry season was also among the top predictors for explaining irrigated area suitability. These results are consistent with the importance of small reservoirs as a buffer against variable rainfall conditions and limited surface runoff in semi-arid regions such as northern Ghana, with reservoirs providing a critical storage buffer to support dry-season irrigated crop production (Annor et al., 2009; Ghansah et al., 2018; Liebe et al., 2005). For example, dams and small reservoirs represented 21% of the water sources used for irrigation in the survey data considered in the modelling (see Figure S1, supplementary material 1 for water sources for irrigation in the study area). Our results could support the Government of Ghana's ''one village one dam'' program and Ghana's irrigation policy to prioritize areas that need additional dams and reservoirs for small-scale irrigation development.

Many small-scale irrigation plots in the Upper East Region of Ghana are irrigated through treadle and motorized pumps, river diversion, and watering cans pumping technologies (Tafesse, 2003). Although 19% of the irrigated plots considered in the modelling used rivers and streams as water sources for irrigation, accessibility to rivers only showed relative importance. Nevertheless, proximity to rivers and streams represents an important indicator for small-scale irrigation as shown by the partial response curve of accessibility to rivers and suitability level. Similarly, the areas modeled as suitable for small-scale irrigation in the present study largely coincided with groundwater potential zones classified as 'good' in terms of groundwater potential by Gumma & Pavelic, (2013). Groundwater represented 60% of the water source for irrigation in the study area. Groundwater yield and distance to the water table, which represented groundwater potential, showed an increasing and decreasing trend with suitability respectively but showed relatively low importance in the ranking of the main predictors. These results suggest that the availability of water alone is not a sufficient indicator for the expansion and/or adoption of small-scale irrigation. The potential for development should be put in the local context of market availability, demographic indicators, and infrastructure development.

Soil quality also plays an important role in cropland development, and therefore the suitability of land for irrigation. The exchangeable sodium percentage was the third overall most important predictor but showed an increasing trend with suitability, while in practice, ESP > 15 % is considered as problematic soil. This may be explained by poor soil fertility and the high sodicity of the study area. Soil fertility analysis of irrigation schemes in the study area showed a relatively high level of salinity and sodicity (Adongo et al., 2015). In general, low soil fertility is a major constraint for smallholder farmers in northern Ghana (Becx et al., 2012) with soil erosion as a major factor affecting soil fertility in the region (Amegashie et al., 2012). Besides, topography plays an important role in irrigation especially for water availability and transport. The positions of irrigated plots in this study were fringes (5%), inland valleys (16%), lowlands/floodplains (30%), and uplands (49%). The discrimination between these landscapes was represented by the elevation percentile; which showed an increasing trend with suitability and one of the top predictors. This showed that lowlands, which generally have relatively higher soil fertility and higher soil moisture compared to the surrounding upland may be more favourable for irrigation. This was supported by the trapezoidal partial response shape of topographic wetness index (although this variable was not a top predictor).

4.3 Uncertainties in the data used and limitations

The mapping of the suitability for small-scale irrigation mapping as presented in this research followed a rigorous modelling framework that has been previously been applied successfully for analyzing irrigated areas elsewhere in West Africa (Akpoti et al., 2021). Nevertheless, some limitations exist in our analysis and underlying data that warrant further discussion. The practice of small-scale, informal irrigation occurs in small cropland areas which are sometimes less than 2 ha. These croplands are in heterogeneous landscapes with varying properties. Even within the same irrigated plots, soil physio-chemical properties, water, and landscape features may vary. These variations are not always captured in the predictors' data used for the estimation of the suitability for small-scale irrigation, in which spatial resolutions varied from 30 m to 1 km. Although the simulations of the suitability for expansion of small-scale irrigation were computed at a pixel size of 30 m by 30 m, the aggregation and resampling of the predictor’s layers to this finer grain size did not necessarily improve the predictor's quality. For example, a small-scale irrigation plot located in an inland valley bottom where the groundwater table is high may have different water requirement compared to a plot located at the upland in the land scale (Rodenburg et al., 2014; Abe et al., 2010; Schmitter et al., 2015). However, using a water table depth predictor of 1 km by 1 km spatial resolution may show the same pixel value for the two irrigated plots in the example cited above, representing a limitation in the models’ predictions. Although the majority of the farmers used bucket/watering can in distributed irrigation settings, indicators that account for the energy costs involved in the different types of water resources transportations are not well captured. For example, in a non-flat area, farmers will develop irrigation preferably downstream of the dam/small reservoirs rather while upstream locations will be associated with higher energy costs and thus less favorable. Other limitations related to soil properties layers in Africa Soil Information Service (AfSIS) 250 m database were also reported by Hengl et al., (2015). Although the initial survey of farmer-led irrigation plots data showed some level of bias to roads and high population density areas, we applied spatial filtering to reduce sampling bias (Boria et al., 2014). This did not, however, exclude completely the bias to roads.

5 CONCLUSIONS AND OUTLOOK FOR FUTURE RESEARCH

We applied an ensemble of three machine learning algorithms to map the suitability for dry season small-scale irrigation in a semi-arid Upper East Region of Ghana. The results showed suitability of 179,584 ± 49,853 when biophysical variables only are considered in the modelling and 158,470 ± 27,222 ha when socio-economic variables are included along with the biophysical predictors, representing 77.3-88.8% of the current rainfed-cropland. The results suggested that the availability of water alone is not a sufficient indicator for the expansion and/or adoption of small-scale irrigation. Rather, the potential for development should be put in the local context of market availability, demographic indicators, and infrastructure development. It is important to note that causal relationships between the predictors and irrigated areas are complex with different feedback loops. This is because irrigation systems are complex social-ecological systems where many variabilities (e.g., demography) have an impact on the area of irrigation systems. In all, the approach developed in this research can be deployed in other countries in Africa to support food security interventions. All models' predictive performances were better than random, both at the training and validation stage, providing an indication of the models' generalizability or transferability to new regions.

Based on the water lifting technologies and water sources reported, there is a possibility to extend this work on cost-benefit analysis, thus providing an avenue for the private-public partnership for business development. Also, future research should consider a one-class classification problem approach to identify actual irrigation areas by using crops calendars, frequency of irrigation reported by the farmers, and spectral indices of remote sensing data based on the current small-scale irrigation survey data. The mapping procedure could help develop an operational irrigation monitoring tool based on remotely sensed temporal and spectral indices only. The maps are useful for identifying areas that are suitable areas for irrigation. Suitability does not however necessarily translate into high productivity or income gains. Areas may be suitable for irrigation development based on biophysical and socio-economic conditions, but additional supports may be required to help farmers to use irrigation productively and to intensify water use on existing irrigated land. Sustainable intensification is a key challenge alongside expanding irrigated areas.

ACKNOWLEDGEMENTS

This research was supported by the "FutureDAMS: Design and Assessment of resilient and sustainable interventions in water-energy-food-environment Mega-Systems" project (grant code: ES/P011373/1), through the Global Challenges Research Fund from the United Kingdom Research and Innovation (UKRI). The authors highly appreciate the constructive comments and suggestions from the two anonymous reviewers that significantly improved the earlier versions of the paper.

Appendix 1. Supplementary results

Figure A1. Continuous prediction maps with biophysical predictors only.

Figure A2. Continuous prediction maps with biophysical and socio-economic predictors.

Figure A3. Jackknife test of AUC for small-scale irrigation.

REFERENCES

Abe, S. S., Buri, M. M., Issaka, R. N., Kiepe, P., & Wakatsuki, T. (2010). Soil fertility potential for rice production in West African lowlands. *Japan Agricultural Research Quarterly*, *44*(4), 343–355. https://doi.org/10.6090/jarq.44.343

Adongo, T. A., Abagale, F. K., & Kranjac-Berisavljevic, G. (2015). Soil Quality of Irrigable Lands of Irrigation Schemes in Northern Ghana. *International Journal of Innovative Science, Engineering & Technology*, *2*(8), 314–326.

Akomeah, E., Odai, S. N., Annor, F. O., Adjei, K. A., & Barry, B. (2009). Mapping Irrigation Potential in the Upper East Region of Ghana. *2010 International Conference on Remote Sensing (ICRS)*, *11*(October 2016), 9666.

Akpoti, K., Kabo-bah, A. T., Dossou-Yovo, E., Groen, T., & Zwart, S. J. (2020). Mapping suitability for rice production in inland valley landscapes in Benin and Togo using environmental niche modelling. *Science of the Total Environment*, *709*(20 March 2020), 136165. https://doi.org/10.1016/j.scitotenv.2019.136165

Akpoti, K., Dossou-yovo, E. R., Zwart, S. J., & Kiepe, P. (2021). The potential for expansion of irrigated rice under alternate wetting and drying in Burkina Faso. *Agricultural Water Management*, *247*(31 March 2021), 106758. https://doi.org/10.1016/j.agwat.2021.106758

Akpoti, K., Kabo-bah, A. T., & Zwart, S. J. (2019). Agricultural land suitability analysis: State-of-the-art and outlooks for integration of climate change analysis. *Agricultural Systems*, *173*(February), 172–208. https://doi.org/10.1016/j.agsy.2019.02.013

Al-hassan, S. (2015). Food Security in the Upper East Region of Ghana : a Situational Analysis. *UDS International Journal of Development*, *2*(1), 69–85. http://www.udsijd.org

Al-shammary, A. A. G., Kouzani, A. Z., Kaynak, A., Khoo, S. Y., Norton, M., & Gates, W. (2018). Soil Bulk Density Estimation Methods : A Review. *Pedosphere: An International Journal*, *28*(4), 581–

596. https://doi.org/10.1016/S1002-0160(18)60034-7

Alemayehu, A., & Bewket, W. (2017). Smallholder farmers ' coping and adaptation strategies to climate change and variability in the central highlands of Ethiopia. *Local Environment*, *22*(7), 825–839. https://doi.org/10.1080/13549839.2017.1290058

Amegashie, B., Quansah, C., Agyare, A., Bonsu, M., & Odai, S. (2012). On-Site Effects and Cost of Fertility Erosion from Five Small Reservoir Catchments in the Upper East Region of Ghana. *Journal of Science and Technology (Ghana)*, *32*(2), 78–93. https://doi.org/10.4314/just.v32i2.10

Amjath-Babu, T. S., Krupnik, T. J., Kaechele, H., Aravindakshan, S., & Sietz, D. (2016). Transitioning to groundwater irrigated intensified agriculture in Sub-Saharan Africa: An indicator based assessment. *Agricultural Water Management*, *168*, 125–135. https://doi.org/10.1016/j.agwat.2016.01.016

Annor, F. O., van de Giesen, N., Liebe, J., van de Zaag, P., Tilmant, A., & Odai, S. N. (2009). Delineation of small reservoirs using radar imagery in a semi-arid environment: A case study in the upper east region of Ghana. *Physics and Chemistry of the Earth*, *34*(4–5), 309–315. https://doi.org/10.1016/j.pce.2008.08.005

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, *22*(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Assefa, T., Jha, M., Reyes, M., Srinivasan, R., & Worqlul, A. W. (2018). Assessment of suitable areas for home gardens for irrigation potential, water availability, and water-lifting technologies. *Water (Switzerland)*, *10*(4), 1–21. https://doi.org/10.3390/w10040495

Balana, B. B., Bizimana, J., Richardson, J. W., Lefore, N., Adimassu, Z., & Herbst, B. K. (2020). Economic and food security effects of small-scale irrigation technologies in northern Ghana. *Water Resources and Economics*, *29*(March 2019), 100141. https://doi.org/10.1016/j.wre.2019.03.001

Beck, J. (2013). Predicting climate change effects on agriculture from ecological niche modelling: Who profits, who loses? *Climatic Change*, *116*(2), 177–189. https://doi.org/10.1007/s10584-012-0481-x

Becx, G. A., Mol, G., Eenhoorn, J. W., van der Kamp, J., & van Vliet, J. (2012). Perceptions on reducing constraints for smallholder entrepreneurship in Africa: The case of soil fertility in Northern Ghana.

Current Opinion in Environmental Sustainability, *4*(5), 489–496. https://doi.org/10.1016/j.cosust.2012.10.012

Beekman, W., Veldwisch, G. J., & Bolding, A. (2014). Identifying the potential for irrigation development in Mozambique: Capitalizing on the drivers behind farmer-led irrigation expansion. *Physics and Chemistry of the Earth*, *76–78*, 54–63. https://doi.org/10.1016/j.pce.2014.10.002

Benjamin Ghansah, S. Z. (2020). *Technical report: Implementation of a Field Survey on Informal Irrigation in Upper East Region, Ghana* (Issue July).

Bjornlund, H., Rooyen, A. Van, & Stirzaker, R. (2017). Profitability and productivity barriers and opportunities in small-scale irrigation schemes. *International Journal of Water Resources Development*, *33*(5), 690–704. https://doi.org/10.1080/07900627.2016.1263552

Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, *275*, 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H., & Bollmann, K. (2013). Selecting from correlated climate variables : a major source of uncertainty for predicting species distributions under climate change. *Ecography*, *36*(9), 1–13. https://doi.org/10.1111/j.1600-0587.2013.00138.x

Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (2011). Package randomForest. In *Software available at: http://stat-www. berkeley. edu/users/breiman/RandomForests.* https://doi.org/10.1023/A

Breiman, L. (2011). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, J. L., Bennett, J. R., & French, C. M. (2017). SDMtoolbox 2.0: The next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, *2017*(12). https://doi.org/10.7717/peerj.4095

Burney, J. A., Naylor, R. L., & Postel, S. L. (2013). The case for distributed irrigation as a development priority in sub-Saharan Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(31), 12513–12517. https://doi.org/10.1073/pnas.1203597110

Castro-Llanos, F., Hyman, G., Rubiano, J., Ramirez-Villegas, J., & Achicanoy, H. (2019). Climate change favors rice production at higher elevations in Colombia. *Mitigation and Adaptation*

Strategies for Global Change, *24*(8), 1401–1430. https://doi.org/10.1007/s11027-019-09852-x

Center for International Earth Science Information Network (CIESIN) Columbia University. (2017). NASA Socioeconomic Data and Applications Center (SEDAC) Documentation for Gridded Population of the World (GPW), v4. *Center for International Earth Science Information Network*, *4*(845), 1–51. https://doi.org/10.7927/H4B56GPT

Cooper, P. J. M., & Coe, R. (2011). Assessing and addressing climate-induced risk in sub-Saharan rainfed agriculture. *Experimental Agriculture*, *47*(2), 179–184. https://doi.org/10.1017/S0014479711000019

Corwin, D. L., & Lesch, S. M. (2003). Application of soil electrical conductivity to precision agriculture: theory, principles, and guideline. *Agronomy Journal*, *95*(3), 455–471.

Crimmins, S. M., Dobrowski, S. Z., & Mynsberge, A. R. (2013). Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling*, *266*, 126–130. https://doi.org/10.1016/j.ecolmodel.2013.07.006

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In *Ensemble machine learning* (pp. 157–175). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9326-7

Dadhich, G., Patel, P. R., & Kalubarme, M. H. (2017). Agriculture land suitability evaluation for wheat cultivation using geomatics for Patan District, India. *International Journal of Agricultural Resources, Governance and Ecology*, *13*(1), 91–108. https://doi.org/10.1504/IJARGE.2017.084038

de Bont, C., Komakech, H. C., & Veldwisch, G. J. (2019). Neither modern nor traditional: Farmer-led irrigation development in Kilimanjaro Region, Tanzania. *World Development*, *116*, 15–27. https://doi.org/10.1016/j.worlddev.2018.11.018

Dietz, T., Millar, D., Dittoh, S., Obeng, F., & Ofori-Sarpong, E. (2004). Climate and livelihood change in North East Ghana. In *The impact of climate change on drylands* (pp. 149–172). Springer, Dordrecht.

Dittoh, S., Awuni, J. A., & Akuriba, M. A. (2013). Small pumps and the poor: A field survey in the Upper East Region of Ghana. *Water International*, *38*(4), 449–464. https://doi.org/10.1080/02508060.2013.819454

Domènech, L. (2015). Improving irrigation access to combat food insecurity and undernutrition : A review. *Global Food Security*, *6*, 24–33. https://doi.org/10.1016/j.gfs.2015.09.001

Dorosh, P., Wang, H. G., You, L., & Schmidt, E. (2012). Road connectivity, population, and crop production in Sub-Saharan Africa. *Agricultural Economics*, *43*(1), 89–103. https://doi.org/10.1111/j.1574-0862.2011.00567.x

Drechsel, P., Graefe, S., Sonou, M., & Cofie, O. O. (2006). *Informal Irrigation in Urban West Africa: An Overview* (I. W. M. Institute. (ed.); 40.p. (IWM).

Duong, T. (2015). ks : Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R . *Journal of Statistical Software*, *21*(7). https://doi.org/10.18637/jss.v021.i07

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., ... & Li, J. (2006). Novel methods improve prediction of species ' distributions from occurrence data. *Ecography*, *29*(January), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elith, Jane, & Graham, C. H. (2009). Do they ? How do they ? WHY do they differ ? On finding reasons for differing performances of species distribution models. *Ecography*, *33*(December 2008), 66–77. https://doi.org/10.1111/j.1600-0587.2008.05505.x

Elith, Jane, & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Evans, J., Murphy, M. A., & Ram, K. (2020). Package "spatialEco." *Cran.R-Project*. https://cran.r-project.org/web/packages/spatialEco/index.html

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science*, *339*(6122), 940–943. https://doi.org/10.1126/science.1229881

FAO. (2008). *AgWA: Partnership for agricultural water for Africa*. http://www.fao.org/agwa/background/what-is-agwa/en/

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Florczyk, A. J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., Sabo, F., & Zanchetta, L. (2019). *GHSL Data Package 2019 - Technical report by the Joint Research Centre (JRC), European Union*. https://doi.org/10.2760/0726

Foster, T., Brozović, N., & Butler, A. P. (2015). Analysis of the impacts of well yield and groundwater depth on irrigated agriculture. *Journal of Hydrology*, *523*, 86–96. https://doi.org/10.1016/j.jhydrol.2015.01.032

Fuka, D. R., Walter, M. T., Macalister, C., Degaetano, A. T., Steenhuis, T. S., & Easton, Z. M. (2014). Using the Climate Forecast System Reanalysis as weather input data for watershed models. *Hydrological Processes*, *28*(22), 5613-5623. https://doi.org/10.1002/hyp.10073

Gallant, J. C., & Dowling, T. I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*, *39*(12). https://doi.org/10.1029/2002WR001426

Ghana Irrigation Development Authority. (2011). *National Irrigation Policy, Strategies and Regulatory Measures* (Issue May). Ghana Irrigation Development Authority of the Ministry of Food and Agriculture. http://extwprlegs1.fao.org/docs/pdf/gha149500.pdf (Accessed on December 13, 2020)

Ghansah, B., Forkuo, E. K., Osei, E. F., Appoh, R. K., Asare, M. Y., & Kluste, N. A. B. (2018). Mapping the spatial distribution of small reservoirs in the White Volta Sub-basin of Ghana. *Remote Sensing Applications: Society and Environment*, *9*(November 2017), 107–115. https://doi.org/10.1016/j.rsase.2017.12.003

Glitse, P., Nyamadi, B. V., Mintah, K. A. and Feruta-Benee, C. (2018). Small Scale Irrigation in Ghana: Challenges and Prospects in the Face of Climate Variability. *International Journal of Irrigation and Agricultural Development (IJIRAD)*, *2*(1), 133–143.

Gumma, M. K., & Pavelic, P. (2013). Mapping of groundwater potential zones across Ghana using remote sensing, geographic information systems, and spatial modelling. *Environmental Monitoring and Assessment*, *185*(4), 3561–3579. https://doi.org/10.1007/s10661-012-2810-y

Gumma, M. K., Thenkabail, P. S., Hideto, F., Nelson, A., Dheeravath, V., Busia, D., & Rala, A. (2011). Mapping irrigated areas of Ghana using fusion of 30 m and 250 m resolution remote-sensing data. *Remote Sensing*, *3*(4), 816–835. https://doi.org/10.3390/rs3040816

Havlin, J. L. (2005). Ferility. *Encyclopedia of Soils in the Environment*, 10–19. https://doi.org/https://doi.org/10.1016/B0-12-348530-4/00228-9

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., De Jesus, J. M., Tamene, L., & Tondoh, J. E. (2015). Mapping Soil Properties of Africa at 250 m Resolution : Random Forests Significantly Improve Current Predictions. *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0125814

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. In *Plos One* (Vol. 12, Issue 2). https://doi.org/10.1371/journal.pone.0169748

Heumann, B. W., Walsh, S. J., & McDaniel, P. M. (2011). Assessing the application of a geographic presence-only model for land suitability mapping. *Ecological Informatics*, *6*(5), 257–269. https://doi.org/10.1016/j.ecoinf.2011.04.004

Higginbottom, T. P., Adhikari, R., Dimova, R., Redicker, S., & Foster, T. (2021). Performance of large-scale irrigation projects in sub-Saharan Africa. *Nature Sustainability*. https://doi.org/https://doi.org/10.1038/s41893-020-00670-7

Hirzel, A. H., Helfer, V., & Metral, F. (2001). Assessing Habitat-Suitability Models with a virtual species. *Ecological Modelling*, *145*(2–5), 111–121.

Hochman, Z., Gobbett, D., Holzworth, D., McClelland, T., van Rees, H., Marinoni, O., Garcia, J. N., & Horan, H. (2013). Reprint of “Quantifying yield gaps in rainfed cropping systems: A case study of wheat in Australia.” *Field Crops Research*, *143*, 65–75. https://doi.org/10.1016/j.fcr.2013.02.001

Ismail, M., Abdel Ghaffar, M. K., & Azzam, M. A. (2012). GIS application to identify the potential for certain irrigated agriculture uses on some soils in Western Desert, Egypt. *Egyptian Journal of Remote Sensing and Space Science*, *15*(1), 39–51. https://doi.org/10.1016/j.ejrs.2012.03.001

Jarnevich, C., Young, N., Sheffels, T., Carter, J., Sytsma, M., & Talbert, C. (2017). Evaluating simplistic methods to understand current distributions and forecast distribution changes under climate change scenarios: an example with coypu (Myocastor coypus). *NeoBiota*, *32*, 107–125. https://doi.org/10.3897/neobiota.32.8884

Jiménez-Valverde, A. (2014). Threshold-dependence as a desirable attribute for discrimination assessment: Implications for the evaluation of species distribution models. *Biodiversity and Conservation*, *23*(2), 369–385. https://doi.org/10.1007/s10531-013-0606-1

Jiménez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, *31*(3), 361–369. https://doi.org/10.1016/j.actao.2007.02.001

Juhos, K., Czigány, S., Madarász, B., & Ladányi, M. (2019). Interpretation of soil quality indicators for land suitability assessment – A multivariate approach for Central European arable soils. *Ecological Indicators*, *99*(November 2018), 261–272. https://doi.org/10.1016/j.ecolind.2018.11.063

Kong, C., Lan, H., Yang, G., & Xu, K. (2016). Geo-environmental suitability assessment for agricultural land in the rural–urban fringe using BPNN and GIS: a case study of Hangzhou. *Environmental Earth Sciences*, *75*(15). https://doi.org/10.1007/s12665-016-5956-z

Lefore, N., Giordano, M., Ringler, C., & Barron, J. (2019). *Viewpoint – Sustainable and Equitable Growth in Farmer-led Irrigation in Sub-Saharan Africa : What Will it Take ? 12*(1), 156–168.

Liebe, J., van de Giesen, N., & Andreini, M. (2005). Estimation of small reservoir storage capacities in a semi-arid environment. *Physics and Chemistry of the Earth*, *30*(6-7 SPEC. ISS.), 448–454. https://doi.org/10.1016/j.pce.2005.06.011

Lindsay, J. B. (2016). Whitebox GAT: A case study in geomorphometric analysis. *Computers and Geosciences*. https://doi.org/10.1016/j.cageo.2016.07.003

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*(4), 778–789. https://doi.org/10.1111/jbi.12058

Lowder, S. K., Skoet, J., & Raney, T. (2016). The Number , Size , and Distribution of Farms , Smallholder Farms , and Family Farms Worldwide q. *World Development*, *87*, 16–29.

https://doi.org/10.1016/j.worlddev.2015.10.041

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, *15*(1), 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Mendes D. M., Paglietti L., Jackson D., & A. A. G. (2014). Ghana: Irrigation market brief. In *FAO Investment Centre, Rome*.

Misra, A. K. (2014). Climate change and challenges of water and food security. *International Journal of Sustainable Built Environment*, *3*(1), 153–165. https://doi.org/10.1016/j.ijsbe.2014.04.006

Morisette, J. T., Jarnevich, C. S., Holcombe, T. R., Talbert, C. B., Ignizio, D., Talbert, M. K., Silva, C., Koop, D., Swanson, A., & Young, N. E. (2013). VisTrails SAHM: Visualization and workflow management for species habitat modelling. *Ecography*, *36*(2), 129–135. https://doi.org/10.1111/j.1600-0587.2012.07815.x

Müller, D., Leitão, P. J., & Sikor, T. (2013). Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. *Agricultural Systems*, *117*, 66–77. https://doi.org/10.1016/j.agsy.2012.12.010

Muyanga, M., & Jayne, T. S. (2014). Effects of rising rural population density on smallholder agriculture in Kenya. *Food Policy*, *48*, 98–113. https://doi.org/10.1016/j.foodpol.2014.03.001

Nachtergaele, F., Velthuizen, H. van, & Verelst, L. (2010). The harmonized world soil database. verson 1.0. *Fao/Iiasa/Isric/Iss-Cas/Jrc*.

Nakawuka, P., Langan, S., Schmitter, P., & Barron, J. (2018). A review of trends, constraints and opportunities of smallholder irrigation in East Africa. *Global Food Security*, *17*(July), 196–212. https://doi.org/10.1016/j.gfs.2017.10.003

Namara, R. E.; Sally, H. (2014). *Proceedings of the Workshop on Irrigation in West Africa: Current Status and a View to the Future, Ouagadougou, Burkina Faso, December 1-2, 2010* (H. Namara, R. E.; Sally (ed.); 1st-2 Decemb ed.). International Water Management Institute (IWMI). https://doi.org/10.5337/2014.218

Namara, R. E., Awuni, J. A., Barry, B., Giordano, M., Hope, L., Owusu, E. S., & Forkuor, G. (2011). Smallholder Shallow Groundwater Irrigation Development in the Upper East Region of Ghana I n t e r n a t i o n a l Water Management I n s t i t u t e. In *International Water Management Institute (IWMI). 3p. (AgWater SolutionsIWMI Research Report* (Vol. 139). https://doi.org/10.5337/2011.214

Namara, R. E., Gebregziabher, G., Giordano, M., & De Fraiture, C. (2013). Small pumps and poor farmers in Sub-Saharan Africa: An assessment of current extent of use and poverty outreach. In *Water International* (Vol. 38, Issue 6, pp. 827–839). https://doi.org/10.1080/02508060.2014.847777

Namara, R. E., Horowitz, L., Kolavalli, S., Kranjac-berisavljevic, G., Dawuni, B. N., Barry, B., & Giordano, M. (2010). *Typology of Irrigation Systems in Ghana* (IWMI Working Paper 142). https://doi.org/10.5337/2011.200

Namara, R. E., Horowitz, L., Nyamadi, B., & Barry, B. (2011). Irrigation Development in Ghana: Past experiences, emerging opportunities, and future directions. In *Ghana Strategy Support Program (GSSP) GSSP Working Paper No. 0027* (No. 27). http://agriskmanagementforum.org/sites/agriskmanagementforum.org/files/Documents/IFPRI - Irrigation in Ghana.pdf

Neitsch, S. ., Arnold, J. ., Kiniry, J. ., & Williams, J. . (2011). *Soil & Water Assessment Tool Theoretical Documentation Version 2009*. 1–647.

NEPAD (New Partnership for Africa's Development). (2003). *Compherensive Africa Agriculture Development Programme (CAADP)*. http://www.nepad.org/foodsecurity/ agriculture/land

Obour, P. B., Lamandé, M., Edwards, G., Sørensen, C. G., & Munkholm, L. J. (2017). Predicting soil workability and fragmentation in tillage : a review. *Soil Use and Management*, *33*(June), 288–298. https://doi.org/10.1111/sum.12340

Ohana-Levi, N., Bahat, I., Peeters, A., Shtein, A., Netzer, Y., Cohen, Y., & Ben-Gal, A. (2019). A weighted multivariate spatial clustering model to determine irrigation management zones. *Computers and Electronics in Agriculture*, *162*(February), 719–731. https://doi.org/10.1016/j.compag.2019.05.012

Panda, R. K., Behera, S. K., & Kashyap, P. S. (2004). Effective management of irrigation water for maize under stressed conditions. *Agricultural Water Management*, *66*, 181–203.

https://doi.org/10.1016/j.agwat.2003.12.001

Peterson, A. T. (2006). Uses and Requirements of Ecological Niche Models and Related Distributional Models. *Biodiversity Informatics*, *3*(0), 59–72. https://doi.org/10.17161/bi.v3i0.29

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). Ecological Niches and Geographic Distributions (MPB-49). In *Ecological Niches and Geographic Distributions (MPB-49)*. https://doi.org/10.23943/princeton/9780691136868.001.0001

Phillips, S. J., & Dudík, M. (2008). Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, *32*(2), 161-175. https://doi.org/10.1111/j.2007.0906-7590.05203.x

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*(1), 181–197. https://doi.org/10.1890/07-2153.1

Puy, A. (2018). Irrigated areas grow faster than the population. *Ecological Applications*, *28*(6), 1413–1419. https://doi.org/10.1002/eap.1743

Puy, A., Muneepeerakul, R., & Balbo, A. L. (2017). Size and stochasticity in irrigated social-ecological systems. *Scientific Reports*, *7*(January), 1–9. https://doi.org/10.1038/srep43943

Ramírez-Gil, J. G., Morales, J. G., & Peterson, A. T. (2018). Potential geography and productivity of "Hass" avocado crops in Colombia estimated by ecological niche modelling. *Scientia Horticulturae*, *237*(October 2017), 287–295. https://doi.org/10.1016/j.scienta.2018.04.021

Rebelo, L. M., McCartney, M. P., & Finlayson, C. M. (2010). Wetlands of Sub-Saharan Africa: Distribution and contribution of agriculture to livelihoods. *Wetlands Ecology and Management*, *18*(5), 557–572. https://doi.org/10.1007/s11273-009-9142-x

Reddy, S., & Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, *30*(11), 1719–1727. https://doi.org/10.1046/j.1365-

2699.2003.00946.x

Ringler, C., Mekonnen, D. K., Xie, H., & Uhunamure, A. M. (2020). Irrigation to Transform Agriculture and Food Systems in Africa South of the Sahara. In *2020 Annual trends and outlook report: Sustaining Africa's agrifood system transformation: The role of public policies* (pp. 57–70). International Food Policy Research Institute (IFPRI).

Rodenburg, J., Zwart, S. J., Kiepe, P., Narteh, L. T., Dogbe, W., & Wopereis, M. C. S. (2014). Sustainable rice production in African inland valleys: Seizing regional potentials through local approaches. *Agricultural Systems*, *123*, 1–11. https://doi.org/10.1016/j.agsy.2013.09.004

Safari, Y., Esfandiarpour-boroujeni, I., Kamali, A., Salehi, M. H., & Bagheri-bodaghabadi, M. (2013). Qualitative Land Suitability Evaluation for Main Irrigated Crops in the Shahrekord Plain, Iran: A Geostatistical Approach Compared with Conventional Method. *Pedosphere*, *23*(6), 767–778. https://doi.org/10.1016/S1002-0160(13)60068-5

Sahour, H., Gholami, V., & Vazifedan, M. (2020). A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. *Journal of Hydrology*, *591*(May), 125321. https://doi.org/10.1016/j.jhydrol.2020.125321

Schmidhuber, J., & Tubiello, F. N. (2007). Global food security under climate change. *Proceedings of the National Academy of Sciences*, *104(50)*, 19703–19708.

Schmidt, F., & Persson, A. (2003). Comparison of DEM data capture and topographic wetness indices. *Precision Agriculture*, *4*(2), 179–192. https://doi.org/10.1023/A:1024509322709

Schmitter, P., Kibret, K. S., Lefore, N., & Barron, J. (2018). Suitability mapping framework for solar photovoltaic pumps for smallholder farmers in sub-Saharan Africa. *Applied Geography*, *94*(March), 41–57. https://doi.org/10.1016/j.apgeog.2018.02.008

Schmitter, P., Zwart, S. J., Danvi, A., & Gbaguidi, F. (2015). Contributions of lateral flow and groundwater to the spatio-temporal variation of irrigated rice yields and water productivity in a West-African inland valley. *Agricultural Water Management*, *152*(2015), 286–298. https://doi.org/10.1016/j.agwat.2015.01.014

Singh, K., McClean, C. J., Büker, P., Hartley, S. E., & Hill, J. K. (2017). Mapping regional risks from climate change for rainfed rice cultivation in India. *Agricultural Systems*, *156*, 76–84.

https://doi.org/10.1016/j.agsy.2017.05.009

Stohlgren, T. J., Ma, P., Kumar, S., Rocca, M., Morisette, J. T., Jarnevich, C. S., & Benson, N. (2010). Ensemble habitat mapping of invasive plant species. *Risk Analysis*, *30*(2), 224–235. https://doi.org/10.1111/j.1539-6924.2009.01343.x

Tafesse, M. (2003). *Small - scale irrigation for food security in sub - Saharan Africa* (No. 8031; CTA Working Document). https://hdl.handle.net/10568/63612

Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*, *December*. https://doi.org/10.1016/j.aci.2018.08.003

Turner, B. (1994). Small-scale irrigation in developing countries. *Land Use Policy*, *11*(4), 251–261. https://doi.org/10.1016/0264-8377(94)90051-5

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). R package “corrplot”: Visualization of a Correlation Matrix. *Statistician*, *56*, 316–324. https://github.com/taiyun/corrplot

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C. D., Howes, R. E., Tusting, L. S., Kang, S. Y., Cameron, E., Bisanzio, D., … Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, *553*(7688), 333–336. https://doi.org/10.1038/nature25181

Wiggins, S., & Lankford, B. (2019). *Farmer-led irrigation in sub-Saharan Africa: synthesis of current understandings* (Issue July). https://degrp.odi.org/wp-content/uploads/2019/07/DEGRP-Synthesis-Farmer-led-Irrigation.pdf (Accessed on August, 7th 2020)

Worqlul, A. W., Dile, Y. T., Jeong, J., Adimassu, Z., Lefore, N., Gerik, T., Srinivasan, R., & Clarke, N. (2019). Effect of climate change on land suitability for surface irrigation and irrigation potential of the shallow groundwater in Ghana. *Computers and Electronics in Agriculture*, *157*(August 2018), 110–125. https://doi.org/10.1016/j.compag.2018.12.040

Worqlul, A. W., Jeong, J., Dile, Y. T., Osorio, J., Schmitter, P., Gerik, T., Srinivasan, R., & Clark, N. (2017). Assessing potential land suitable for surface irrigation using groundwater in Ethiopia. *Applied Geography*, *85*, 1–13. https://doi.org/10.1016/j.apgeog.2017.05.010

Xie, H., You, L., Dile, Y. T., Worqlul, A. W., Bizimana, J. C., Srinivasan, R., Richardson, J. W., Gerik, T., & Clark, N. (2021). Mapping development potential of dry-season small-scale irrigation in Sub-Saharan African countries under joint biophysical and economic constraints - An agent-based modelling approach with an application to Ethiopia. *Agricultural Systems*, *186*(January 2021), 102987. https://doi.org/10.1016/j.agsy.2020.102987

Xie, H., You, L., & Takeshima, H. (2017). Invest in small-scale irrigated agriculture: A national assessment on potential to expand small-scale irrigation in Nigeria. *Agricultural Water Management*, *193*, 251–264. https://doi.org/10.1016/j.agwat.2017.08.020

Xie, H., You, L., Wielgosz, B., & Ringler, C. (2014). Estimating the potential for expanding smallholder irrigation in Sub-Saharan Africa. *Agricultural Water Management*, *131*, 183–193. https://doi.org/10.1016/j.agwat.2013.08.011

Yadav, J. S. P., Massoud, F. I., & Abrol, I. P. (1988). Salt-Affected Soils and their Management. In *Fao Soils Bulletin 39*.

You, L., Ringler, C., Wood-Sichra, U., Robertson, R., Wood, S., Zhu, T., Nelson, G., Guo, Z., & Sun, Y. (2011). What is the irrigation potential for Africa? A combined biophysical and socioeconomic approach. *Food Policy*, *36*(6), 770–782. https://doi.org/10.1016/j.foodpol.2011.09.001

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-arroita, G., Guisan, A., Lahoz-monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, *43*(9), 1261–1277. https://doi.org/10.1111/ecog.04960

Table 1. List of potential predictors. Some predictors were dropped from the final modelling due to their high correlation with other predictors (total nitrogen and clay) or least contribution in the variable importance such as solar radiation, precipitation of the wettest quarter, silt content, and total nitrogen. Distance to cities predictor was dropped due to biased it induced in the predictions.

No	Name	Definition	Resolution	Unit	Included in the final modelling (Yes/No)
	Socio-economic predictors				
1	D_ROADS	Proximity to main roads [a]	30m	m	Yes
2	D_TOWN	Proximity to towns [a]	30m	m	No
3	ACCESS	Travel time to major cities [b]	1km	min	Yes
4	POP	Population density [c]	1km	persons.km^{-2}	Yes
	Water availability				
5	WTD	Depth to water table [d]	1km	m	Yes
6	YGW	Aquifer productivity [e]	30m	l.s^{-1}	Yes
7	D_RIVERS	Proximity to stream/rivers [a]	30m	m	Yes
8	SURQ	Cumulative 5-years average surface runoff for the months of July, August, September and October [f]	30m	mm	Yes
9	D_RESER	Proximity to small reservoirs [a]	30m	km	Yes
Climate and energy					
10	BIO16	Precipitation of the wettest quarter [g]	1km	mm	No
11	SRD	Solar radiation [g]	1km	kJ.m^{-2}.day^{-1}	No
(a)	**Topography**				
12	SLOPE	Slope [h]	30m	%	Yes
13	ELVPERC	Elevation percentile [i]	30	%	Yes
14	TWI	Topographic wetness index [i]	30m	Index	Yes
15	FLACC	Flow accumulation			No
Soil physical properties					
16	BLD	Bulk density [j]	250m	kg.dm^{-3}	Yes
17	WWP	Available soil water capacity [j]	250m	%	Yes
18	CRFVOL	Coarse fragments volumetric fraction [j]	250m	%	Yes
19	CLAY	Soil texture fraction clay [j]	250m	%	No
20	SILT	Soil texture fraction silt [j]	250m	%	No
21	SAND	Soil texture fraction sand [j]	250m	%	Yes
Soil Chemical properties					
22	CEC	Cation exchange capacity [j]	250m	cmol.kg^{-1}	Yes
23	ECN	Electrical conductivity [j]	250m	dS.m $^{-1}$	Yes
24	BSP	Base saturation percentage [k]	250m	%	Yes
25	ESP	Exchangeable sodium percentage [k]	250m	%	Yes
26	EXKX	Exchangeable K [j]	250m	cmol.kg^{-1}	Yes
27	TPHOS	Total phosphorus [j]	250m	mg.kg^{-1}	Yes
28	NTO	Total nitrogen [j]	250m	g.kg^{-1}	No
29	ORC	Soil organic carbon content [j]	250m	g.kg^{-1}	Yes
30	PH	Soil pH in water [j]	250m	g.kg^{-1}	Yes

Data source: [a] Computed using Euclidean Distance tool in ArcGIS based on OpenStreetMap data; [b] Weiss et al., 2018; [c] CIESIN, 2016; [d] Fan et al., 2013; [e] Computed using an inverse distance weighted (IDW) on borehole data on yield; [f] Modelled using Soil and Water Assessment Tool (SWAT); [g] Obtained from the WorldClim data based (Fick & Hijmans, 2017); [h] Computed using SLOPE tool in ArcGIS based on ArcGIS; [i] Computed in R using the whitebox package; [j] Hengl et al., 2017; [k] Computed using raster calculator in ArcGIS.

Table 2. Evaluation of the models.

Evaluation metrics	BRT		MaxEnt		RF	
	Train	Test	Train	Test	Train	Test
Area Under the Curve (AUC)	0.95	0.90	0.93	0.88	0.88	0.89
Percentage Correctly Classified (PCC)	87.7	80.5	84.6	79.4	80.6	80.4
Sensitivity	0.88	0.81	0.84	0.78	0.81	0.80
Specificity	0.88	0.81	0.85	0.79	0.81	0.80
True Skill Statistics (TSS)	0.76	0.62	0.69	0.58	0.61	0.60

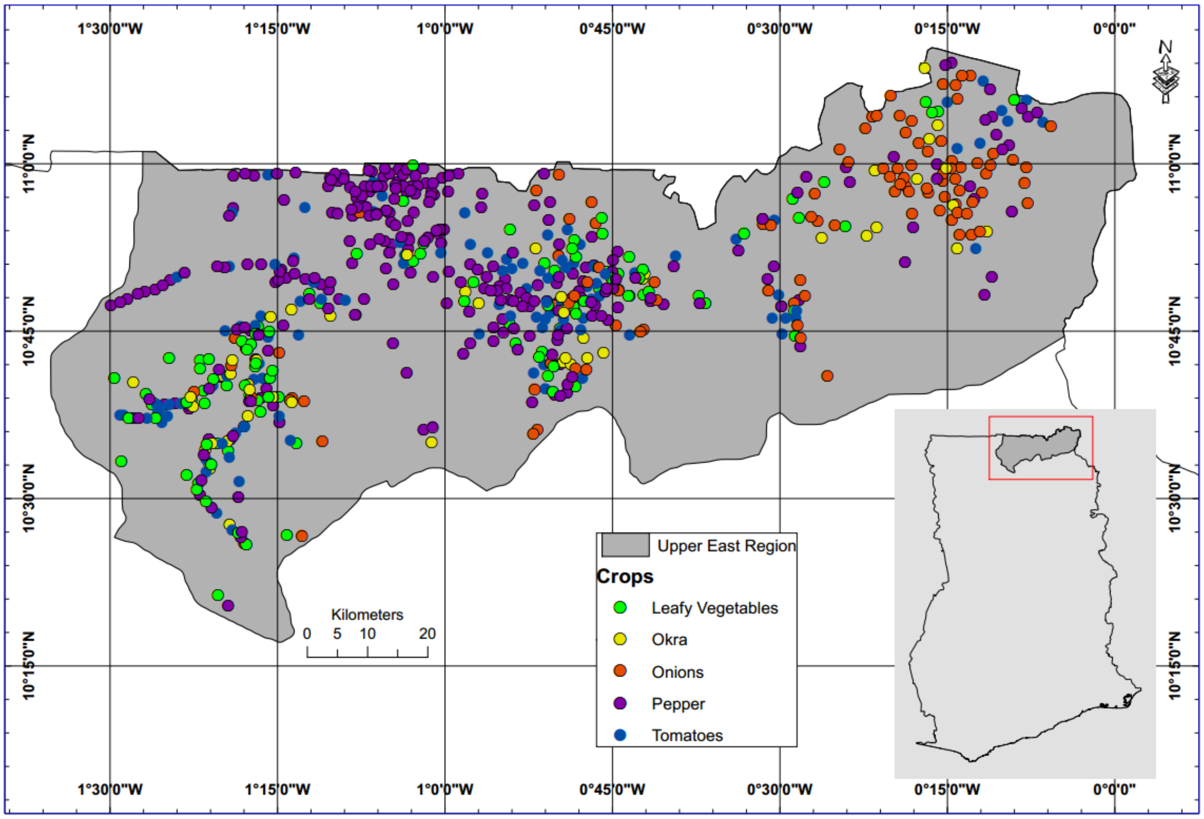

Figure 1. Study area map showing the spatial distributions of crops such as leafy vegetables, okra, onions, peppers, tomatoes among others under informal and farmer-led dry season irrigation in the Upper East Region of Ghana. The data was collected during a plot-level survey, implemented by the International Water Management Institutes (IWMI) and the University of Manchester between May and July 2020. Information in the data collected include geographic coordinates of the plots, nearest community, plot size, physical characteristic of the landscape, irrigation facilities, plot usage, crop calendars, seasonal irrigation practices, irrigation constraints, and available investment opportunities.

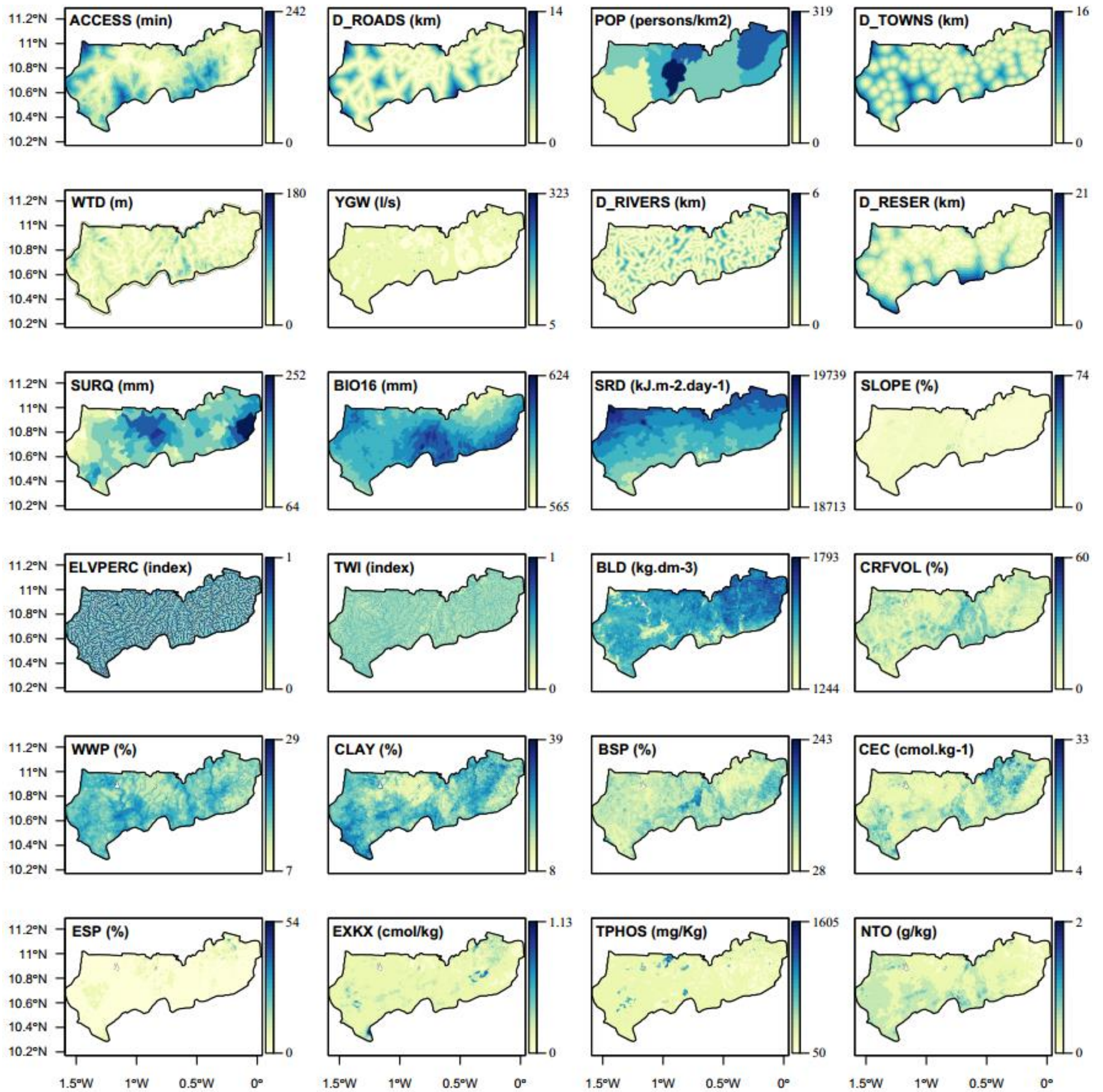

Figure 2. List of selected predictors. ACCESS—accessibility or travel time to cities, D_ROADS—distance to roads, POP_D—population density, D_TOWNS—proximity to towns, WTD—depth to water table, YGW—aquifer productivity, D_RIVERS—proximity to stream/rivers, D_RESRV—proximity to small reservoirs, SURQ—surface runoff, BIO16—precipitation of the wettest quarter, SRQ—solar radiation, SLOPE—slope, ELVPERC—elevation percentile, TWI— topographic wetness index, BLD—bulk density, CRFVOL—coarse fragments volumetric fraction, WWP—available soil water capacity, CLAY—soil texture fraction clay, BSP—base saturation percentage, CEC—cation exchange capacity, ESP—exchangeable

sodium percentage, EXKX—exchangeable potassium, TPHOS—total phosphorus, NTO—total nitrogen, PH—soil pH in water.

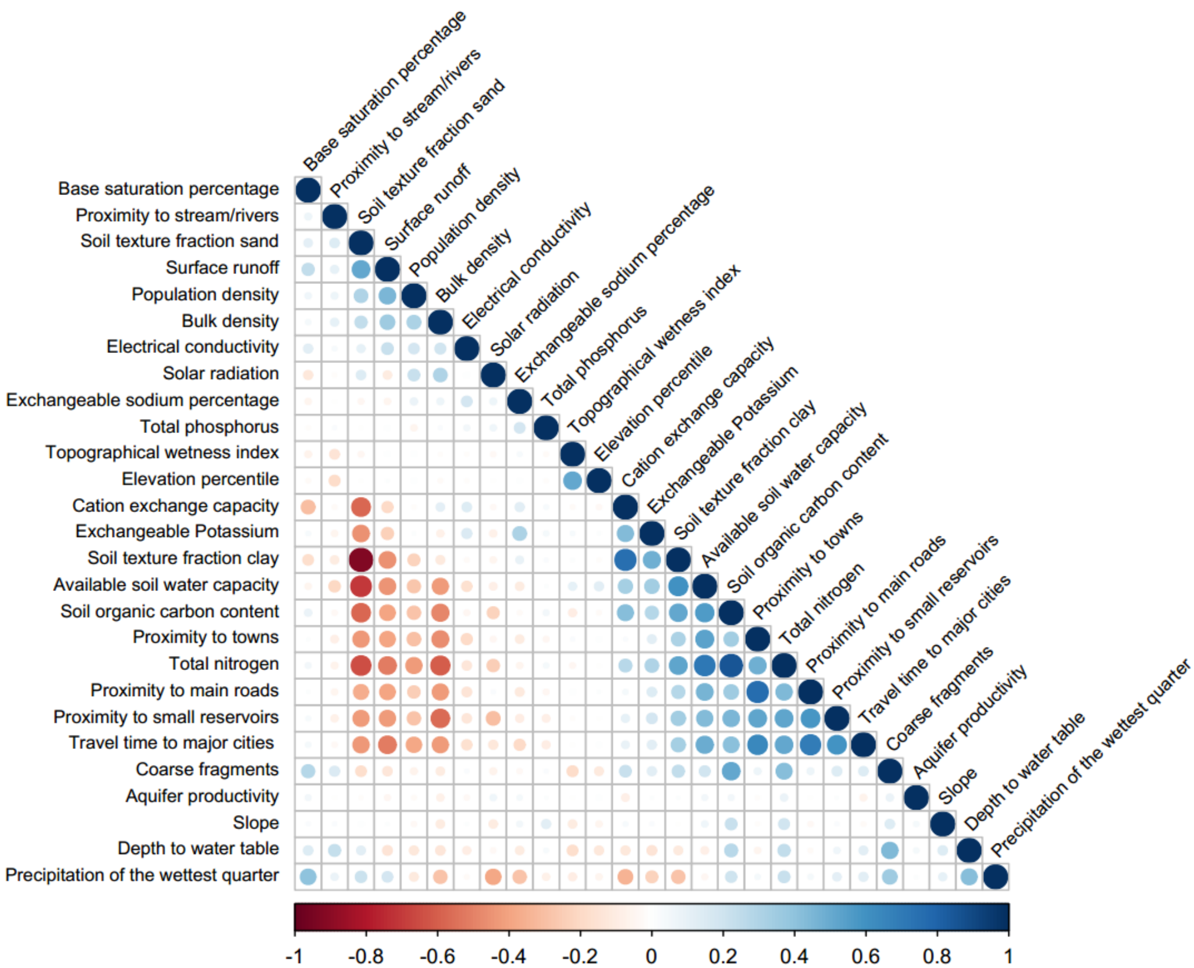

Figure 3. Pearson correlation coefficient among predictors.

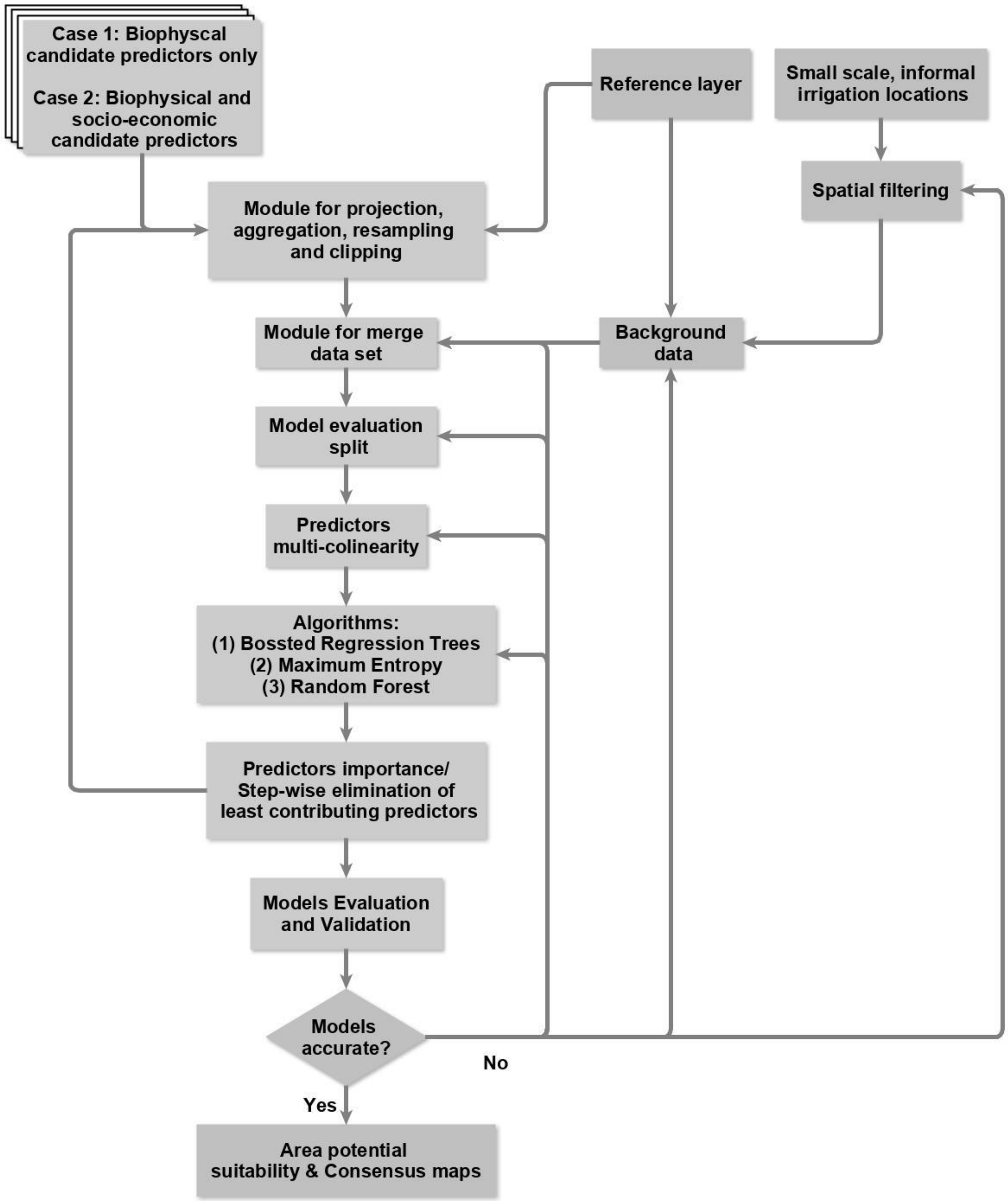

Figure 4. Modelling framework used for mapping the potential for informal small-scale irrigation.

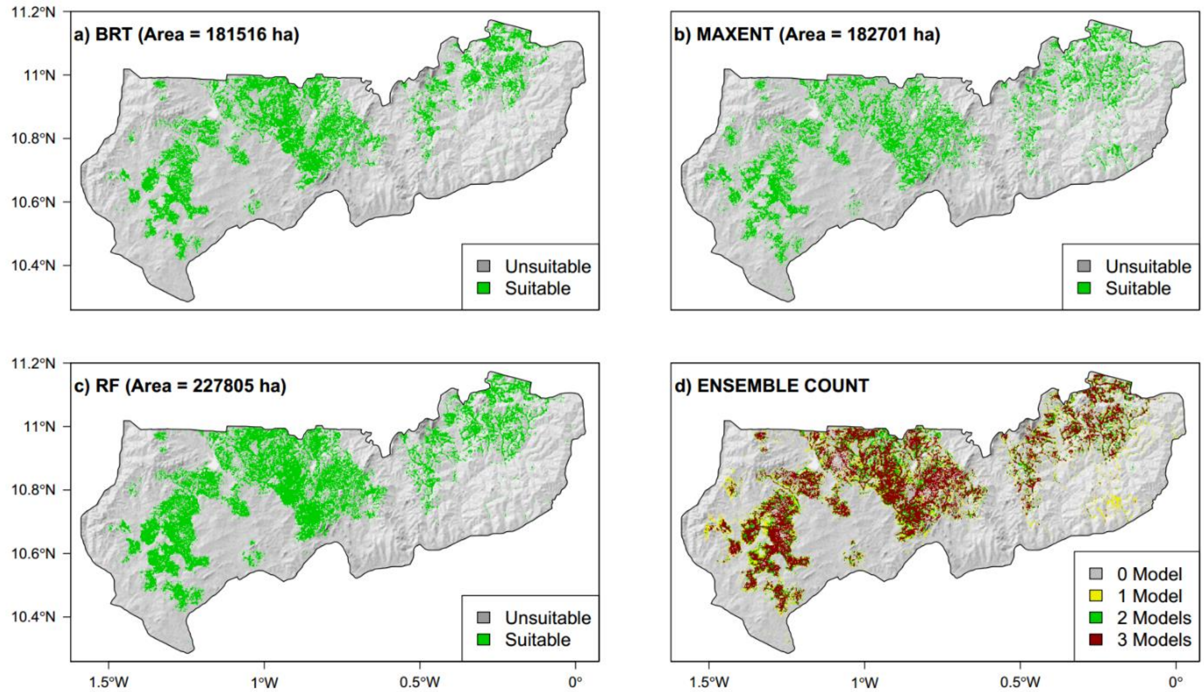

Figure 5. Binary suitability predictions using biophysical predictors only. The ensemble count shows the number of models agreeing on a location as suitable.

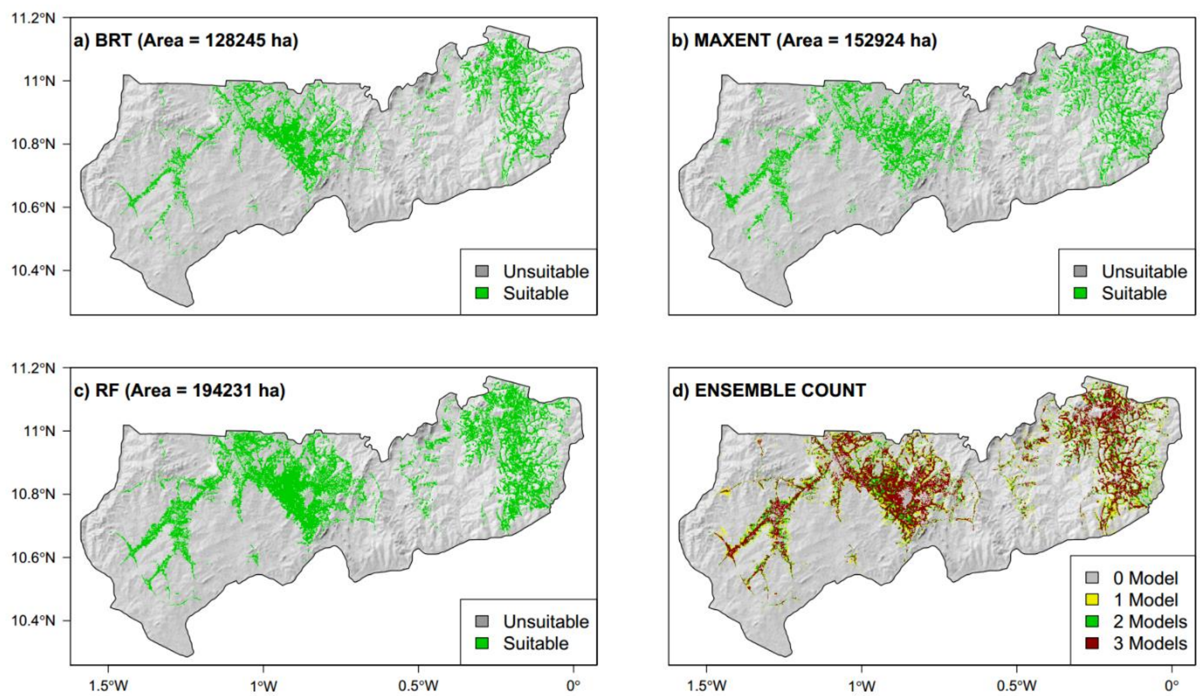

Figure 6. Binary suitability predictions using both biophysical and socio-economic predictors. The ensemble count shows the number of models agreeing on a location as suitable.

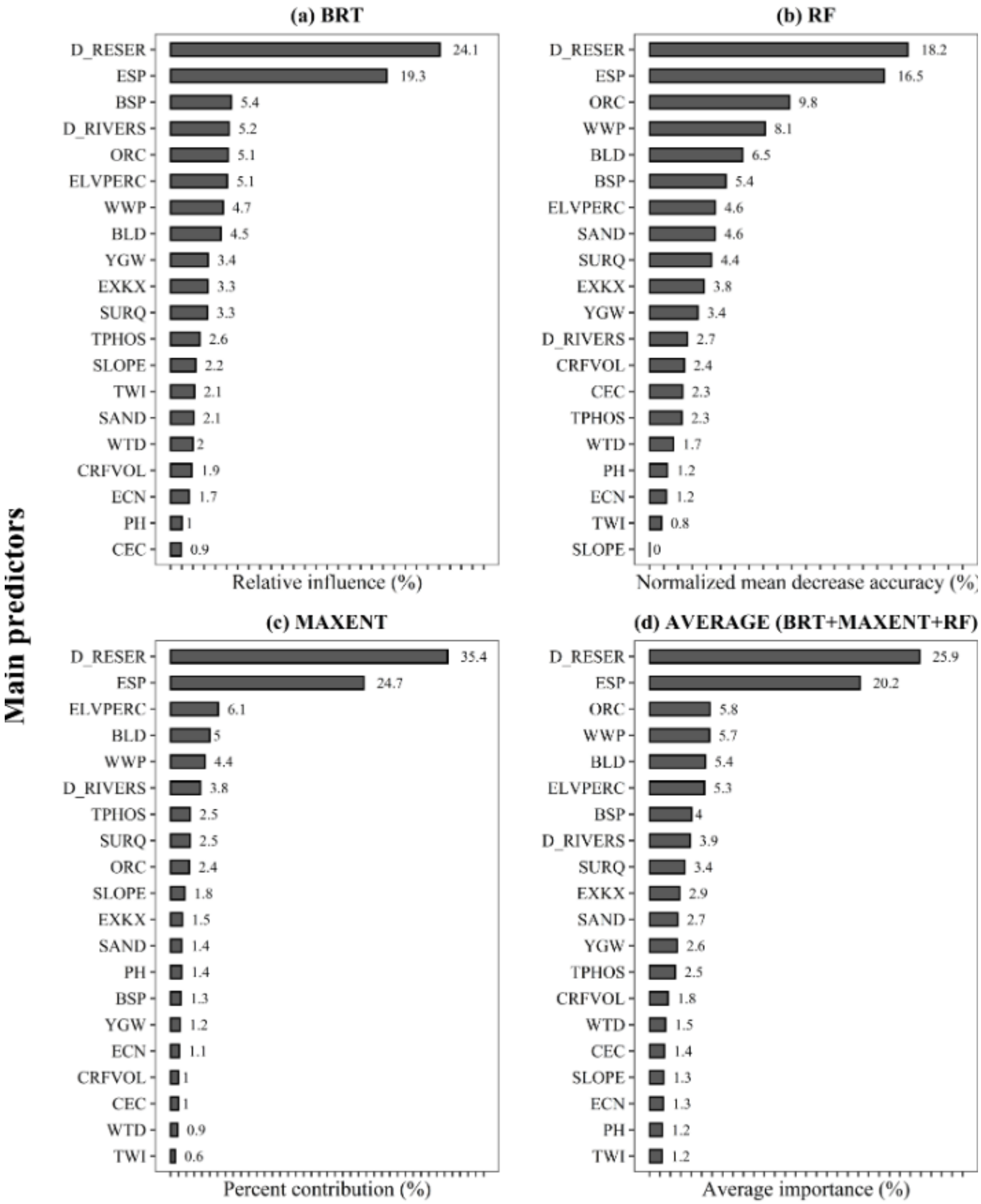

Figure 7. Predictors importance in the computation of the potential for small-scale irrigation with the biophysical predictors only. WTD—depth to water table, YGW—aquifer productivity, D_RIVERS—proximity to stream/rivers, D_RESRV—proximity to small reservoirs, SURQ—surface runoff, SLOPE—slope, ELVPERC—elevation percentile, TWI—topographical wetness index, BLD—bulk density, CRFVOL—coarse fragments volumetric fraction, WWP—available soil water capacity, SAND—soil texture fraction

sand, BSP—base saturation percentage, CEC—cation exchange capacity, ESP—exchangeable sodium percentage, EXKX—exchangeable potassium, TPHOS—total phosphorus, NTO—total nitrogen, PH—soil pH in water.

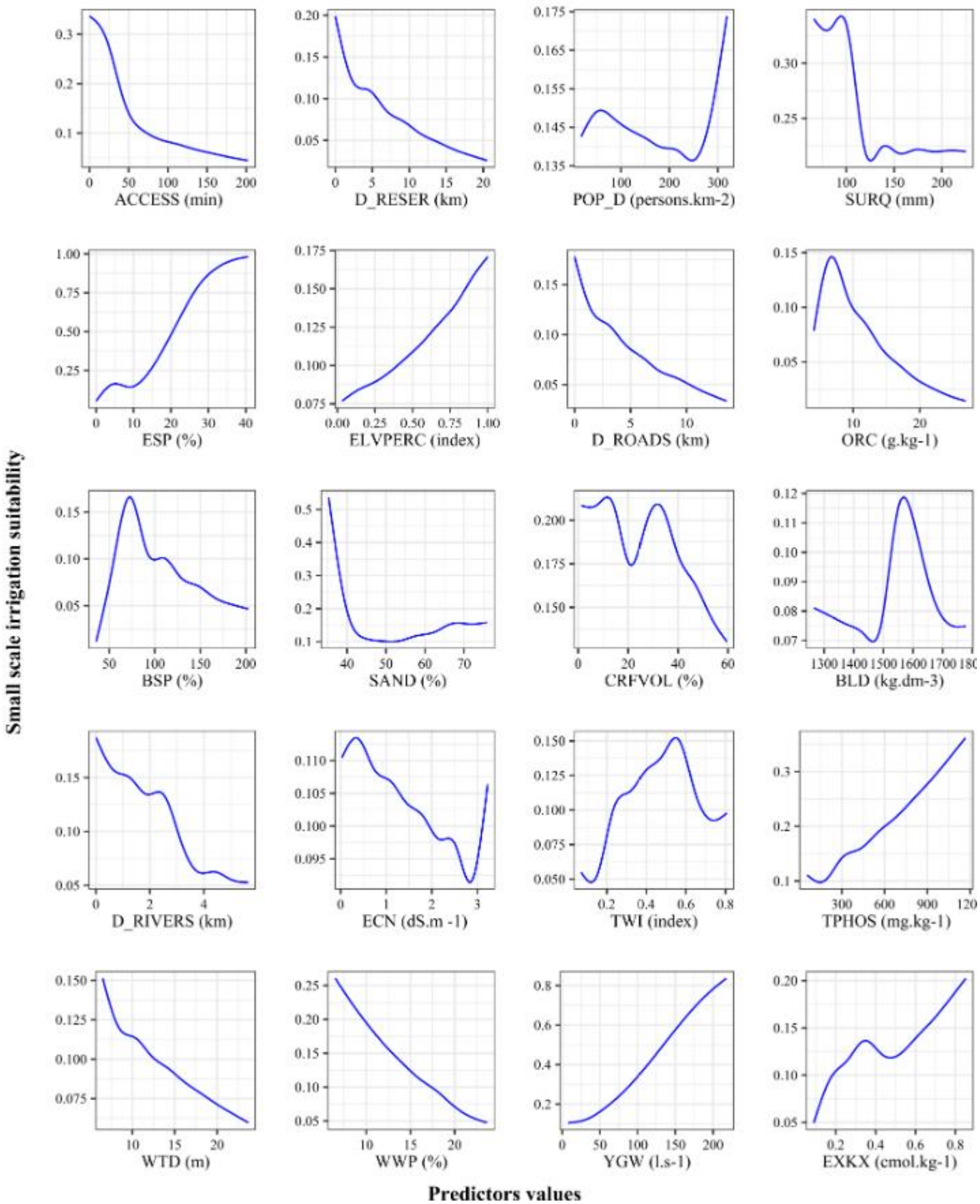

Figure 8. Response curves. D_ROADS—distance to roads, POP_D—population density, D_TOWNS—proximity to towns, WTD—depth to water table, YGW—aquifer productivity, D_RIVERS—proximity to stream/rivers, D_RESRV—proximity to small reservoirs, SURQ—surface runoff, SLOPE—slope, ELVPERC—elevation percentile, TWI—topographic wetness index, BLD—bulk density, CRFVOL—coarse fragments volumetric fraction, WWP—available soil water capacity, SAND—soil texture fraction sand, BSP—base saturation percentage, CEC—cation exchange capacity, ESP—exchangeable sodium percentage, EXKX—exchangeable potassium, TPHOS—total phosphorus, NTO—total nitrogen, PH—soil pH in water.

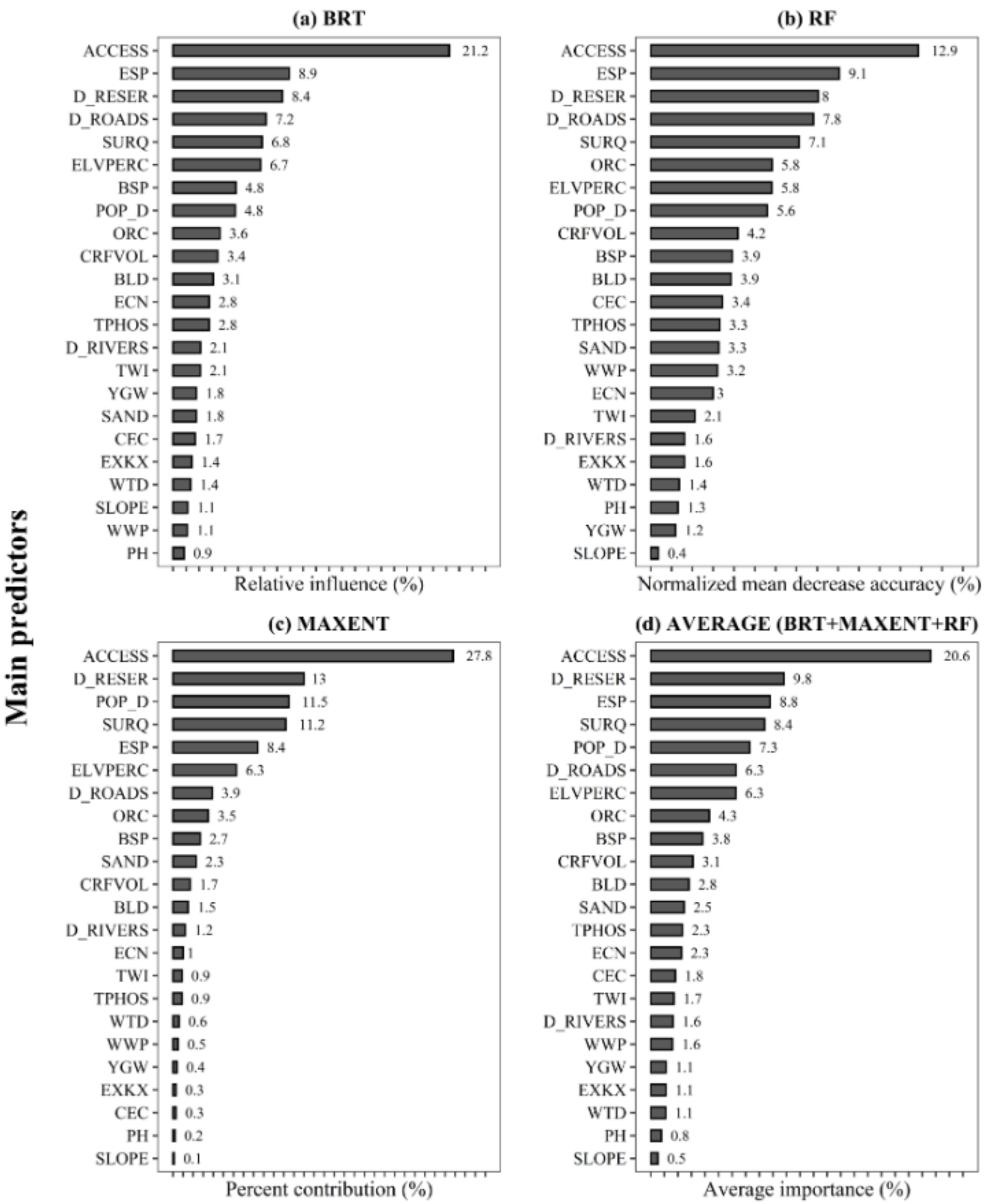

Figure 9. Predictors importance. ACCESS—accessibility or travel time to cities, D_ROADS—distance to roads, POP—population density, D_TOWNS—proximity to towns, WTD—depth to water table, YGW—aquifer productivity, D_RIVERS—proximity to stream/rivers, D_RESRV—proximity to small reservoirs, SURQ—surface runoff, SLOPE—slope, ELVPERC—elevation percentile, TWI—topographic wetness index, BLD—bulk density, CRFVOL—coarse fragments volumetric fraction, WWP—available soil water

capacity, SAND—soil texture fraction sand, BSP—base saturation percentage, CEC—cation exchange capacity, ESP—exchangeable sodium percentage, EXKX—exchangeable potassium, TPHOS—total phosphorus, NTO—total nitrogen, PH—soil pH in water.

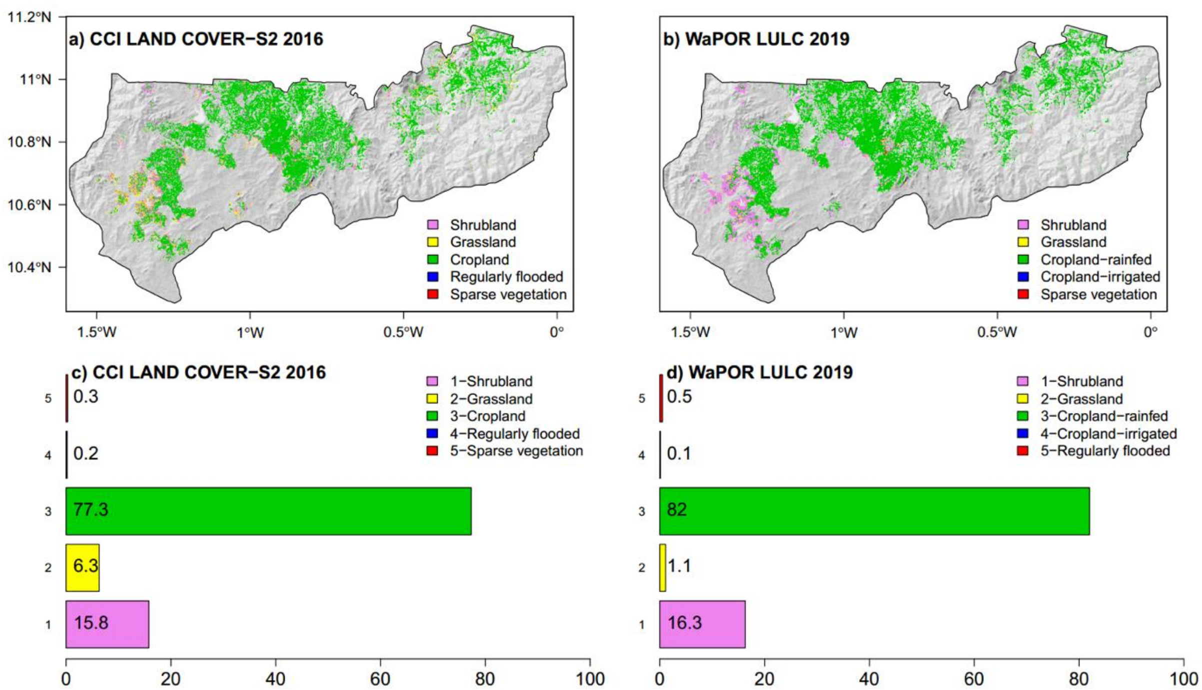

Figure 10. Comparison between the predicted potential for small-scale irrigation and existing land use land cover products when only biophysical predictors are considered in the modelling. CCI LAND COVER-S2 2016 is the land cover from CCI Land Cover (LC) team (http://2016africalandcover20m.esrin.esa.int/) and WaPOR LULC 2019 corresponds to WaPOR land cover data (https://wapor.apps.fao.org/catalog/WAPOR_2/2/L2_LCC_A). The numbers 1-5 represents the land cover

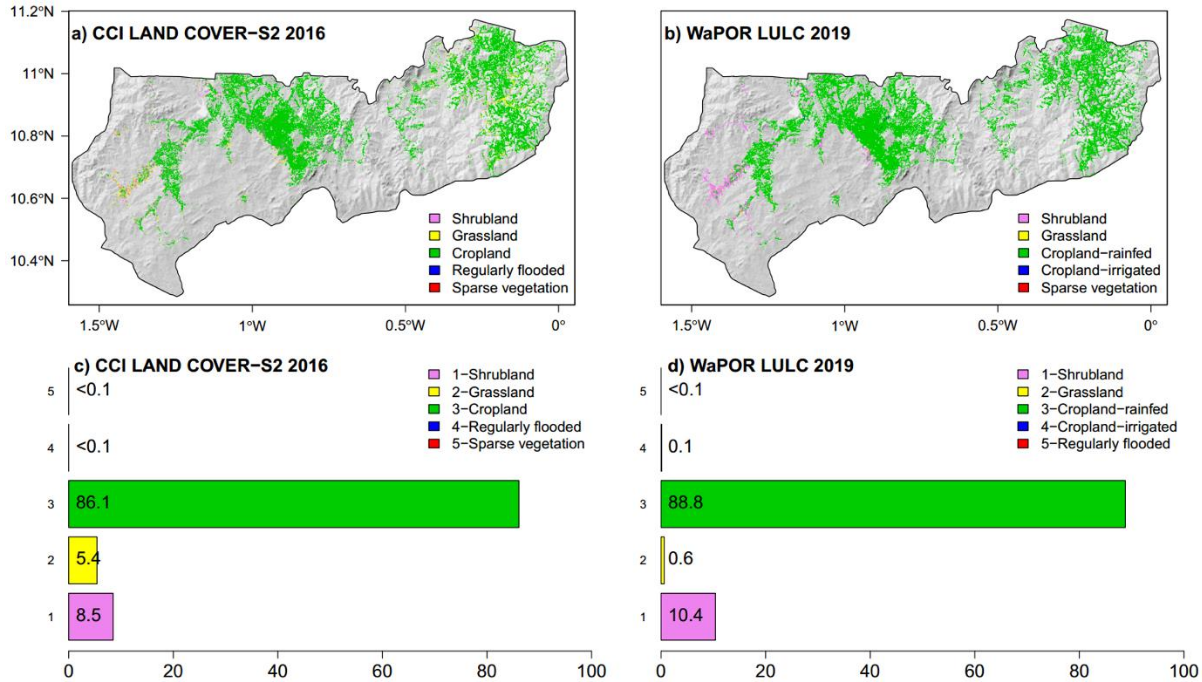

Figure 11. Comparison between the predicted potential for small-scale irrigation and existing land use land cover products when both biophysical and socio-economic predictors are considered in the modelling.

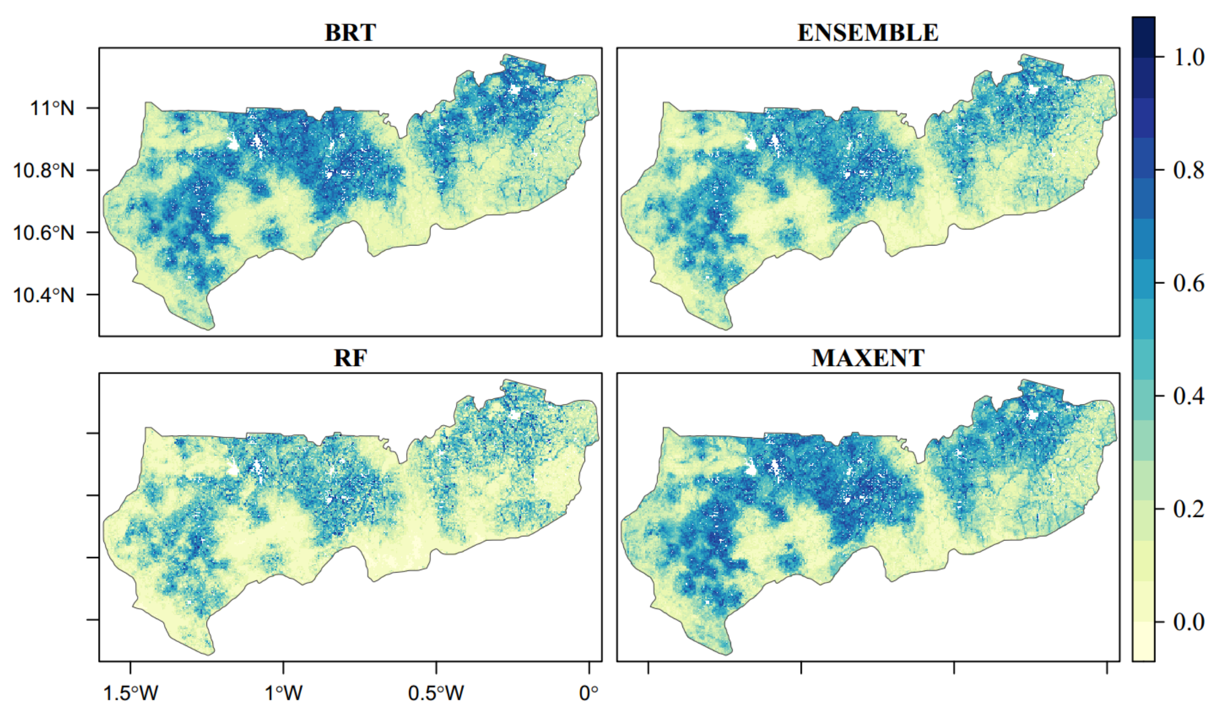

Figure A1. Continuous prediction maps with biophysical predictors only.

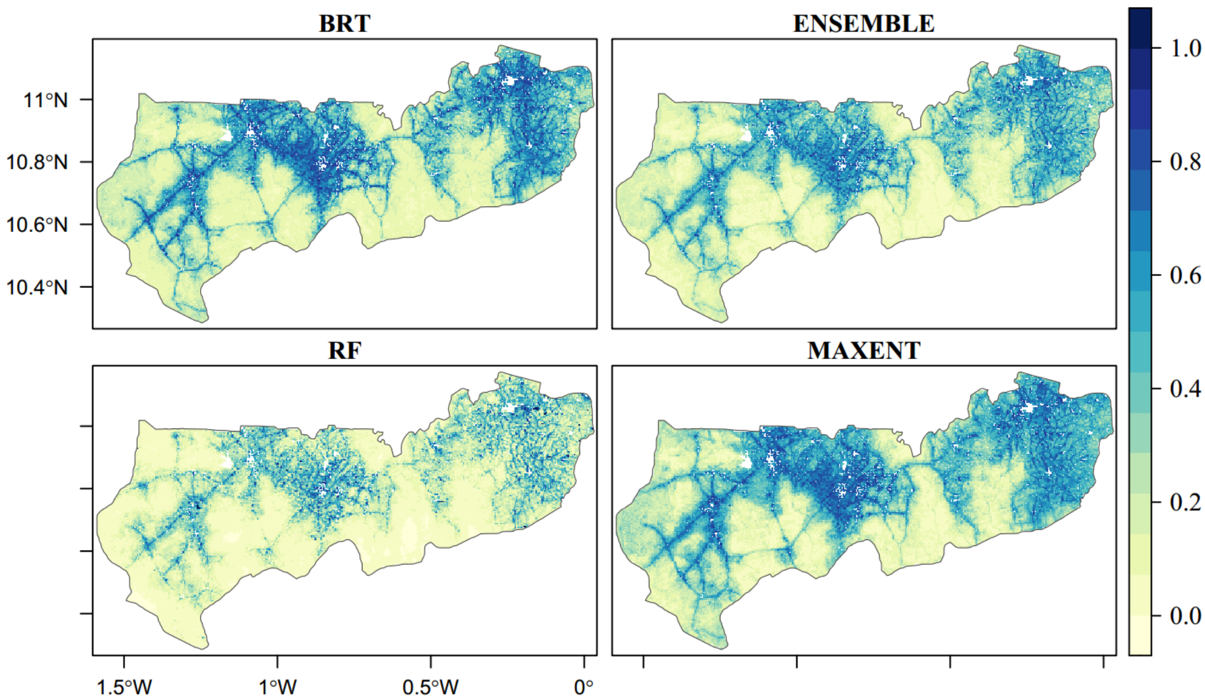

Figure A2. Continuous prediction maps with biophysical and socio-economic predictors.

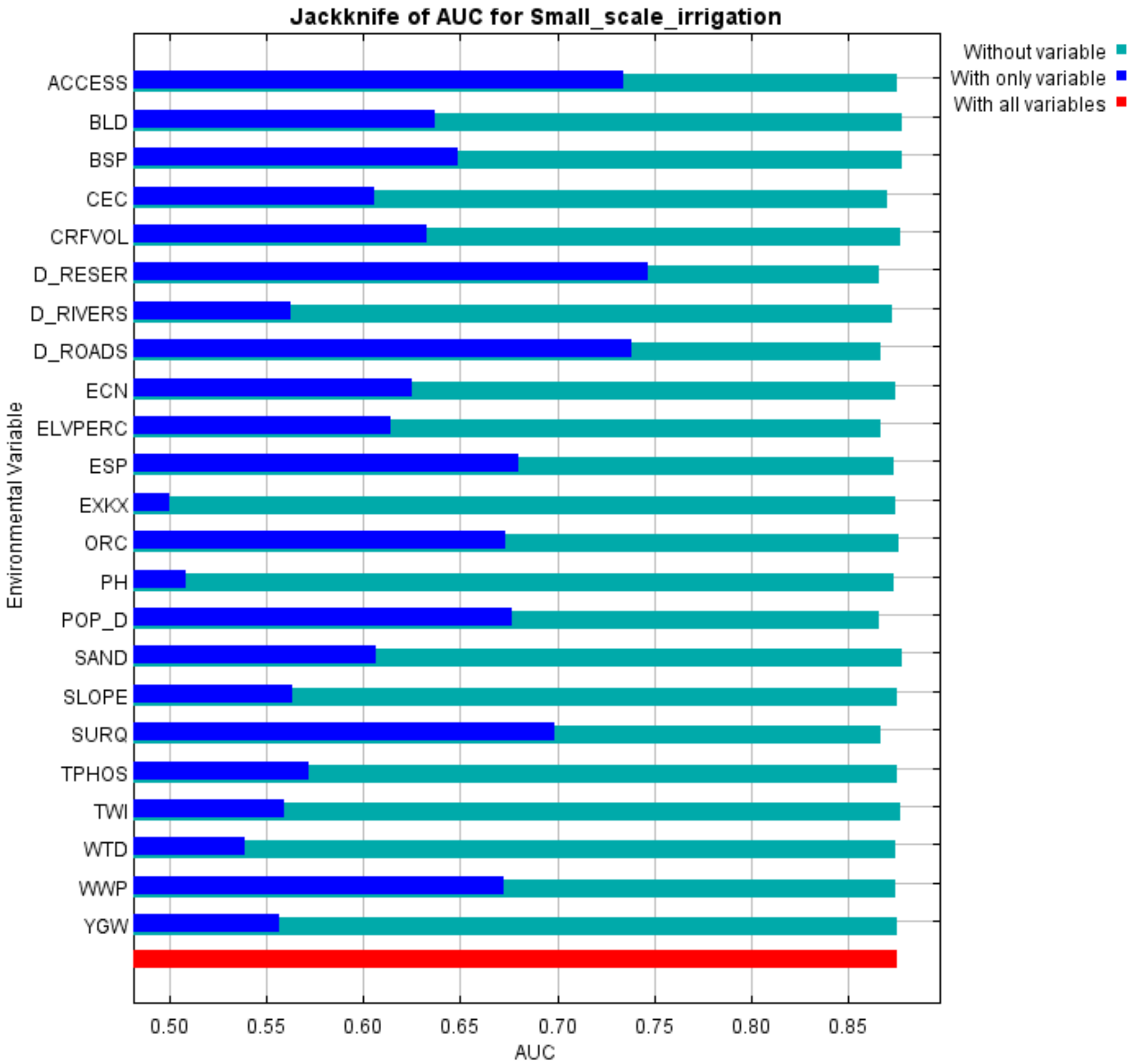

Figure A3. Jackknife test of AUC for small-scale irrigation. predictors importance for individual variable (blue bars), without variable (light blue bars) and all environmental variables (red bar) for MaxEnt model. D_ROADS—distance to roads, POP—population density, D_TOWNS—proximity to towns, WTD—depth to water table, YGW—aquifer productivity, D_RIVERS—proximity to stream/rivers, D_RESRV—proximity to small reservoirs, SURQ—surface runoff, SLOPE—slope, ELVPERC—elevation percentile, TWI—topographical wetness index, BLD—bulk density, CRFVOL—coarse fragments volumetric fraction, WWP—available soil water capacity, SAND—soil texture fraction sand, BSP—base saturation percentage, CEC—cation exchange capacity, ESP—exchangeable sodium percentage, EXKX—exchangeable potassium, TPHOS—total phosphorus, NTO—total nitrogen, PH—soil pH in water.

Credit Author Statement

- Research design (Sander J. Zwart, Komlavi Akpoti)
- Data collection and preparation (Komlavi Akpoti, Thomas P. Higginbottom, Timothy Foster, Roshan Adhikari)
- Data analysis (Komlavi Akpoti, Sander J. Zwart)
- Manuscript writing (Komlavi Akpoti)
- Manuscript review (Sander J. Zwart, Thomas P. Higginbottom, Timothy Foster, Roshan Adhikari)

Declaration of Interest Statement

The authors declare that they have no known conflict of interest.