Space-time modelling of seasonal soil moisture for improved crop production – the case of the Guinea savannah region, Ghana

DISSERTATION

to acquire the doctoral degree in mathematics and natural science

"Doctor rerum naturalium"

of the Georg-August-Universität Göttingen

within the doctoral degree program of Geoscience and Geography

of the Georg-August University School of Science (GAUSS)

submitted by

KWABENA ABREFA NKETIA

from Kumasi, Ghana

Göttingen, 2020

THESIS COMMITTEE

1. Prof. Dr. Daniela Sauer (Main supervisor)
   Department of Physical Geography, Institute of Geography, University of Göttingen.
2. Dr. Stefan Erasmi (Co-supervisor)
   Thünen-Institute of Farm Economics, Braunschweig, Germany.

REFEREES

1. Prof. Dr. Daniela Sauer (1st Referee)
   Department of Physical Geography, Institute of Geography, University of Göttingen.
2. Prof. Dr. Martin Kappas (2nd Referee)
   Department Cartography, GIS and Remote Sensing, Institute of Geography, University of Göttingen.

FURTHER MEMBERS OF THE EXAMINATION BOARD

1. Dr. Stefan Erasmi
   Thünen-Institute of Farm Economics, Braunschweig, Germany.
2. Prof. Dr. Christoph Dittrich
   Department of Human Geography, Institute of Geography, University of Göttingen.
3. Prof. Dr. Heiko Faust
   Department of Human Geography, Institute of Geography, University of Göttingen.
4. PD Dr. Steffen Herbold
   Institute of Computer Science, University of Göttingen.

Date of oral examination: 03rd July, 2020.

Space-time modelling of seasonal soil moisture for improved crop production – the case of the Guinea savannah region, Ghana

Photo: Author

To my family

"THE SOIL COMES FIRST – It is the basis, the foundation of farming, without it nothing; with poor soil, poor farming, poor living; with good soil, good farming and living. An understanding of good farming begins with an understanding of the soil"

Henry L. Ahlgren (EMER. Prof.)

Acknowledgements

I express my deepest gratitude to my supervisor, Prof. Dr. Daniela Sauer, who agreed to supervise this thesis that is geared towards improving the Ghanaian agriculture. Not only did she provide research advice, assistance, encouragement and mentorship, but also, she gave me room to be innovative. Prof. Dr. Sauer has immensely contributed to shaping my career path. I am also grateful to my co-supervisor, Dr. Stefan Erasmi, for his contributions and valuable comments for this thesis. Prof. Dr. Sauer and Dr. Erasmi have both left an indelible mark in my career path. "*Meda mo ase*".

I would also like to thank my thesis committee: Prof. Dr. Martin Kappas, Prof. Dr. Christoph Dittrich, Prof. Dr. Heiko Faust and PD Dr. Steffen Herbold for their key suggestions and insightful contributions for this thesis.

I deeply express my appreciation to the German Academic Exchange Service (DAAD) and the Ministry of Education, Ghana, for the PhD scholarship award granted me. I thank the Georg-August University School of Science (GAUSS), Department of Physical Geography of the University of Göttingen, and Council for Scientific and Industrial Research – Soil Research Institute, Ghana, for financial support for the field campaign of my research tasks. In addition, I owe my earnest thanks to Prof. Joseph Opoku Fening, Eric Asamoah, Samson Owusu, Alex Owusu Ansah and Anim Boafo, and the Ministry of Food and Agriculture, Ghana, for field assistance. I thank the German Research Foundation and the Open Access Publication Funds of the University of Göttingen for funding the publication of manuscripts of this thesis.

My thanks go to all my colleagues and staff of the Department of Physical Geography of the University of Göttingen for their contribution in my academic path. My sincere thanks go to Stephen Asabere, who has been more than a brother to me during these past years. Stephen, I cannot thank you enough! Ms Hesse, thank you for your support these past years.

Most importantly, this thesis would not have been a success without the love, care and prayers of my family: wife – Jacinta, kids – Nhyira and Awura-Ama, parents – Kwasi Osei and Adwoa Pokuaa, and siblings – Kwame, Agyemang, Poku and Ida. You have all brought me this far. No words can express how appreciative and grateful I am to have all of you. *"Nyame nhyira mo".*

List of abbreviations

API	Antecedent precipitation index
cLHC	Cost-constrained conditioned Latin hypercube algorithm
DEM	Digital elevation model
DSM	Digital soil mapping
EBM	Ensemble-based model
ET_o	Evapotranspiration
GWPCA	Global weighted principal component analysis
LS	Local slope
MLA	Machine learning algorithm
PAWC	Plant available water content
RZ-SWS	Root-zone soil water storage
SM	Soil moisture
SM_g	Gravimetric soil moisture
SM_p	PR2/60 handheld probe soil moisture
SM_e	Model estimated soil moisture
SSA	sub-Saharan Africa
SWS	Soil water storage
TWI	Topographic wetness index
RF-RFE	Random-forest recursive feature elimination algorithm
LLTO-CV	Leave location time out cross validation

Table of Contents

List of Tables

Section A

Chapter 3

Section B

Chapter 4 (Study 1)

Chapter 5 (Study 2)

Chapter 6 (Study 3)

Chapter 7 (Study 4)

List of Figures

Section A

Chapter 1

Chapter 2

Section B

Chapter 4 (Study 1)

Chapter 5 (Study 2)

Chapter 6 (Study 3)

Chapter 7 (Study 4)

Summary

The upsurge in advocacy for food security in SSA implies the urgent need for improved sustainable adaptation measures that can boost food-crop production. This is of utmost concern, because, over the past decades, food security targets in SSA have remained unmet due to food-crop production limitations. One of the key adaptation measures which is identified to address these food-crop production limitations has been the urgent call for DSM. DSM is essential to address concerns on site-specific soil information that guides fertilizer application, improve data availability of soil fertility parameters and fill the gaps of spatially explicit soil maps. SSA is one of the regions in the world with a large terra incognita ahead of its DSM initiatives. Thus far, there is a paucity of data omission on seasonal SM and SWS, which reflect the size of water reservoir of agricultural soils and its water storage adequacy. It is imperative that such critical soil information is made available. This is because, studies have demonstrated that in rain-fed agriculture, which dominates the agricultural landscape of SSA, ~50% of total crop yield loss can be implicated by weather-induced water stress. Already, studies are reporting declining crop yields due to water-deficit conditions. Against this backdrop, meeting the globally-projected 60% increase in food demand by 2050, of which SSA is deemed to play a pivotal role, is recognized as a major challenge.

This thesis fills the knowledge gap by employing state of the art approaches on spatio-temporal scale analyses in order to complement existing DSM initiatives, which guide sustainable agriculture, crop intensification, modelling agricultural systems and site-specific farm management recommendations. We investigated and modelled the spatio-temporal seasonal SM and SWS of arable benchmark soils of the Guinea savannah zone of Ghana. The Guinea savannah zone is of importance because, it is a key reminiscent of the arable landscapes of SSA. To make our findings useful to the SSA region, this research specifically targeted smallholder farming communities, as they constitute ~80% of the farmers in the region. Also, these farmers have farm sizes < 1 ha that can easily adopt improved management practices. Here, we selected major arable benchmark soils along three main soil toposequences of the Guinea savannah zone. Specific objectives undertaken to fill the knowledge gap were to: (1) design a new soil sampling stratification that adequately represented the soil toposequences, defined local structures and accounted for localized spatial autocorrelation in explaining SM and SWS variability, (2) analyze and assess the spatio-

temporal dynamics of SM of soils of the area, (3) investigate the potentials for using high-spatial and -temporal resolution remote sensing images to estimate SM at detailed scale and (4) functionally map, at 100 m spatial resolution, the four-dimensional root zone SWS of soils of the Guinea savannah zone.

Addressing these specific objectives, key implications that can improve food-crop production, especially for the Guinea savannah zone, are recommended. Firstly, SM and SWS in the shallow soil depths (≤ 15 cm) were highly variable, unstable and consistently dry as compared to the bottom soil layers (≥ 20 cm). These observed high temporal instability were as a result of, on the one hand, the influence of internal soil factors such as clay and silt contents, and bulk density, and on the other hand, external factors such as slope, precipitation and evapotranspiration. In the bottom layer soils, clay content increased with increasing soil depth which kept SM and SWS for longer periods by promoting time-stable wet cluster of locations. Secondly, time-stable locations where crop water requirements can be met during crop growing periods is explicitly identified for use. An outcome of this research is that almost all benchmark soils of the Guinea savannah zone (except for the Kumayili series) have SWS potentials that match the water requirements of at least some drought-tolerant crops of the area. Furthermore, we found that the use of high spatial resolution multi-temporal radar and optical remote sensing images opens new perspective to estimate and adequately understand the spatio-temporal variabilities of SM in sparse *in situ* measurements network. This finding brings the advantage over existing SM and SWS point-based analysis and also improves the use of SM and SWS information in semi-arid farming landscapes. Also, the estimation of SM at detailed spatio-temporal global scales while preserving a short revisit time is possible.

Through this thesis, we connected several scales of analyses and initiatives regarding an improved food-crop production system in SSA. Possible adoptable recommendations drawn from this thesis include, e.g., the possibility and practicability to either prolong the existing major single farming window and the identification of locations and durations where additional crop-specific farming is applicable. In addition, the outcomes of the thesis can be used to enhance the adaptive capacity of smallholder farmers to increase food-crop production, yields, income and diversify livelihood alternatives of the local farming communities. Therefore, the findings from this thesis forms a core support system that is necessary to guide the implementation of drought-adaptation measures, dual farming system and complement existing DSM initiatives around the world.

Section A: Thesis overview

Chapter 1: Introduction

1.1 General background and justification

It is widely known that the agricultural sector continues to play a pivotal role in many national economies of sub-Sharan Africa (SSA) – either ensuring food security or providing other livelihood alternatives for the region (Xie et al. 2018; AGRA 2017). Although SSA has over 50% of the world's arable lands (Hengl et al. 2017b), improved food-crop production is highly challenged by poor adoption of integrated soil fertility management principles (Guilpart et al. 2017) and inherent low soil fertility (Bado and Bationo 2018). Despite strong advocacy to boost crop production, these production limitations still linger and are far from reaching an improved state (AGRA 2017). Consequently, meeting the globally-projected 60% increase in food demand by 2050, of which SSA is deemed to play a critical role, is recognized as a major challenge (Godfray et al. 2010). Also, across SSA and other parts of the globe, there is an increase in advocacy efforts geared towards underscoring the importance of food security targets – the key being the policy on sustainable development goals (Griggs et al. 2013). However, external factors such as limited site-specific soil information, upsurge in land grabbing earmarked for agriculture due to population increase and price volatility of crops continue to undermine the food security targets (van Ittersum et al. 2016; Koning et al. 2008). Nevertheless, future climate projections which indicates ~40% reduction in rainfall and an increase in severe drought conditions will exacerbate the challenges to reaching the targets of food security (Howell 2001). Although the implementation of existing adaptation measures such as irrigation schemes could address some of these challenges, only 3% supplemental irrigation is currently being practiced in SSA, due to high cost of operation and implementation (van Ittersum et al. 2016; Adeboye et al. 2017). Hence, farmers adaptive capacities that could couple the rainy and dry season farming windows to boost food-crop production is largely decreasing (Armah et al. 2011; Adolwa et al. 2017). These backdrops coupled with the teeming human population further exacerbate the uncertainties associated with rainfed agriculture and food demand in SSA (Fisher et al. 2015; Xie et al. 2018). Thus, it places SSA in a highly vulnerable state and limits the prospects for its economic development.

On the global scale, these crop production limitations and its adverse effects appear enormous, especially in SSA. However, studies have rather suggested to meet the global food demands at the national level (Foley et al. 2011; Tilman et al. 2011; Pradhan et al. 2015).

Consequently, adopting measures that can close actual and potential yield gaps of existing smallholder farming communities would be a key step to addressing these global constraints. This is mainly because, e.g. in SSA alone, smallholder farmers make up ~80% of the population and have farm sizes < 1 ha, which can easily adopt improved management practices (Lowder et al. 2016; Leenaars et al. 2018). It is therefore highly important to curtail food-crop production limitations at the smallholder farming scales in order to meet food security targets and limit vulnerabilities of SSA.

Recently, the number of adaptation options in SSA have increased. For instance, a number of African countries have accented to the Abuja Declaration on fertilizer intended to tackle problems of soil fertility. While other key adaptation option has been the recent flagship projects of SSA, i.e. the Africa Soil Information Service project (AfSIS; http://africasoils.net/), the West African Agricultural Productivity Program (WAAPP; https://www.waapp-ppaao.org) and the Global Yield Gap Atlas project (GYGA; http://www.yieldgap.org/), which address site-specific soil information gaps and accelerate the adoption of improved farming technologies. Despite the importance and contributions of these improved food-crop production initiatives in SSA, one thing that is clear is the paucity of data on seasonal soil moisture (SM), the size of water reservoir of agricultural soils and its water storage adequacy (Vågen et al. 2010; Dewitte et al. 2013; Hengl et al. 2017b; Leenaars et al. 2018). The missing data on root zone SM could be detrimental to the attainment of food security targets in SSA. This is because, in Rossato et al. (2017) and IPCC (2007), it was reported that in rain-fed agriculture ~50% of total crop yield loss can be implicated by weather-induced water stress, which suggests a strong contribution to the yield losses recorded over the past decades in SSA (Adeboye et al. 2017; Leenaars et al. 2018). Thus far, no regional initiative or study has developed spatial-temporal information on SM and soil water storage (SWS) of root zone of soils to address water requirements of crops in SSA, especially for the water deficit-landscapes (Chivenge et al. 2015; Leenaars et al. 2018). This observation raises farmers concern on where (spatial location), when (temporal), quantity (requirements) and stability of seasonal SM and SWS of the root zone of soils to support crop production and, also to inform other on-farm water management schemes (Dai et al. 2004; Armah et al. 2011; Chivenge et al. 2015). Furthermore, there is little evidence in SSA about the spatio-temporal variabilities of SM and SWS, how its occurrence is being influenced by soil-environmental factors, and how its potentials can be tapped to boost food-crop production. Indeed, rain-fed agriculture would be severely hit in a water-stressed and nutrient deficient SSA. Already, studies are demonstrating and reporting declining crop yields and short falls due to water-deficit conditions (Jayne et al. 2010; Hengl et al. 2017b). In addition, there is decrease in dry season farming practices due to limited site-specific information, which can supplement yield losses and production deficits of the main rainy cropping season (Armah et al. 2011). Given that rainfed agriculture will continue to play a pivotal role in the foreseeable future of SSA, the availability of such critical site-specific soil information will be an important contribution to support other on-going adaptation options, which thrive to improve food-crop production.

The case of the Guinea savannah is indifferent to those of SSA and it is of utmost concern in terms of food security targets and smallholder livelihood empowerment (Armah et al. 2011; Adolwa et al. 2017; Antwi-Agyei et al. 2018; Williams et al. 2018). This is mainly because, not only does it has a vast prospect for commercial agriculture in Africa, but also, it has accounted for about 60% of the total value of agricultural outputs for the past two decades (Morris et al. 2009; OECD 2016). In addition, the Guinea savannah region is the breadbasket of many arable lands of SSA. It is therefore imperative to enhance existing adaptation options such as digital soil mapping initiatives, which can adequately support the cultivation of drought tolerant crops and the implementation of sustainable dual-farming systems in SSA.

In this thesis, we focused on bridging the knowledge gap on the missing data on the size of water reservoir of the root zone of agricultural soils and its water storage adequacy. For this purpose, we develop the first spatio-temporally explicit seasonal root zone SM and SWS information of soils of the Guinea savannah zone, which can guide site-specific sustainable agriculture, crop intensification and the modelling of agricultural systems. This thesis brings an additional advantage of an improved site-specific soil information package for arable soils, which forms part of the needed adaptation options identified to partly address food-crop production limitations in SSA. We used spatio-temporal scale analyses coupled with empirical and secondary datasets, which bring an advantage over point-based analysis and the practicability of understanding seasonal SM and SWS trends in water deficit landscapes. Due to large amounts of small farm sizes in the study area, which is reminiscent of the general farm size of SSA (Lowder et al. 2016), we use high spatial resolution multi-temporal radar and optical remote sensing images to infer secondary datasets for use in our analyses. The use of remote sensing images also addresses the limitation of sparse *in situ* measurements network, which is a challenge of adequate data representation of SM statuses at the regional scale. As a preliminary test of methodologies and applications, we focused on the Guinea savannah semi-arid zone of Ghana, which is suggestive of the arable landscapes of SSA and its agro-ecological zones (Figure 1.1) (Sebastian 2009).

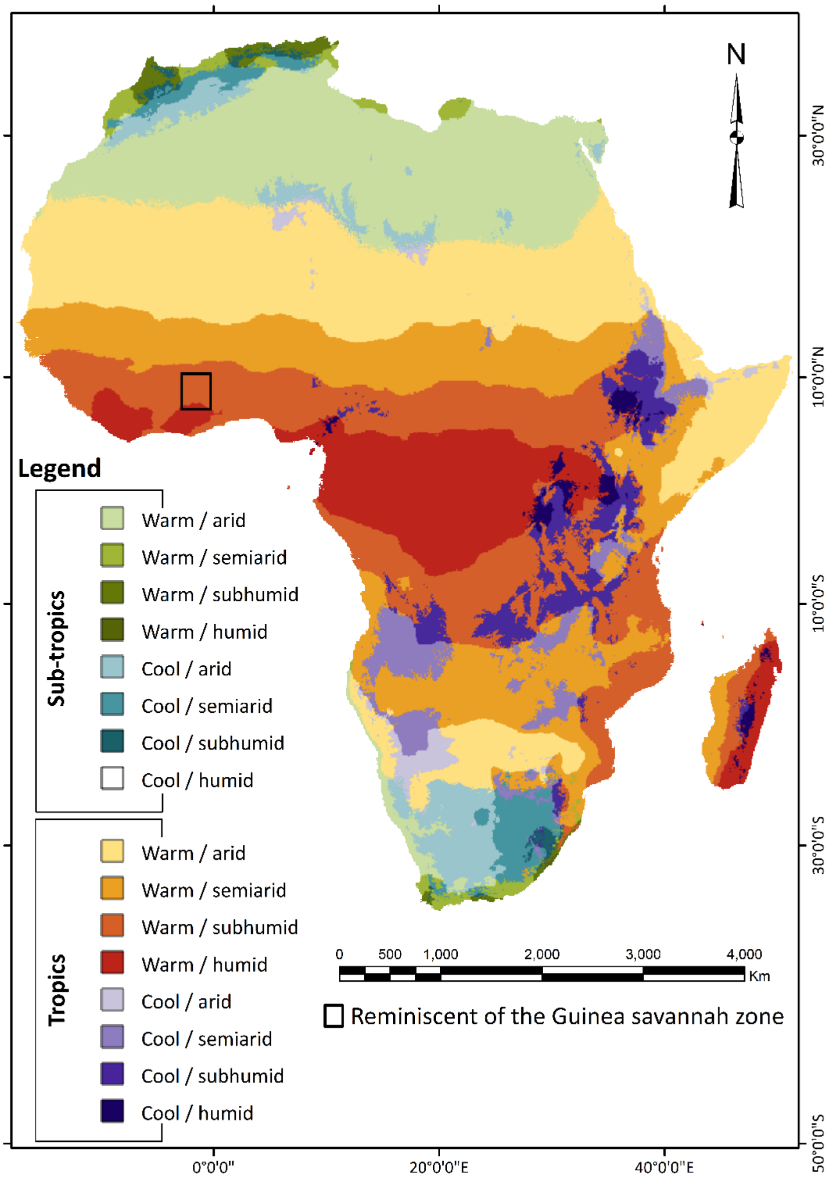


Figure 1.1. Map showing the extent of the Guinea savannah zone (i.e. covers the arid, semi-arid, sub-humid and part of the humid tropics) of SSA. Modified from Sebastian (2009).

1.2 Research objectives

The main goal of this thesis was to investigate and model the spatio-temporal seasonal SM and SWS of benchmark soils of the Guinea savannah zone of Ghana. In this context, two specific objectives were undertaken to;

1. Investigate the spatio-temporal variability of root zone SM of soils and how it is being influenced by soil-environmental factors. This objective was achieved by two main research tasks:
    a. develop a new soil sampling stratification approach that adequately represent the soil toposequences, define local structures and account for localized spatio-temporal autocorrelation in explaining SM variability (**Study 1; Chapter 4**).
    b. analyze and assess the spatial and temporal dynamics of SM of key benchmark soils of the Guinea savannah zone of Ghana (**Study 2; Chapter 5**).
2. Explicitly model surface SM and map the four-dimensional (i.e. space, soil depth and time) distribution of root zone SWS. Two main research tasks were used to achieve this objective:
    a. investigate the potentials of using high-spatial and -temporal resolution remote sensing imagery to estimate SM at fine spatio-temporal global scale (**Study 3; Chapter 6**).
    b. assimilate and functionally map, at 100 m spatial resolution, the four-dimensional root zone SWS potentials of the Guinea savannah zone (**Study 4; Chapter 7**).

Chapter 2: Study area and methods

2.1 Overview of agriculture in Ghana

Similar to most economies of SSA, the agricultural sector is one of the mainstays of Ghana's economy and employs over 50% of the working population (World Bank 2017b). It contributes about 20–25% of Ghana's gross domestic product (World Bank 2017b). The agricultural sector has been estimated to grow, in the medium term, at an average of ~3.3% yearly (Ghana Ministry of Finance 2017), as against the proposed growth rate of 6–8% (Banson et al. 2016). In Ghana, agricultural outputs vary with agro-ecological zones, however, smallholder farmers (~80% of the total agriculture chain) are discernable throughout the arable landscapes as compared to those involved in commercial agriculture. It is these smallholder farmers that are also noted for traditional and rain-fed agriculture, and the low use of improved agricultural technologies (Chamberlin 2008). For example, only 0.2% of the 42.4% arable lands are under irrigation (World Bank 2017a), and thus, makes improved crop production a major challenge throughout the year.

Tree and cash crops, i.e., cocoa (*Cacao*), oil palm (*Elaeis guineensis*), tobacco (*Nicotiana tabacum*) and rubber (*Ficus elastica*), are significant in the southern part of Ghana. Whereas, in the middle to the northern belt, food crops are significant with maize (Zea mays), legume (*Leguminosae*), plantain (*Musa paradisiaca*), rice (*Oryza glaberrima*), cocoyam (*Xanthosoma sagittifolium*), cassava (*Manihot esculenta*) and yam (*Dioscorea abyssinica*) being of particular importance. Specifically, the northern sector is also well-known for cotton (*Gossypium*), groundnut (*Arachis hypogaea*), sorghum (*Sorghum bicolor*) and millet (*Panicum miliaceum*). Rice is cultivated throughout the country. Amongst these cultivated food crops, maize and rice are the largest-staple caloric intake in Ghana (SRID 2016). Farming practices are diverse, with the prominent being inter-cropping due the large smallholder farming communities (Adolwa et al. 2017).

Within the agro-ecological zones of Ghana, the Guinea savannah zone, which is categorized as the breadbasket of Ghana (Bourne Jr 2014), is the dominant. On the one hand, it is within this region where most farming activities are undertaken, and on the other hand, agricultural challenges are very prominent in the area and beset crop production potentials (Acheampong et al. 2014; Amekudzi et al. 2015). The main challenge is the adverse effect of the single farming window that limits crop production (Abass 2015). For instance, Ghana currently stands at 8–10 t ha^{-1} (maize), 40 t ha^{-1} (cassava), 6–8 t ha^{-1} (rice) and 2.5 t ha^{-1} (soybean), compared to the overall crop yield potentials of SSA and the world (Alidu et al. 2016; World Bank 2017b). Some other significant constraints to sustainable agricultural production in Ghana are low incorporation of integrated soil fertility management technologies, low rainfall and nutrient depletion. Currently, nutrient depletion rates of 35 kg N ha^{-1} yr^{-1}, 4 kg P ha^{-1} $yr^{-1}$and 20 kg K ha^{-1} yr^{-1} have been reported (Fening 2016). It is in this light that Ghana continues to miss its food security targets.

Recognizing these facts and to achieve self-sufficiency in food-crop production, the Government of Ghana is putting in adaptation measures to spur economic growth in order to curtail the challenges faced by smallholder farmers. Instituted adaptation measures include the E-Agricultural initiative, the Medium-Term Agricultural Sector Investment Plan and the Planting for Food and Jobs Program (Asuming-Brempong 2013; Mohammed Tanko et al. 2019). These recent flagship initiatives are geared towards boosting food-crop production, food security and reduction of food imports (Mohammed Tanko et al. 2019). Within these policies, farm inputs and adaption measures outlined includes access to updated site-specific soil information, improved seed, irrigation schemes, fertilizer and sustainable farming practices.

It is against this backdrop that this thesis complements ongoing adaptation efforts (i.e. enhancing site-specific soil information) to address existing crop production challenges in the Ghanaian agriculture chain, especially for the Guinea savannah zone of Ghana.

2.2 Characteristics of the study area

All research tasks of this thesis, i.e. study 1–4, were conducted at the same study area of the Guinea savannah zone and followed a single sampling approach. Hence, I only report the settings of the study area, sampling design and *in situ* measurements in these subsequent sections of the introduction and not with the manuscripts of this thesis.

The Guinea savannah zone experiences erratic rainfall patterns, which is characterized by high evaporation loss and low rate of runoff conservation (Acheampong et al. 2014). For this thesis, we selected the Tamale enclave (Figure 2.1) due to large farming communities, suitable arable soils for crop production and good accessibility. The study extent covered a 150 x 80 km area between Daboya (westward), Sang (eastward), Gaa (northward) and Fufulsu (southward).

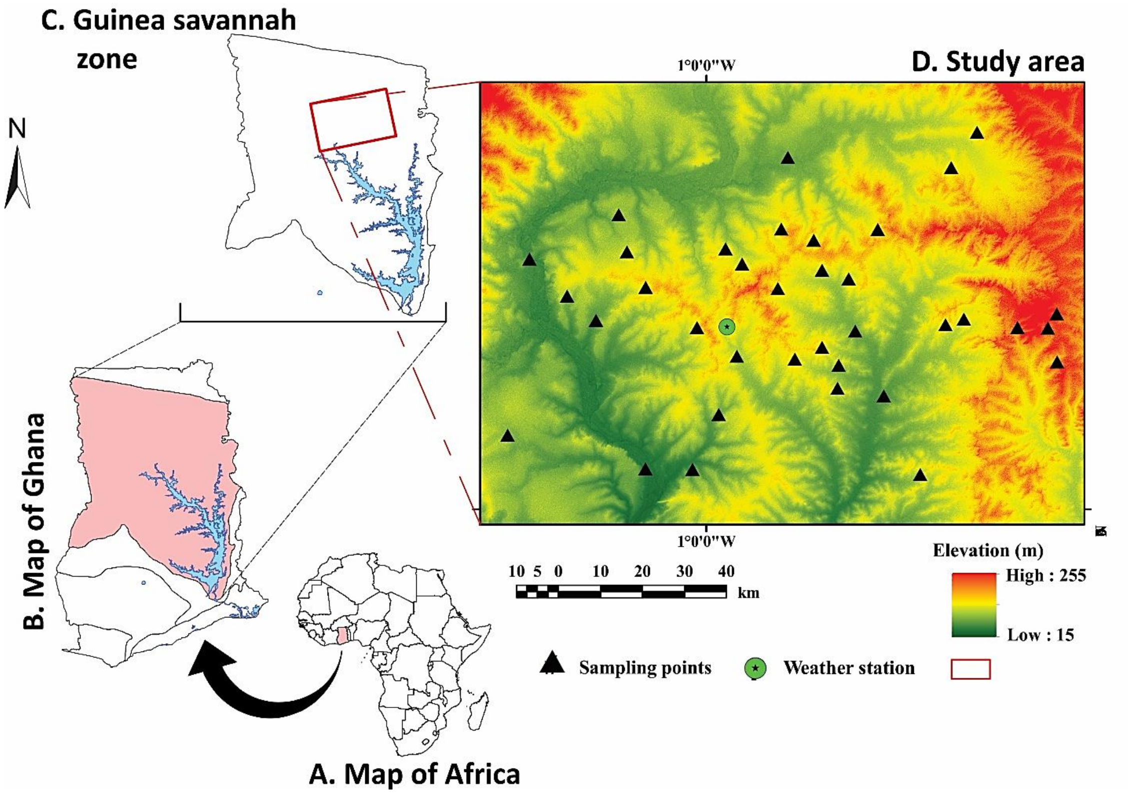


Figure 2.1. Location maps showing the Guinea savannah zone (B and C) and in situ measurement locations (D) overlaid on the digital elevation model of the study area.

2.2.1 Climatic conditions

The Guinea savannah zone experiences a semi-arid climatic condition with mean annual precipitation varying from 900 mm to 1100 mm (1985–2018). Compared to the defined dry season (255 ± 5 days; September–March), the rainy season is shorter (110 ± 5 days; April–August) (Amekudzi et al. 2015) and it is within this period where all major farming activities are undertaken. Daily, relative humidity varies between 18% and 97% while temperature vary from a minimum of 32 °C in August to a maximum of 43 °C in March. Evapotranspiration of the Guinea savannah zone is ~3-times the daily average precipitation of the area (range = 4.9 mm day^{-1}). Retrospective and current climatic information reported in this thesis was obtained from a weather station located in the study area (Figure 2.1D).

2.2.2 Land use/Land cover

Elephant grass (*Pennisetum purpureum*) dominates the vegetation types of the Guinea savannah zone with few interspersed occurrences of *Borassus aethiopum* (Borassus palm), *Khaya senegalensis* (Senegal mahogany) and *Vitellaria paradoxa* (Shea tree). Due to large river networks and low vegetation cover, flooding events are rampant especially in the rainy season. In the Guinea savannah, notable food-crops cultivated includes *Oryza glaberrima* (upland rice), *Zea mays* (maize), *Solenostemon rotundifolius Poir* (potatoes), *Pennisetum americanum* (millet), *Vigna unguiculata* (cowpea), *Dioscorea* (yam) and *Manihot esculenta* (cassava). Other land uses of the area includes irrigated vegetable cropping, dryland farming and pastures.

2.2.3 Geomorphology, geology and soils

While ploughing activities from smallholder farming systems mainly drives the surface roughness of soils of the area, the topography and soil surface is generally flat and relatively homogenous. The highest altitude of the study area varies between 255 m a.s.l. to 39 m a.s.l. with a gentle topography with slope inclination ranging from 0.5% (minimum) to 5% (maximum). *In situ* measurements within the 150 x 80 km study area covered seven key benchmark soils of the Guinea savannah zone (Figure 2.2), which are distinguished along three soil toposequences (Adu 1995). On the upper slopes occur Eutric Plinthosols (Kpelesawgu series, in the Ghanaian soil classification system). Chromic Lixisols (Kumayili series), Gleyic Planosols (Lima series) and Petric Plinthosols (Changnalili series) covers the middle to lower slopes. The toe slopes are covered by Plinthic Lixisols (Siare series), Gleyic Fluvisols (Dagare series) and Fluvic Gleysols (Volta series). The underlying geology from which these benchmark soil types are developed mainly consists of sandstones, shales, mudstones, quartzites and alluvial sediments of the Voltain platforms (Junner 1940). See Adu (1995) for details on these benchmark soils.

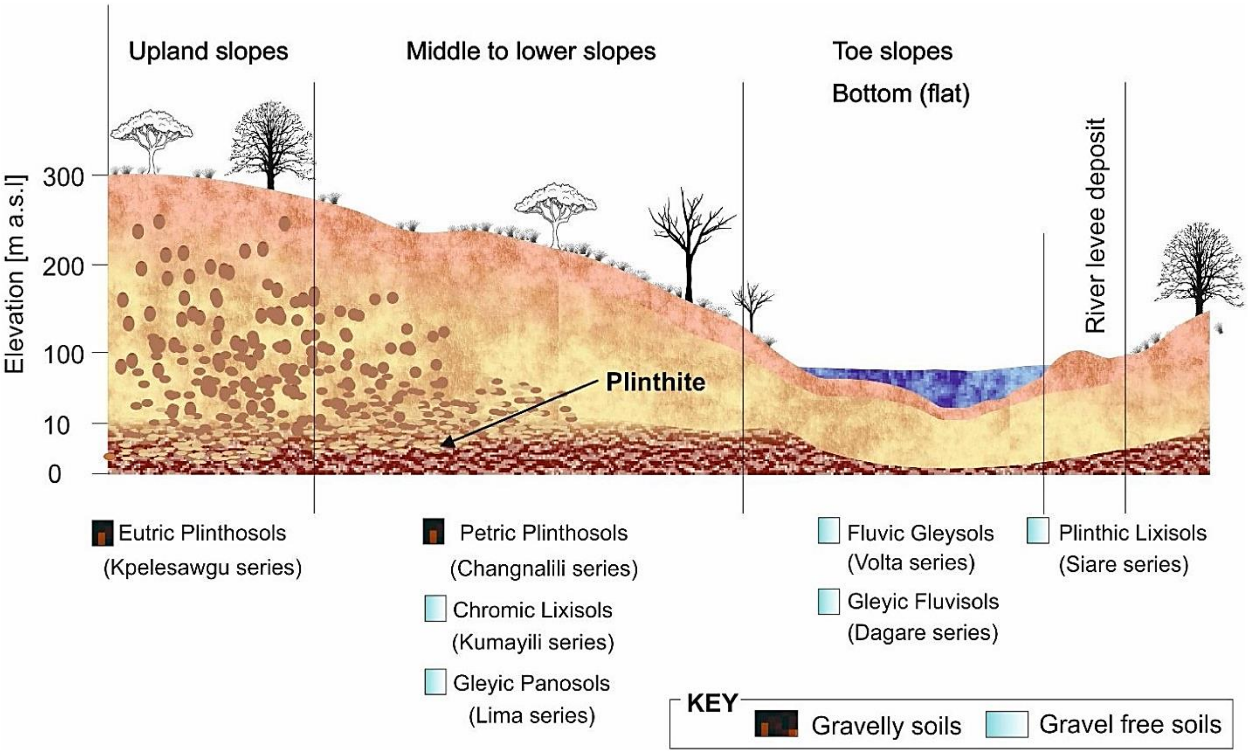


Figure 2.2. Key benchmark soil types of the Guinea savannah zone distinguished along three soil toposequences. Soil names are in Ghanaian soil classification system. Chart not drawn to scale. Source: Nketia et al. (2019).

2.3 Sampling stratification

The sampling stratification used for all research tasks of this thesis (i.e. study 1–4), followed the approach of Nketia et al. (2019). In the sampling design, we combined the global weighted principal component analysis with a cost-constrained conditioned Latin hypercube algorithm to select locations that represented the soil toposequences and landscape heterogeneities of the study area (Figure 2.1D). While the global weighted principal component analysis captured defined local structures and localized spatio-temporal autocorrelation at a point scale, the cost-constrained conditioned Latin hypercube algorithm optimized the selection of representative locations that highly explained the variability of SM for our case study areas. In addition, the sampling stratification suppressed the representation and selection of locations that had minimal influence on the variability of SM in the study area. See Nketia et al. (2019) for further description on the sampling stratification.

2.4 *In situ* SM measurement and soil sampling

At each sampling location and along the three topographical units, we inserted an access tube into the soil to a depth of 100 cm for SM measurements (Figure 2.3A–C). From each access tube, SM was measured at six soil depth intervals (i.e. 0–10, 10–20, 20–30, 30–40, 40–60 and 60–100 cm) with the use of a calibrated moisture probe (PR2/60, Delta-T Devices) (Figure 2.3D). In each PR2/60 SM measurement, the probe was turned clockwise ~45°, and thus, it represented a quadrant of measurement for the volumetric SM. An average of four repetitive *in situ* SM measurement represented each measuring soil depth at each location. This ensured a reduction in field-scale uncertainties in the measured SM. In total, we obtained 10-sets of measurements at a time interval of 12 days from February to June 2018. The temporal resolution of 12 days was designed to coincide with Sentinel-1 satellite overpass time for the study area, which was also part of our objectives to assimilate SM in the root zone of soils through Sentinel-1 and -2 satellite images.

We collected undisturbed soil samples from a soil profile, with a depth of 60 cm, at close proximity (~1 m away) to each model selected sampling location. The undisturbed soil samples were collected at five soil depth intervals (i.e. 0–10, 10–20, 20–30, 30–40 and 40–60 cm) at each sampling location. We took soil samples, vertically, in a stainless-steel cylinder with a volume of 110.78 cm^3 (Figure 2.3E–F). Soil samples from the 60–100 cm depth interval were not included due to inherent petro-plinthite soil layer, which made sampling impossible. In total, we collected 170 soil samples and analyzed for gravimetric SM, bulk density and particle size distribution.

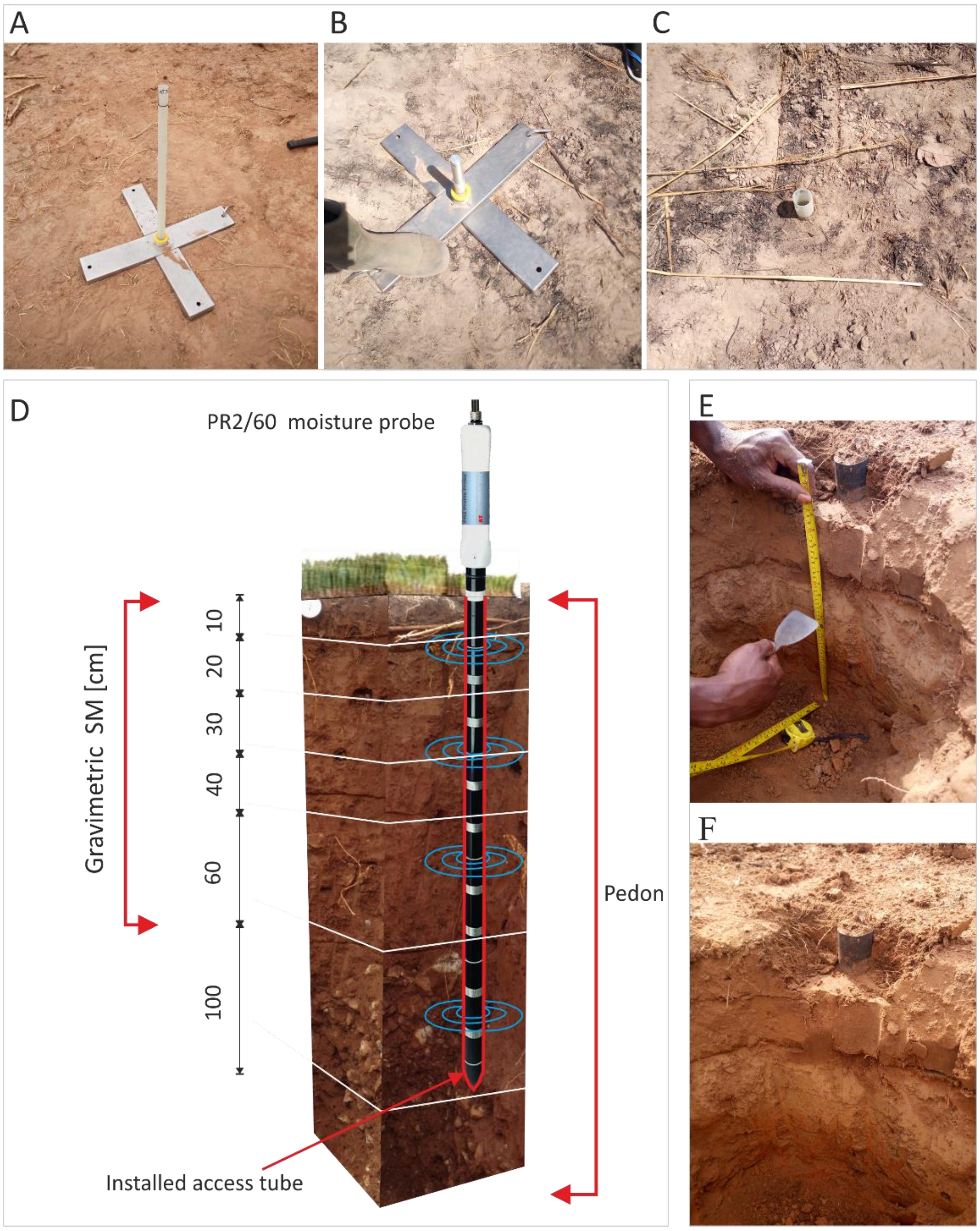


Figure 2.3. In situ SM measurement and undisturbed soil sampling. A–C represents the installation of PR2/60 capacitance profile probe access tube (Delta-T Devices) to a depth of 100 cm. D – Three-dimensional representation of soil profile with installed access tube and PR2/60 capacitance profile probe. E and F shows soil sampling with a stainless-steel cylinder at specific soil depths. PR2/60 image source: Delta-T Devices.

Chapter 3: Thesis structure and included publications

3.1 Structure of the following parts of the thesis

This thesis is a compilation of methods, results and findings from four main research tasks, which contributes to the reliability of using *in situ* SM, remote sensing imagery and soil predictive models to develop root zone SM and SWS information critical for improved food-crop production in water deficit landscapes. This thesis is further structured with two main sections: (B) publications and manuscripts and (C) general conclusions and outlook.

The other two sections are described as:

- **Section B** is categorized into two sub-sections: (1) research and (2) data. The research sub-section presents synthesis studies from the four specific research tasks while the data sub-section describes the associated open-source data repositories for both raw and analyzed datasets used for the specific research tasks:
    - **Research:**
        - **Study 1 (Chapter 4)** focuses and summarizes the results from a case study that was used to validate a newly designed soil sampling method, which was necessary to assess the spatio-temporal variability of SM and SWS of our case study area.
        - **Study 2 (Chapter 5)** investigates whether there are any unused potentials (in this case SM and SWS reserves) that can be tapped for food-crop production in the Guinea savannah zone, especially during the dry-season.
        - **Study 3 (Chapter 6)** presents and illustrates results from an improved remote sensing SM estimation technique via integrated remote sensing data sources and machine-learning algorithms.
        - **Study 4 (Chapter 7)** combines approaches, methods and datasets from research task 2 and 3 to innovatively assimilate SM information from 0 cm to 100 cm soil depth in order to explicitly map the four-dimensional distribution of root zone SWS at 100 m spatial resolution.
    - **Data:**
        - **Data repository 1 (Chapter 8)** describes and make publicly available the datasets on the spatio-temporal *in situ* SM and laboratory analyzed results, i.e. gravimetric SM and physical soil properties for the sampling locations.

    - **Data repository 2 (Chapter 9)** also describes and make publicly available the spatio-temporal root zone SWS datasets also for the sampling locations.
- **Section C** summarizes scientific contributions made, recommendations for improve food-crop production and their associated implications for policy advice. In addition, limitations that necessitates future research for this thesis are provided.

In order to ensure consistent presentation format and style for our findings throughout this thesis, we edited and reformatted scientific publications and manuscripts, which may differ from the styles of the journals. However, publications and manuscripts are duly acknowledged. Editing and modifications includes:

1. Layout of text, fonts and tables,
2. Numbering of pages, sections, figures and tables, and
3. Citations and references style.

3.2 Contributions to publications and manuscripts

From section B (Chapter 4–9), various contributions were made and they are described in Table 3.1–3.6.

Table 3.1. Contributions to study 1 (Chapter 4)

Title: A new method for selecting sites for soil sampling, coupling global weighted principal component analysis and a cost-constrained conditioned Latin hypercube algorithm	
Authors	***Contributions to study***
Kwabena Abrefa Nketia[1, 3]	Conceptualized and designed study, field campaign, data organization and analysis, manuscripts structure and writeup.
Stephen Boahen Asabere[1]	Reviewed draft manuscript.
Stefan Erasmi[2]	Reviewed draft manuscript.
Daniela Sauer[1]	Conceptualized and designed study, and reviewed draft manuscript.
Current status: published	**DOI**: 10.1016/j.mex.2019.02.005

[1]Physical Geography Dept. University of Göttingen, Germany, [2]Department Cartography, GIS and Remote Sensing, Institute of Geography, University of Göttingen. [3]Council for Scientific and Industrial Research-Soil Research Institute, Kumasi-Ghana.

Table 3.2. Contributions to study 2 (Chapter 5)

Title: Spatial and temporal dynamics of soil moisture in benchmark soils of the Guinea savannah zone of Ghana - is there an unused potential for food-crop production?	
Authors	***Contributions to study***
Kwabena Abrefa Nketia[1, 4]	Conceptualized and designed study, field campaign, data organization and analysis, manuscripts structure and writeup.
Joscha Becker[2]	Reviewed draft manuscript.
Stephen Boahen Asabere[1]	Reviewed draft manuscript.
Stefan Erasmi[3]	Reviewed draft manuscript.
Daniela Sauer[1]	Conceptualized and designed study, and reviewed draft manuscript.
Current status: under review	To be submitted to SOIL

[1]Physical Geography Dept. University of Göttingen, Germany, [2]Institute of Soil Science, Universität Hamburg, Germany, [3]Thünen-Institute of Farm Economics, Braunschweig, Germany. [4]Council for Scientific and Industrial Research-Soil Research Institute, Kumasi-Ghana.

Table 3.3. Contributions to study 3 (Chapter 6)

Title: Estimating soil moisture from integrated Sentinel-1, -2 and DEM-derived datasets using an ensemble of forest, neuron and vector-based machine-learning algorithms	
Authors	***Contributions to study***
Kwabena Abrefa Nketia[1, 3]	Conceptualized and designed study, field campaign, data organization and analysis, manuscripts structure and writeup.
Stefan Erasmi[2]	Reviewed draft manuscript.
Daniela Sauer[1]	Conceptualized and designed study, and reviewed draft manuscript.
Current status: under review	To be submitted to Remote Sensing of Environment

[1]Physical Geography Dept. University of Göttingen, Germany, [2]Thünen-Institute of Farm Economics, Braunschweig, Germany. [3]Council for Scientific and Industrial Research-Soil Research Institute, Kumasi-Ghana.

Table 3.4. Contributions to study 4 (Chapter 7)

Title: Spatio-temporal mapping of soil water storage in a semi-arid landscape of Northern Ghana	
Authors	***Contributions to study***
Kwabena Abrefa Nketia[1, 5]	Conceptualized and designed study, field campaign, data organization and analysis, manuscripts structure and writeup.
Amanda Ramcharan[2]	Reviewed draft manuscript and supported data analysis.
Stephen Boahen Asabere[1]	Reviewed draft manuscript.
Steffen Herbold[3]	Reviewed draft manuscript and checked mathematical syntax.
Stefan Erasmi[4]	Reviewed draft manuscript.
Daniela Sauer[1]	Conceptualized and designed study, and reviewed draft manuscript.
Current status: under review	To be submitted to Geoderma

[1]Physical Geography Dept. University of Göttingen, Germany, [2]Penn State University, United States, [3]Institute of Computer Science. University of Göttingen, Germany, [4]Thünen-Institute of Farm Economics, Braunschweig, Germany. [5]Council for Scientific and Industrial Research-Soil Research Institute, Kumasi-Ghana.

Table 3.5. Contributions to data archiving 1 (Chapter 8)

Title: Data on seasonal soil moisture and physical soil properties from the breadbasket of Ghana	
Authors	***Contributions to data archiving***
Kwabena Abrefa Nketia[1, 2]	Conceptualized and designed data repository, data organization, manuscripts structure and writeup.
Daniela Sauer[1]	Reviewed draft manuscript.
Current status: under review	To be submitted to Data in Brief

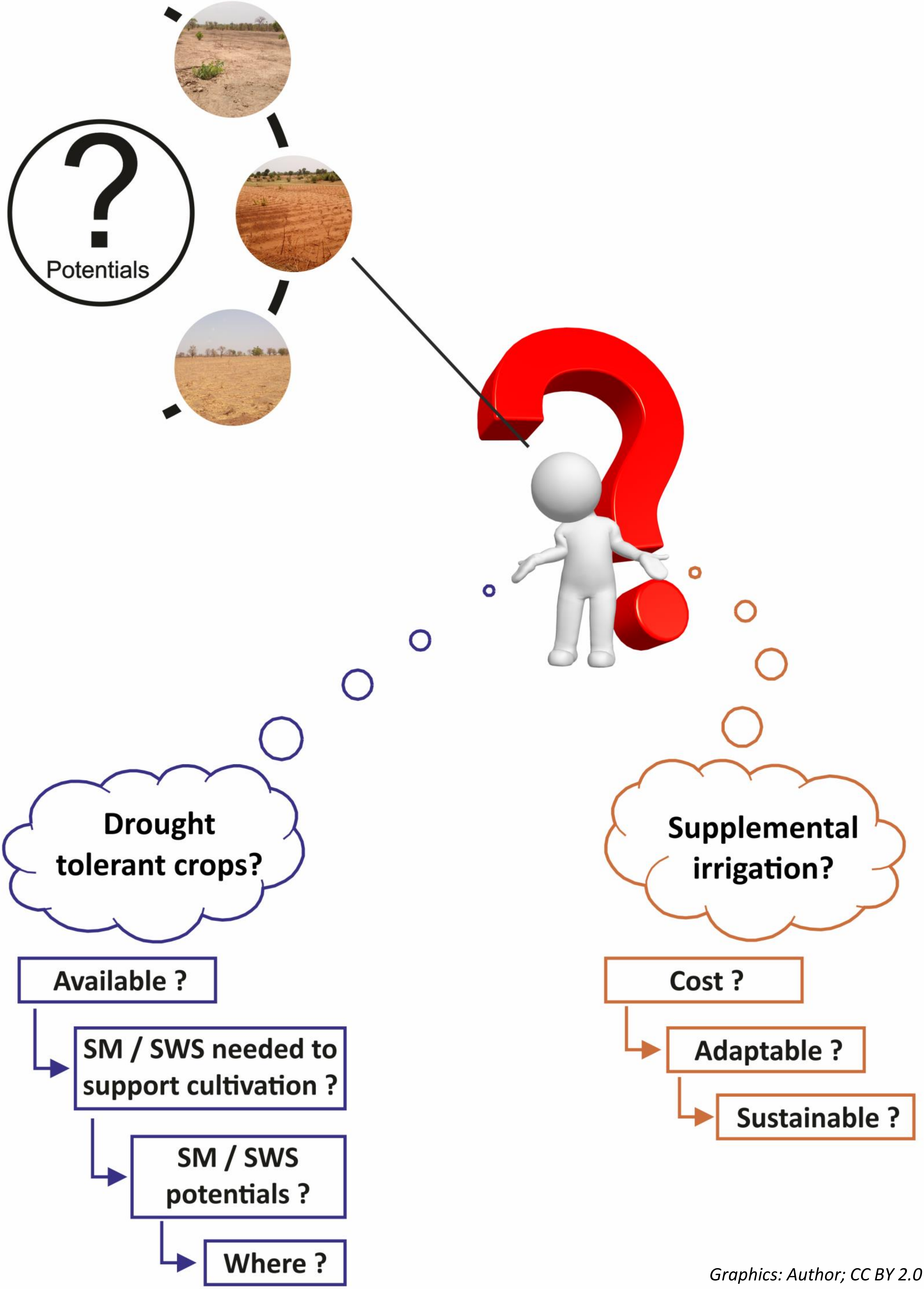
[1]Physical Geography Dept. University of Göttingen, Germany. [2]Council for Scientific and Industrial Research-Soil Research Institute, Kumasi-Ghana.

Table 3.6. Contributions to data archiving 2 (Chapter 9)

Title: Data on the spatio-temporal root zone soil water storage – the Guinea savannah zone of Ghana	
Authors	***Contributions to data archiving***
Kwabena Abrefa Nketia[1, 2]	Conceptualized and designed data repository, data organization, manuscripts structure and writeup.
Daniela Sauer[1]	Reviewed draft manuscript.
Current status: under review	To be submitted to Data in Brief

[1]Physical Geography Dept. University of Göttingen, Germany. [2]Council for Scientific and Industrial Research-Soil Research Institute, Kumasi-Ghana.

Section B: Publications and manuscripts



Graphics: Author; CC BY 2.0

Chapter 4: A new method for selecting sites for soil sampling, coupling GWPCA and cLHC algorithms

This manuscript is published as:

Nketia KA, Asabere SB, Erasmi S, Sauer D. A new method for selecting sites for soil sampling, coupling global weighted principal component analysis and a cost-constrained conditioned Latin hypercube algorithm. MethodsX 2019; 6:284–99. DOI: 10.1016/j.mex.2019.02.005.

Abstract

Analyzing spatial patterns of soil properties in a landscape requires a sampling strategy that adequately covers soil toposequences. In this context, we developed a hybrid methodology that couples global weighted principal component analysis (GWPCA) and cost-constrained conditioned Latin hypercube algorithm (cLHC). This methodology produces an optimized sampling stratification by analyzing the local variability of the soil property, and the influence of environmental factors. The methodology captures the maximum local variances in the global auxiliary dataset with the GWPCA, and optimizes the selection of representative sampling locations for sampling with the cLHC. The methodology also suppresses the subsampling of auxiliary datasets from areas that are less representative of the soil property of interest. Consequently, the method stratifies the geographical space of interest in order to adequately represent the soil property. We present results on the tested method (*R^2 = 0.90 and RMSE = 0.18 m*) from the Guinea savannah zone of Ghana.

Keywords: auxiliary dataset; cLHC; GWPCA; localized spatial soil variability; optimized soil sampling design

4.1 Method details

Sampling designs aim at representing either the investigated soil property or the geographical space of a study area, or both (Hengl et al. 2003; Minasny and McBratney 2006). However, there are still shortcomings in the geographical-space sampling designs (Biswas and Zhang 2018). Minasny and McBratney (2006) proposed the conditioned Latin hypercube (cLHC) method as a feature-sampling approach, to address the shortcomings of the geographical-space sampling design. To ensure that a sampling strategy adequately represents both the geographical and the feature space, Minasny and McBratney (2006) recommended considering the whole range of auxiliary data available for a study area. Based on this suggestion, Levi and Rasmussen (2014) proposed a hybrid approach, in which they coupled an iterative principal component analysis (iPCA) with the cLHC. However, the iPCA hardly accounted for local spatial effects in their auxiliary datasets. The iPCA algorithm generally obscured the localized spatial effects in the auxiliary datasets (Harris et al. 2011; Kumar et al. 2012; Harris et al. 2015). Generally, the iPCA algorithm is unable to correlate the measured soil property and its local environment, which depicts the geographic variations in the soil and environmental characteristics across space. Hence, iPCA ignores spatial characteristics. Therefore, several authors suggested to correct this shortcoming by including a localized weighted spatial auto-correlated principal component analysis (Harris et al. 2011; Harris et al. 2015; Comber et al. 2016). However, even coupling iPCA with cLHC still does not account for geographical weightings that provide principal component scores and loadings at all data locations (Kumar et al. 2012). Therefore, in this paper we propose a global weighted principal component analysis (GWPCA) as an alternative to the iPCA. The advantage of the GWPCA is that it is able to recover the known dimensional spatial structures. Hence, it accounts for localized spatial autocorrelations in the algorithm that can explain the variability of auxiliary datasets (Charlton et al. 2010). Consequently, we propose a new method, in which we couple GWPCA and a cost-constrained cLHC, to optimize the representation of both the feature and the geographical space (Figure 4.1).

Similar to the *scorpan* concept (McBratney et al. 2003), auxiliary datasets, represented as indices, were used to explain the local spatial heterogeneities and the soil property of interest at the selected sampling locations. We evaluated all localized spatial effects, trends and variabilities in the auxiliary datasets by GWPCA, adopting an automatic bandwidth in the GWPCA calibration. Next, using the selected GWPCA principal components as model input parameters, we selected optimal sampling locations using the cLHC algorithm executing 1e+04–5e+04 iterations. We incorporated a cost layer in the cLHC algorithm to suppress subsampling from areas that had only minor influence on the soil variable. Finally, we evaluated the model selections using root mean square error and correlation coefficient between model selected and actual locations. We chose SM to test the method. In all stages of the proposed hybrid approach, we used software (R Core Team 2017), specifically the packages (Kassambara and Mundt 2016), (Brenning et al. 2018), (Revelle 2018) and (Roudier et al. 2012). Source codes (i.e. scripts) for full method implementation is publicly accessible on DOI: [10.5281/zenodo.3662411.](https://doi.org/10.5281/zenodo.3662411)

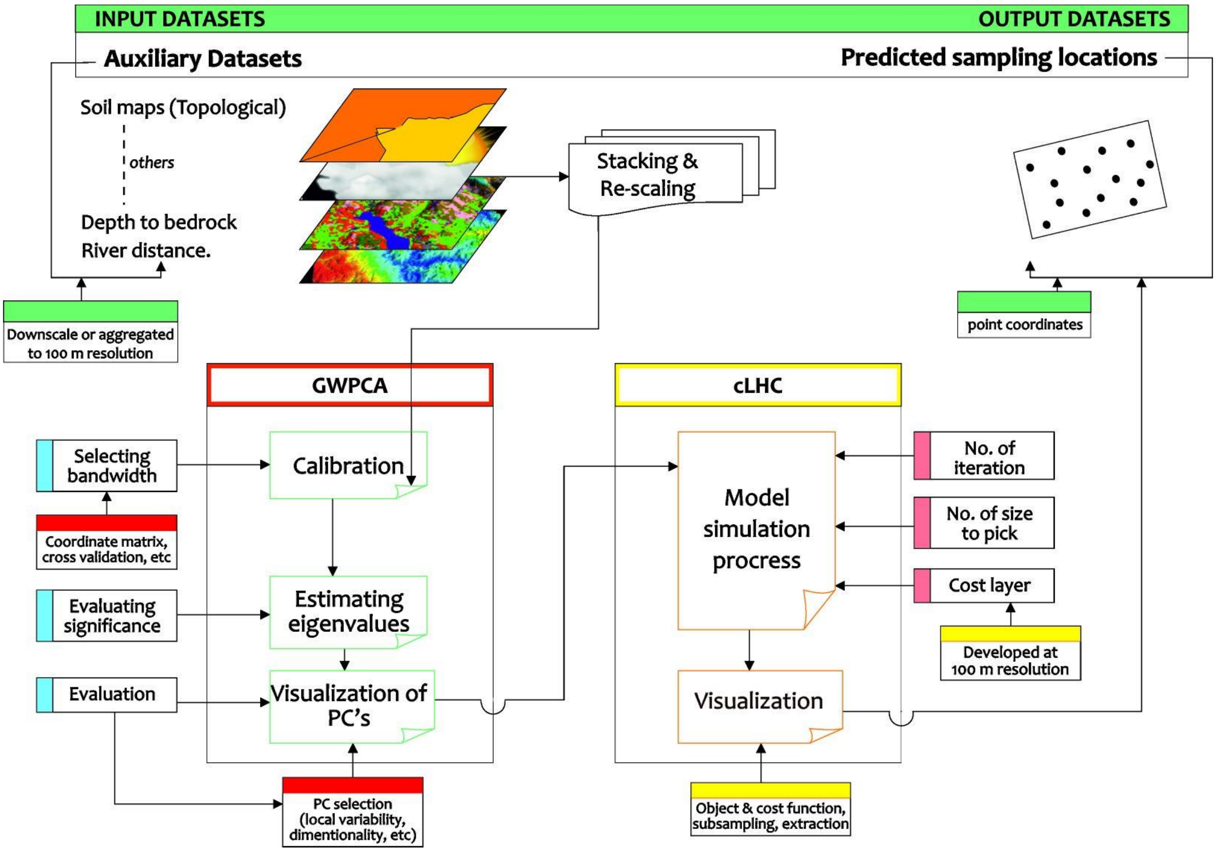


Figure 4.1. Methodological flowchart of the ensembled global weighted principal component analysis and a cost-constrained conditioned Latin hypercube algorithm.

4.1.1 GWPCA

The GWPCA is an add-on to the standard principal component analysis (Harris et al. 2011; Kumar et al. 2012; Harris et al. 2015). We evaluated all localized spatial effects, trends and variabilities in the auxiliary datasets by the GWPCA. Geographical weights (GW) used in the GWPCA were determined by a bi-square function (Equation 4.1).

$$GW_{ij} = \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2 \quad (4.1)$$

where d_{ij} is the distance between the spatial location *i* and *j* at a bandwidth *b* in determining the kernel size of the PCA.

Finally, we estimated the global weighted principal components at each location *(x_i, y_i)* (Equation 4.2).

$$LVL^T \mid (x_i, y_i) = \sum(x_i, y_i) \quad (4.2)$$

where LVL is the local eigenstructure, and $\sum(x_i, y_i)$ is the GW variance-covariance matrix for location (x_i, y_i).

We used a component matrix of the loadings to explore the local variations in the auxiliary data used in this study. Positive/negative signs associated with the loadings indicate, how each auxiliary data is associated with other auxiliary data. The geographically weighted standard deviation of auxiliary datasets was estimated by equation 4.3.

$$\sqrt{\sigma_i} = \left[\sum_{j=1}^{n}(x_1 - \bar{x}_1)^2\, GW_{ij}\right]^{0.5} \quad (4.3)$$

where $x_1 - \bar{x}_1$is the auxiliary data and its mean and GW_{ij} is the geographical weights between the spatial location *i* and *j*.

4.1.2 cLHC algorithm

A cost layer was introduced into the simulated annealing process within the cLHC algorithm (Roudier et al. 2012) as equation 4.4. The cost layer suppressed the subsampling of selected PCs of the GWPCA auxiliary data from areas that had only minor influence on SM in the cLHC algorithm outputs.

$$C_{cost_{(j)}} = e^{\left(-\frac{\Delta cost_j}{T}\right)} \quad (4.4)$$

where $\Delta cost_{(j)} = cost(j) - cost(j-1)$. Within the sampling schemes, $C_{cost_{(j)}}$was the sum of the cost layers of individual locations at j iterations in the simulated annealing. Details on applying the standard and cost-constrained cLHC algorithms can be obtained from the studies of Minasny and McBratney (2006) and Roudier et al. (2012).

4.1.3 Evaluation of the accuracy of the model selected sampling locations in the field

We tested the performance of the hybrid approach in the field, using six covariates, including soil type, parent material, landform, drainage, effective soil thickness, and the possibility to fix access tubes without impedance (Table 4.1). The selection of these indicators was corroborated by studies of Adu (1995) and Adu and Asiamah (2003) in the Guinea savannah zone. We evaluated each model selected sampling location in the field, by assigning either a value of 0 (= unsuitable) or 1 (= suitable) to each of the six parameters. Subsequently, the total score of each selected sampling location was averaged and expressed as percentage. For the selected sampling locations, the root mean square error (RMSE) and the correlation between the predicted and actual location was estimated. In this way, we evaluated the average error and the suitability of the model to select representative sampling locations (Equation 4.5).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[P_t - A_t]^2} \qquad (4.5)$$

where P is the model selected sampling location at a feature space t, n is the number of model selected sampling locations and A is the confirmed/actual *in situ* sampling location.

Table 4.1. Evaluation form to confirm the suitability of predicted sampling location

Locations	**Field Conditions**						
	Soil type	**Geology**	**Land form**	**Possibility to fix access tube**	**Drainage**	**Effective soil depth > 100 cm**	**0 | 1 Score**
AT01							
AT02							
…							
…							
…							
AT38							

4.2 Validation of the proposed hybrid methodology

4.2.1 Auxiliary datasets used in the study

We selected twenty auxiliary datasets, which we considered to represent factors that contribute to the local spatial variability of SM. We thus expected these datasets to be suitable to explain the SM dynamics at the model selected sampling locations. They were also included in the modelling process, in order to optimize the selection of sampling locations. Collinearity between the auxiliary datasets was accounted for with the GWPCA algorithm (Harris et al. 2011). Table 4.2 and Figure 4.2 present summaries of the auxiliary datasets used in this study. Each auxiliary dataset was either sourced or resampled at 100 m resolution via the cubic spline model of the $GDAL$ package (Mitchell and Developers 2014). DEM-derived surfaces were obtained in R via the $SAGA$- GIS morphometry and hydrology functions (Conrad et al. 2015).

Table 4.2. Auxiliary datasets used in the study

Abbreviation	**Description**	**Remarks**	**Units**
DEM	Digital elevation model	Representation of the terrain surface, steepness, wetness and to represent other geomorphological parameters	m
Slope	Slope inclination		degree
SAVI	Soil adjusted vegetation index		-
TWI	SAGA topographic wetness index		-
Landforms	USGS topographical landform classes		-
AWC	Available water content	Legacy information on moisture content	%
BD	Bulk density	restriction to root growth, infiltration, percolation and the ability of roots to reach moist zones in the soil	$Mg\ m^{-3}$
Bedrck	Depth to bedrock	Depth to impenetrable layer	cm
Clay	Clay content of the soil	Legacy information on clay content	%
Sand	Sand content of the soil	Legacy information on sand content	%
Silt	Silt content of the soil	Legacy information on silt content	%
Drainage	FAO soil drainage classes	-	-
Geology	Geological formation	-	-
Lithology	FAO lithological classes	-	-
Riverdist	River distance		km
Temp	Spatial pattern of temperature	-	°C
Precip	Spatial pattern of precipitation	-	mm
Sent1A_VH	Calibrated sigmaO Sentinel-1A radar backscatter coefficients in VH polarization	Proxy for SM	dB
Sent1A_VV	Calibrated sigmaO Sentinel-1A radar backscatter coefficients in VV polarization	Proxy for SM	dB
Soil_types	Mapping units at series level	-	-
WatBal	Water balance	-	%
WatCov	Spatial coverage of hydrology networks	-	ha

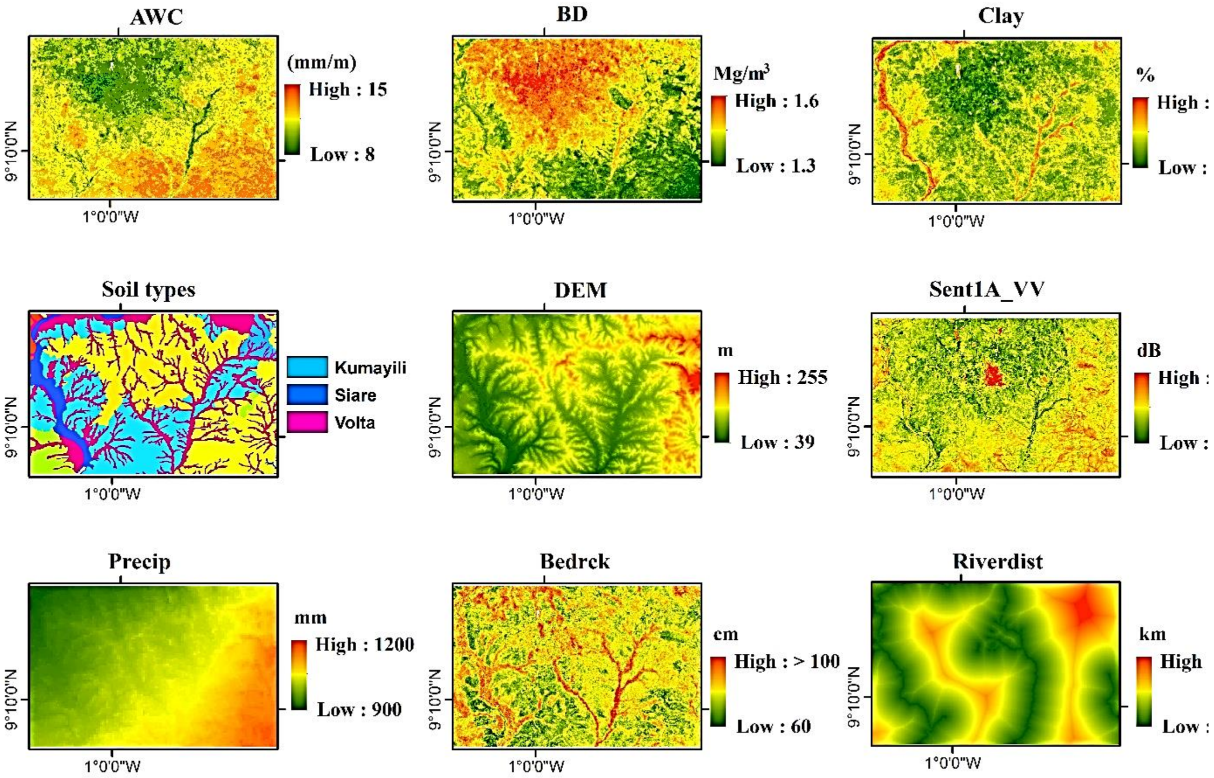


Figure 4.2. Extract from list of auxiliary datasets used to analyze the soil-landscape heterogeneity in representing SM. AWC = available water content, BD = bulk density, Clay = clay content, Sent1A_VV = calibrated Sentinel-1 radar backscatter coefficient in VV polarization, Precip = precipitation, Bedrck = depth to bedrock, Riverdist = river distance.

4.2.2 Selecting bandwidth for the GWPCA

Selecting an optimal bandwidth was the key to achieve an optimized GWPCA algorithm. Following the approach of Harris et al. (2015), we adopted an automatic bandwidth selection process. The stack of auxiliary datasets had dimensions of 315, 579 and 21 (number of rows, columns and layers respectively) and 182,385 pixels. It was thus impossible within a reasonable time-span to automatically select a bandwidth using the entire auxiliary datasets, which had GWPCA loadings at each 182,385 pixel sites. This challenge was related to the required computing power and processing time in the cross-validation algorithm, because each observation omitted was reconstructed using the principal components (PC) derived from the observations of the entire stack of auxiliary datasets. Therefore, we randomly selected 10% of the auxiliary datasets for use in the automatic bandwidth selection process. The selection of 10% of the datasets was guided by a series of cross validation evaluations.

We identified an adaptive bandwidth of 16.2 km at k = 10 (principal components) as the optimum minimized fit between the score and auxiliary data.

4.2.3 Development of the cost layer and the cost-constrained cLHC algorithm

In the cLHC annealing simulation process, model optimization was implemented by executing 1e+04–5e+04 iterations (increments of 1e+04). Because the user can define the number of cLHC selection outputs, we defined 38 locations in this study, as this was the maximum number of access tubes we had for the SM measurements. Thus, depending on the objectives and resources available for a particular study, users of this methodology can assign any maximum number of outputs. Roudier et al. (2012) used rough terrain, surface gradient and distance to road or trail network as criteria to design their cost constraint in simulating the annealing process. In this study, we used a similar approach but a different key criterion, namely the topographic wetness index (TWI), as a suitable constraint indicator directly affecting SM. The TWI is generated with an upslope contribution area (Western et al. 1999) and accounts for water redistribution within crest, ridges and depressions in an area (Ballerine 2017). The TWI gives an indication of the potential SM contribution areas within the top and bottom soil layers (Huang et al. 2016) and quantifies the spatial scale effects on hydrological processes (Grabs et al. 2009; Lei et al. 2016). In the TWI, slope inclination is classified at very short ranges to account even for slight changes in topography and local slope. In this study, low values represented crests and ridges, whereas high values represented depressions. We assumed that rainfall, infiltration, percolation and flooding by rivers during the rainy season are the only means by which SM is replenished in the soil layers of the study area. This assumption is based on the fact that in the study area, there is an inherent strong plinthic to petro-plinthic horizon at ~60–100 cm depth, which largely hinders contribution of ground water to SM (Adu 1995; Asiamah and Dedzoe 1999). Hence, we chose the TWI as cost-constraint criterion for assessing SM in the study area in the cLHC annealing simulation. Using the $rsaga$ package (Brenning et al. 2018), we developed a TWI layer at a 100 m resolution (Figure 4.3).

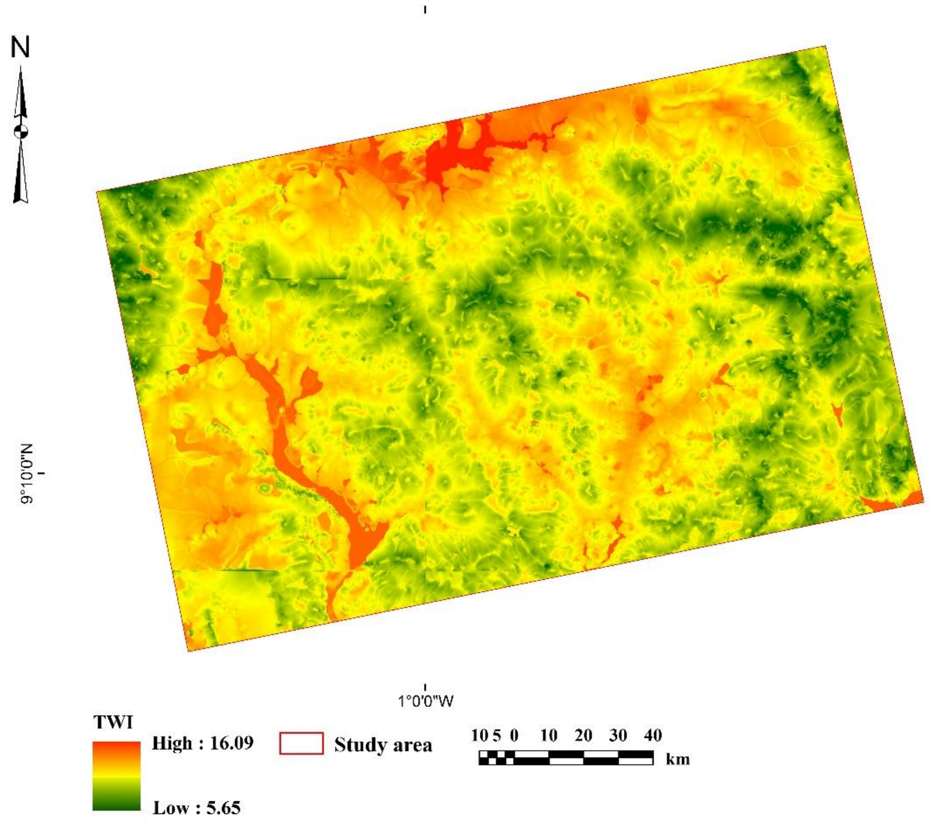


Figure 4.3. SAGA-TWI layer assigned as cost in the cLHC, simulating annealing at cooling temperature at iteration j.

4.3 Method implementation and outputs

4.3.1 GWPCA

Only PCs that accounted for eigenvalues ≥ 1 were considered. Temperature showed the strongest positive correlation, followed by bulk density, silt and clay contents (Figure 4.4). Increasingly negative correlations were found for water balance < precipitation < DEM < drainage < sand content < available water content and sentinel-1A. Within the rotational matrix of the global weighted PCs, available water content always showed the largest contribution in the list of auxiliary datasets, whereas sand content contributed least. Thus, we conclude that most of the variation was explained by available water content.

PC1 to PC6 together explained 91.5% of the observed variance (see appendix Table S4.1). Generally, the highest positive loading of 0.552 was recorded for PC5 and the lowest negative loading of -0.453 was recorded for PC6. Together, PC1 and PC2 accounted for 74.7% of the local variation in the list of auxiliary datasets used. These findings suggest that AWC, clay and silt content are the key factors that need to be considered to explain local variability of SM.

By adding PC3, the explained variability reached 82.1%. PC4 explained an additional 3.5% of the observed local variation. In PC4, soil type showed the highest positive loading of 0.446, whereas river distance exhibited the lowest negative loading of -0.445. We thus conclude that the local soil types, affected by their distance to rivers, markedly influence the spatial variability of SM. Adding PC5 and PC6 resulted in a cumulative explained variability of 88.1% and 91.5%, respectively. River distance exhibited the lowest negative loadings in PC4, PC5 and PC6. AWC showed maximum positive loadings in the GWPCA two times. Thus, among the list of auxiliary datasets used in this study, AWC and river distance showed the strongest influence on the spatial variability of SM. The reason for the important role of distance to a river that we found in this case may be the dense river network and the generally low elevation within the area that leads to a far-reaching riparian influence.

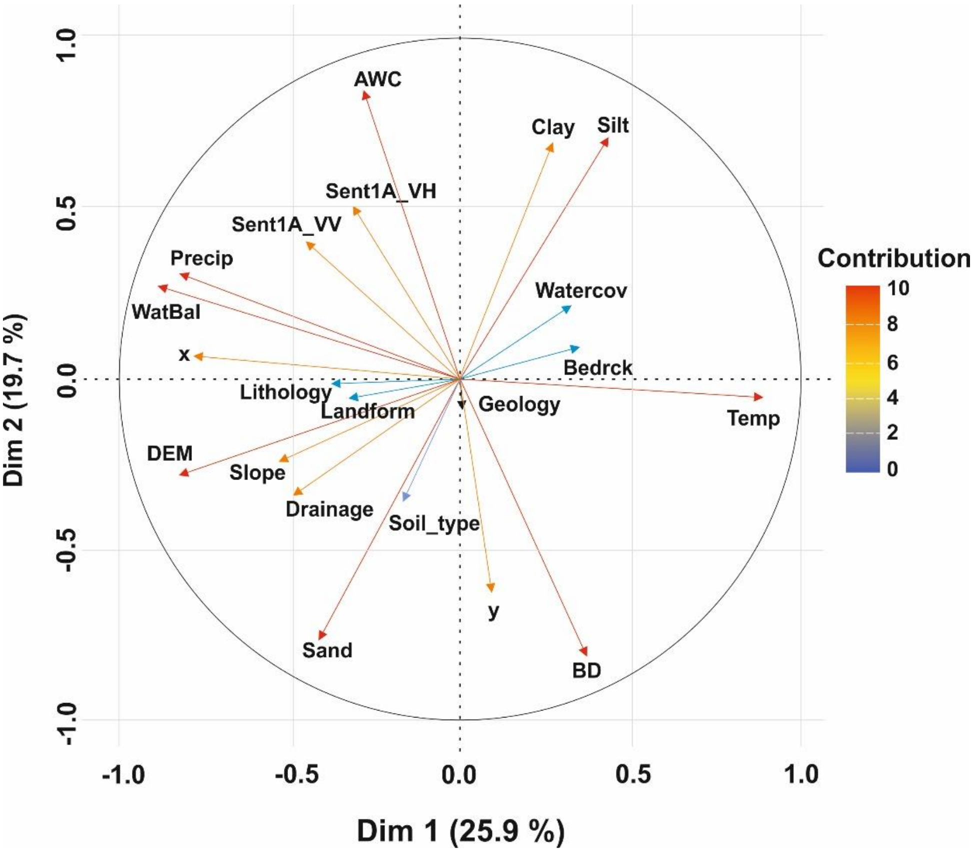


Figure 4.4. Correlation of the variation of list of auxiliary datasets in the feature space used in the GWPCA. AWC = available water content, BD = bulk density, Clay and Silt = clay and silt contents respectively, Sent1A_VV and Sent1A_VH = calibrated Sentinel-1 radar backscatter coefficient in VV and VH polarization respectively, Precip = precipitation, Bedrck = depth to bedrock, DEM = digital elevation model, WatBal = water balance, Watercov = water cover, Soil_type = soil mapping units and Temp = temperature.

While the PCs showed the observed variances in the auxiliary datasets, it also indicated the collinearity between the auxiliary datasets (appendix Table S4.1). The larger or smaller the first or last PC, respectively, the stronger the collinearity between the auxiliary datasets (Harris et al. 2011). Hence, local variables that caused the local collinearity in the auxiliary datasets was identified and eliminated via the multivariate glyph cartogram plot prior to its use in the cLHC (Harris et al. 2011). The local spatial variability was analyzed in the GWPCA using a 16.2 km bandwidth window around each data point in the auxiliary datasets. This bandwidth is equivalent to half of the maximum distance from river networks (minimum = 0 km, maximum = 32.3 km). The selected automatic bandwidth interprets as ~50% of the entire auxiliary datasets were retained each time in the cross-validation algorithm to calibrate the GWPCA. Thus, ~50% of the auxiliary datasets were used each time to explain the localized spatial variabilities of the study area. In addition, the automatic bandwidth ensured a balance between the local variation and locations with less influence on the spatial variability, as reported by Kumar et al. (2012). Our findings suggest that the use of GWPCA will account for the local influences and collinearity of each auxiliary datasets on the proposed SM measurements with regards to their locality. The final output of the cLHC ensures that the proposed sampling locations cover 91.5% of the locally occurring site conditions, represented by the auxiliary datasets.

4.3.2 The cost-constrained cLHC algorithm

To implement the cLHC, the algorithm analyzed the selected PCs of the GWPCA to identify points in the landscape representing a Latin hypercube, similar to the approach of Minasny and McBratney [54]. In the subsequent series of iterations of this spatial representation of site conditions, the level of optimization of the objective function of the cLHC was shown at each iteration. Perturbations were recorded in both the objective and the cost-function optimization processes at < 5e+03 iterations, similar to what Roudier et al. (2012) reported. Contrary to the findings of Roudier et al. (2012), our results showed a clear full model optimization after 5e4 iterations, both in the evolution and in the cost functions of the cLHC algorithm (Figure 4.5). Between 1e+04 and 22e+03 iterations of the objective function, the optimization steadily increased until 5e+04 iterations were completed.

A comparison of the subsamples drawn from the selected PCs of the GWPCA used in the cLHC simulating annealing process is presented in appendix Figure S4.1. Generally, the subsamples were within the first and third quartiles of the selected PCs of the auxiliary datasets. Only very small subsamples were drawn at locations with low TWI (ridges and crests). At sites with high TWI (depressions), large subsamples were drawn for use in the cLHC simulating annealing process. The reason for this difference is that the cLHC simulating annealing process suppressed sampling of locations with little or no influence on spatial SM variability in the study area (in this case, from ridges and crests). Thus, the probability of model selected sampling locations on ridges and crests was low.

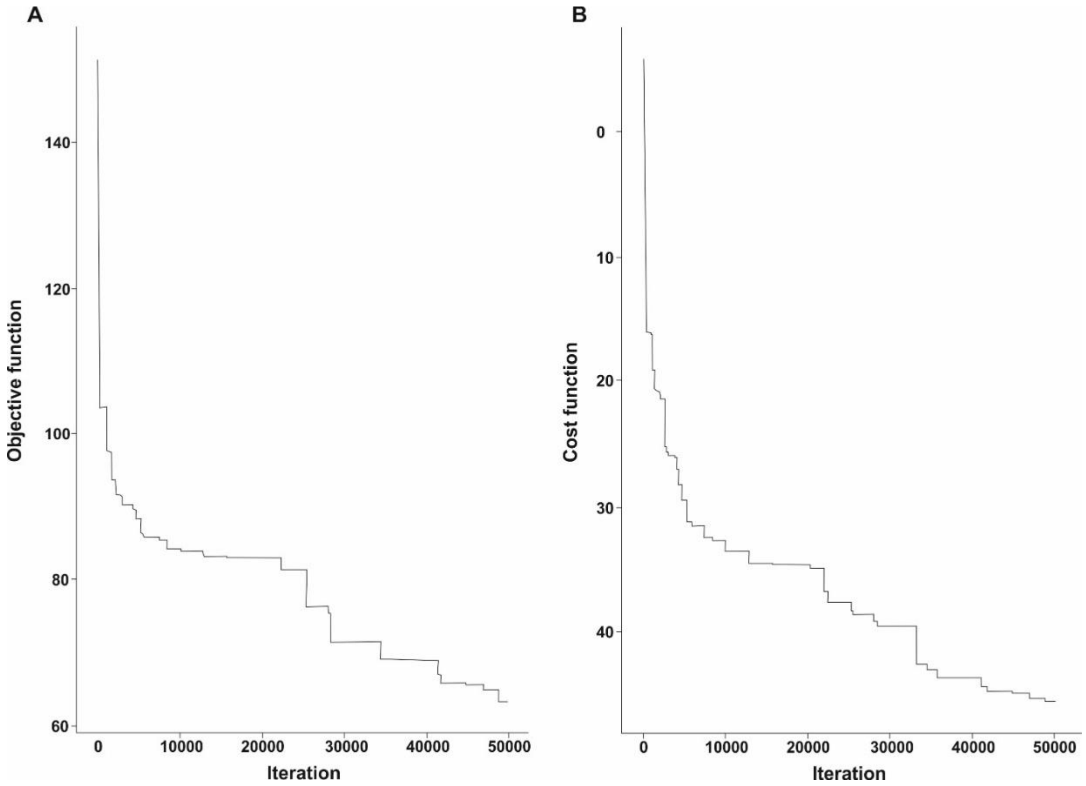


Figure 4.5. Evolution of the objective (A) and cost function (B) for the cLHC simulation from 1e4–5e4 iterations.

A single realization of the cLHC output shows that the cost layer in the cLHC influenced the selection of sampling locations (Figure 4.6). The cost-constrained cLHC stratified the selected sampling locations, based on the influence of TWI (similar to results reported by Levi and Rasmussen (2014)). In our study, the selected sampling locations represented the soil property of interest and in the geographical space, as also reported by Hengl et al. (2003). Locations selected by unconstrained cLHC algorithms were dispersed across the entire study area (Figure 4.6). The reason is that the cLHC optimization process accounted for all key landscape heterogeneities that occurred within the study area. The wide spatial distribution of sampling locations indicates that the cLHC effectively selected sampling locations, in terms of both the soil property of interest and the geographic space (Figure 4.6). However, some locations were selected similarly under both the unconstrained and the cost-constrained cLHC algorithms. This hybrid approach enables a scientist to assess the local variability of a soil property of interest and to derive adequate sampling designs for analyzing that soil property across a defined study area.

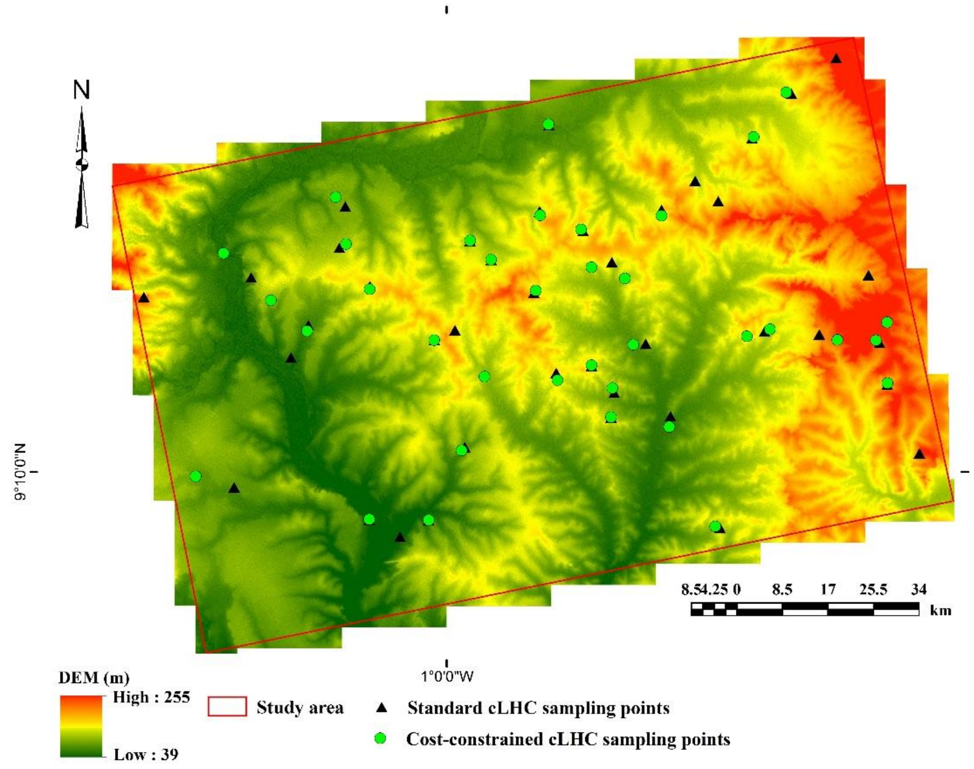


Figure 4.6. Single realization of the cLHC algorithm overlaid on a five-point 'z' exaggerated digital elevation model. Green circles represent the cost-constrained cLHC and black triangles represent unconstrained cLHC.

4.3.3 Accuracy assessment

The highest (1) and lowest (0.3) average scores were achieved by 71% and 2.63%, respectively, of the model selected sampling locations. Test of the method showed a RMSE of 0.18 m with a correlation coefficient (R^2) between actual and model selected location of 0.90. Our findings indicates that the model selected sampling locations were very close to the *in situ* sampling locations. AT14 and AT21 were found not suitable probably due to the main input dataset in the GWPCA, which affected the output of the cLHC algorithm. Although 91.5% of the observed local variability in the auxiliary dataset was explained by the GWPCA, it must be emphasized that incorrect input parameters can limit the performance of the cLHC.

4.4 Conclusion

In this paper, we have presented a new approach for selecting soil sampling locations that adequately represent both the soil property of interest and the geographical space across a study area. We implemented the proposed hybrid approach under a cost-constrained conditioned Latin hypercube algorithm, by using inputs from a global weighted principal component analysis. This method defines the local structure and accounts for localized spatial autocorrelation in explaining soil-landscape variability. The method also suppresses the occurrence of model-selected sampling locations in areas that are less representative of the soil property of interest. In addition, the method provides an appropriate base for selecting adequate sites for a given number of possible measurements. The proposed approach can guide the selection of adequate sites for soil measurements and installations of soil-monitoring stations, in the context of scientific studies and agricultural interventions.

Spatial and temporal dynamics of soil moisture in benchmark soils of the Guinea savannah zone of Ghana - is there an unused potential for food-crop production?

Photo: Author

Chapter 5: Spatial and temporal dynamics of SM in benchmark soils, of the Guinea savannah zone of Ghana - is there an unused potential for food-crop production?

This manuscript is under review as:
Nketia KA, Becker J, Asabere SB, Erasmi S, Sauer D. Spatial and temporal dynamics of soil moisture in benchmark soils of the Guinea savannah zone of Ghana - is there an unused potential for food-crop production?

Conference presentation at EGU General Assembly 2019, Vienna:
Vol. 21, EGU2019-13853-1, 2019: https://meetingorganizer.copernicus.org/EGU2019/EGU2019-13853-1.pdf

Abstract

The demand for increased crop production and food security in arid and semi-arid landscapes of Africa implies the need for adopting sustainable drought-adaptation measures by subsistence smallholder farmers. The efficiency of these adaptation measures strongly relies on local soil moisture (SM) conditions. However, spatio-temporal SM information is still missing in large areas of semi-arid Africa. Hence, the challenge to cope with long dry-spells and the implementation of drought-adaptation measures have yet remained unresolved. In this paper, we report spatio-temporal SM variability, measured over the dry season in the Guinea savannah zone of Ghana. Through this work, we aimed at assessing the potential of certain soils to provide enough SM for utilizing also part of the dry season for crop production, in order to satisfy the demand for increased crop production. With a calibrated PR2/60 moisture probe, we measured SM at six soil depths (0–100 cm) of 34 soil profiles, representing seven key benchmark soil types. *In situ* measurements were at 12-day intervals from February to June 2018 (n = 1,920). For each soil profile and depth, we further analyzed for gravimetric SM, bulk density (BD) and particle size distribution. We analyzed the influence of internal factors such as BD, clay and silt contents, and external factors such as slope, topographic wetness index, antecedent precipitation index (API) and evapotranspiration (ET_o). Multi-factor relationships, as well as direct and indirect interactions between SM and the internal-external factors were also analyzed. SM was low only in the sandy upper parts of the soils (≤ 10 cm), while in the lower parts of the soils (≥ 20 cm), clay content increased with soil depth and SM was kept for longer periods. Internal factors had stronger influence on SM variability than external factors. Clay and silt contents accounted for 91% of SM variability, while BD accounted for 11%. ET_o was about three times the daily mean precipitation. Temporally, ET_o influenced SM only in the sandy top soils and showed a strong negative relationship with SM (R^2 = 0.77). API strongly influenced SM at ≥ 20 cm soil depth, exhibiting a strong positive relationship with SM (R^2 = 0.83). The lowest SM (found at all soil depths) occurred in the beginning of March, whereas the highest SM occurred in the end of April. The main outcome of this work is that almost all soil types in the study area (except for the Kumayili soil series) below 10 cm depth have soil water storage potentials that match the water requirements of at least some drought-tolerant crops (> 9 mm day^{-1}). We conclude that most benchmark soils of the area hold the potential for implementing a dual farming system that can enhance agricultural productivity and increase the income of local farming communities.

Keywords: soil moisture, PR2/60 profile probe, semi-arid, spatio-temporal variability, soil water storage, Guinea savannah zone, food security, West Africa

5.1 Introduction

Soil moisture (SM) is an important factor that influences environmental processes such as the exchange of energy, water and carbon between the atmosphere and land (Brocca et al. 2017; Mohanty et al. 2017). SM also plays an important role in the control of the physical, chemical and biological soil processes within the soil system, which influences agricultural productivity (Weil and Brady 2016; Rossato et al. 2017). Particularly, it is a principal limiting factor in rainfed agriculture and its fluxes heavily drive landscape processes such as ground water recharge, in-soil lateral water movement, erosion, surface runoff and evapotranspiration (Tallon and Si 2015).

In tropical arid and semi-arid regions, climatic conditions are mainly characterized by high temperatures, high evaporation and low rainfalls, which makes water availability a critical input in the rainfed agricultural systems (Cafarelli et al. 2015; Greenland et al. 2018). Corroborating this observation, many studies have demonstrated that in water deficit periods, weather induced water stress can result in up to 50% decrease of total crop yields (Hall et al. 1982; Rossato et al. 2017). Although the importance of SM is widely recognized and underscored, future climate change projections which indicates an increase in dry days, 40% decrease in precipitation and rising risk of severe drought (Antonino et al. 2000), pose severe threat to the sustainability of crop production in these arid and semi-arid landscapes. Thus, it is evident that the prospect of improved crop production and food security in tropical arid and semi-arid regions under changing climatic conditions depends on the sustainable implementation of drought-adaptation measures.

The efficiency of most on-farm drought-adaption measures strongly relies on local SM conditions (Geng et al. 2015). However, an adequate localized spatio-temporal SM information of the root zone of soils is still missing in large areas of semi-arid landscapes of sub-Saharan Africa (SSA) (Vågen et al. 2010). SSA is of concern because, not only does it has ~80% of the population being smallholder farmers, but also, it has over 50% of the world's arable lands (Bourne Jr 2014; Guilpart et al. 2017; Xie et al. 2018). Due to the paucity of data omission on SM, the challenge to cope with the long dry-spells and the implementation of drought-adaptation measures have yet remained unresolved. Among such drought-adaptation measures is the implementation of irrigation schemes (Howell 2001) and the cultivation of drought tolerant crops (Fisher et al. 2015). In order to adequately and sustainably implement such drought-adaptation measures, it is critical to understand the spatial and temporal variability of SM of the root zone of soils of these landscapes (Megersa and Abdulahi 2015). For instance, the implementation of irrigation technologies such as drip and cover irrigation in order to adapt to the impacts of agricultural drought in semi-arid regions of other parts of the world was guided by SM variability information (Veste et al. 2008; Megersa and Abdulahi 2015; Smilovic et al. 2015; Frisvold et al. 2018; Greenland et al. 2018). Even in the absence of SM information, one key challenge SSA still face with irrigation schemes is the implementation and operational cost, hence, only 3% is currently practiced (van Ittersum et al. 2016). While another critical consideration to boost crop productivity in tropical arid and semi-arid regions is the resilience of cultivated crops to drought conditions (El-Sharkawy 1993; Fisher et al. 2015). The main challenge for these drought tolerant crop varieties is the available SM information that is needed to support them through their growing season (Critchley et al. 1991; Chivenge et al. 2015). It is therefore a fundamental step in the implementation of drought-adaptation measures to have a detailed understanding of the spatio-temporal variability of root zone SM, particularly in resource poor arid and semi-arid regions of SSA and Asia (Mashnik et al. 2017).

One region where sustaining crop production and human livelihood under water stress conditions is challenging is the Guinea savannah agro-ecological zone of Ghana (Armah et al. 2011; Adolwa et al. 2017; Antwi-Agyei et al. 2018), which is a reminiscent of the arable landscapes of West Africa and SSA at large (Morris et al. 2009). The Guinea savannah zone is of specific importance because, over the years, it has accounted for about 60% of agricultural outputs of SSA, and also, has the largest prospects for commercial agriculture in SSA (Morris et al. 2009; OECD 2016). The case of Ghana, in terms of rainfed agriculture, is indifferent to those of SSA (Adolwa et al. 2017). Ghanaian agriculture is dominated by resource poor subsistence smallholder farmers with farm sizes < 1 ha (Chamberlin 2008). These smallholder farmers produce ~80% of the total agriculture output and are mainly located in the Guinea savannah zone (Chamberlin 2008). Despite the high crop production potentials, high rural workforce, market potentials and an upsurge in advocacy for the adoption of integrated soil fertility management principles in the area, crop yields still fall short of their potential (Fening et al. 2006; SRID 2016). Partly because farming activities within the Guinea savannah zone are mainly limited to a single cropping season. This single season farming window limits livelihood options of these subsistence smallholder farming communities (Abass 2015; SRID 2016). Hence, identifying and utilizing parts of the dry season, as suggested by Abass (2015), to grow some suitable drought tolerant crops may be a key step to supplement crop production short falls of the main cropping season. Also, such an initiative could provide an option to increase livelihood for the rural communities of the Guinea savannah zone. However, no study has provided baseline information on the spatio-temporal variability of SM of the root zone of soils to support such drought-adaptation measures in the area. In addition, it is important to also investigate whether there is an unused potential (i.e. SM reserves) that can be tap to support food-crop production during water-deficit periods, especially in the dry season. The availability of such detailed spatio-temporal variability SM information of the root zone of soils may also help to adapt agricultural management practices, select suitable areas and periods where additional dry-season cropping is applicable. Furthermore, such an information will be useful to either prolong the single farming season or advocate for a dual farming system to boost agricultural productivity, which can be replicated for the other parts of Guinea savannah zone of SSA.

In this study, it is aimed to investigate the spatial and temporal variability of SM of the root zone of agricultural benchmark soils of the Guinea savannah zone of Ghana during the dry season. A plethora of studies have investigated the variability of SM in other parts of the world and have suggested that the spatio-temporal variability of SM is driven by several localized soil-environmental factors (Rawls et al. 2003; Baskan et al. 2013; Gutierrez-Lopez et al. 2014; Wang et al. 2016). In these studies, the frequently reported factors include soil texture, effective soil depth, land use, evapotranspiration, geomorphology and precipitation. One thing that was clear in these studies is that these multiple factors are site-specific, hence, we also identify and systematically investigate how SM is influenced by these soil-environmental factors for our case study areas. We simultaneously test the direct and indirect interactions between SM and the soil-environmental factors. We individually test both the spatial and temporal dependencies of SM variability as influenced by local soil types and by soil depths. In order to make recommendations for improved crop production, we further quantify the soil water storage potentials of the key benchmark soil types of the area at specific soil depths at different stages of the dry season. Thus, to indicate where crop water requirements for notable crops grown in the area can be met.

5.2 Materials and methods

5.2.1 Estimating PAWC in soils under maize

We adopted the methodology of Burk and Dalgliesh (2013) to investigate the thresholds of SM demands by maize crops (*Zea mays*) in the dry season for the case of our study areas. We selected maize as a test crop because it is the commonly cultivated crop in the area. Here, we quantified the PAWC in soils under maize through its critical stages, i.e. at anthesis and at maturity in two main stages: (1) estimating the drain upper limit via a soil profile wetting experiment and (2) estimating the crop lower limit via a rain exclusion experiment.

In the wetting experiment, we selected an area with a 5 x 5 m dimension, cleared it and created a trench of 10 cm depth around the boundaries of the plot. We laid a drip-irrigating tube (25 m long) that was connected to a water tank in a coil across the plot area to ensure the soil profile was wet uniformly from 0 cm to 100 cm (Figure 5.1A and 5.1B). We covered the laid-out drip-irrigating tube with a black plastic sheet to prevent evaporation from the surface soil layer (Figure 5.1C). The edges of the plastic sheet were placed into the trenches and back-filled with soil (Figure 5.1D). After the wetting process, the soil profile was allowed to drain, and undisturbed soil samples were collected at six soil depth intervals (i.e. 0–10, 10–20, 20–30, 30–40, 40–60 and 60–100 cm) at the center of the plot in order to characterize for the drain upper limit and bulk density.

While we undertook the wetting experiment, we also erected a rain exclusion tent on a maize plot with a 2 x 3 m dimension in order to exclude rain that might prevent maize extracting water to crop lower limit within the anthesis and maturity period (Figure 5.1E). To avoid a microclimate within the tent, parts of the tent were left open. We collected undisturbed soil samples also at six soil depths at close proximity (~1 m away) from the rain exclusion tent at maize anthesis and at the center of the rain exclusion plot at maize maturity. Undisturbed soil samples were further characterized for crop lower limit at anthesis and at maturity.

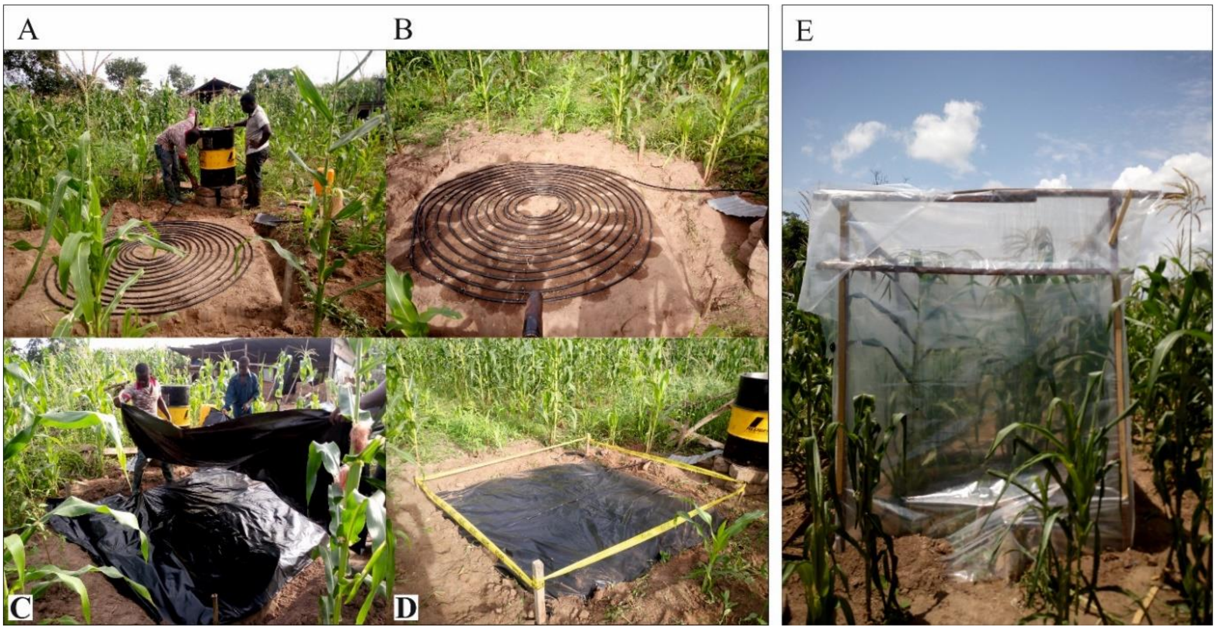


Figure 5.1. Soil profile wetting experiment (A — D) and the rain exclusion tent experiment (E) that were used to quantify PAWC from 0 — 100 cm soil depth under maize plant at its anthesis and maturity stages.

5.2.2 Datasets used in the study

In this study, we considered empirically measured variables as internal factors and those that were derived from other data sources as external factors. The internal factors included gravimetric SM (SM_g), PR2/60 moisture probe SM (SM_p), clay and silt contents and bulk density (BD). External factors also included local slope (LS), topographic wetness index (TWI), antecedent precipitation index (API) and evapotranspiration (ET_o). See Table 5.1 for a detailed description of all parameters used in explaining the spatio-temporal variability of SM in the study area.

Table 5.1. Parameters used in explaining the spatio-temporal variability of soil moisture in the different soil types of the study area and its implication for dry season farming

Parameters	**Abbreviation**	**Description**	**Units**
Spatial variability analysis			
Gravimetric soil moisture	SM_g	Volumetric soil moisture by gravimetric method	%
Bulk density	BD	Bulk density of soil samples	g cm^{-3}
Clay content	Clay	Percentage clay content of soil samples	%
Silt content	Silt	Percentage silt content of soil samples	%
Soil depth intervals	Depth	Standard depth intervals corresponding to PR2 profile probe measuring depths	cm
Local slope	LS	Slope surface	degree
Topographic wetness index	TWI	SAGA topographic wetness index	-
Temporal variability analysis			
PR2 soil moisture	SM_p	Volumetric soil moisture by PR2 capacitance moisture probe	%
Antecedent precipitation index	API	Daily mean antecedent precipitation index corresponding to SM_g and SM_p measurement day	mm
Evapotranspiration	ET_o	Daily mean evapotranspiration corresponding to SM_g and SM_p measurement day	mm day^{-1}
Implication for dry season farming			
Plant available water content	PAWC	Plant available water content for a 90-day maize plant both at anthesis and at maturity	mm
Soil water storage	SWS	Soil water storage estimated at standard depth intervals for the different soil types	mm

5.2.2.1 Internal factors

SM_g, clay and silt contents, and BD were determined for each SM measurement depth from each soil profile. We weighed fresh soil samples and afterwards oven dried at 105 °C for 48 h. We weighed the samples intermittently during the drying process until a constant dry weight was obtained. SM_g was calculated by subtracting the dry weight from the fresh weight. The profile BD and volumetric SM_g of each soil sample were calculated following Reynolds (1970) and Black and Hartge (1986) respectively. By following standard soil analytical procedures, we determined the soil texture (i.e. sand, silt and clay contents) using the hydrometer method (Bouyoucos 1962).

5.2.2.2 External factors

The Shuttle Radar Topographic Mission digital elevation model (DEM) with a spatial resolution of 1 arc-second was used to derive the LS for the study area. LS represented ground surface steepness, slightest change in topography and landforms. TWI, also calculated from the DEM, was used to explain the local flow accumulation, water distribution and accumulation within depression and ridges of the study area. Estimated daily from February 2010 to June 2018 was used to evaluate the changes of SM and soil water storage from the top to bottom soil layers due to gravity drainage over preceding days of rainfall events within the study area. is expressed by equation 5.1.

$$API = k * API_{d-1} + P_d \tag{5.1}$$

where API_d is the antecedent precipitation index for a day d, k is an empirical decay factor and P_d is rainfall for day d.

To assess SM loss from the surface soil layers, we also estimated the daily reference ET$_o$ (Equation 5.2) of the study area from February 2010 to June 2018. ET$_o$ was estimated by the FAO Penman-Monteith equation (Allen et al. 1998).

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{5.2}$$

where Δ is the slope vapour pressure curve, R_n is the net radiation at the crop surface, G is the soil heat flux density, γ is the psychrometric constant, T is the mean daily air temperature at 2 m height, u_2 is the wind speed at 2 m height, e_s is the saturation vapour pressure, e_a is the actual vapour pressure and $e_s - e_a$ is the saturation vapour pressure deficit.

The external factors described above were developed in software (R Core Team 2017) using the $SAGA$ morphometry and hydrology functions (Conrad et al. 2015) and R packages $r - gdal$ (Mitchell and Developers 2014), $plyr$ (Wickham 2018) and $r - saga$ (Brenning et al. 2018).

5.2.2.3 SWS

For each effective soil thickness, the SWS was calculated. In this way, we eliminated the effect of soil depth on SM in order to quantify and relate SM to the available crop water requirements at the six soil depth intervals (Equation 5.3).

$$SWS_i = 0.1 * SM * BD_i * h_i \tag{5.3}$$

where SM is the soil moisture, BD_i is the bulk density and h_i is the soil thickness of a particular soil depth interval.

5.2.3 Statistical data analysis

We tested both the spatial and temporal dependencies of SM variability individually, under two main scenarios: (1) SM as influenced by the local soil types, and (2) SM as influenced by the soil depth intervals. We explained the spatial variability of SM using SM_g and BD, Clay, Silt, Depth, LS and TWI (Table 5.1). In the temporal data analysis, SM variability was explained with SM_p, API and ET_o (Table 5.1). A set of least square mean algorithm (Lenth 2016), correlation coefficients, standard deviations and ranges were used to support the explanation of the causal relationships and interactions that existed between SM and the internal-external factors. Individually, we quantified how each internal and external factor influenced the variability of SM at the local soil types and at the soil depth intervals. For each case, spatial covariance was considered (Lark and Marchant 2018).

In addition, we used a set of linear mixed-effect models (LME) (Bates et al. 2014) to also relate the spatial (SM_g) and temporal (SM_p) measurements under the two main data analysis scenarios. In this way, the individual and interactive importance of the internal-external factors in the spatio-temporal measurements were further evaluated. In addition, this allowed us to account for long-term trends, time of measurement impact and seasonality in the *in situ* measurements (Bates et al. 2014; Hengl et al. 2018). We implemented the LME model under a two-stage approach: (1) an initial model (*Null*) that included only the random intercept for SM_p and (2) a full model (*Full*) which included the additive and interactive effects of key explanatory variables such as clay and silt contents. We executed the LME model under three model options: (1) varying intercept nested by the random effects of soil types and standard soil depth intervals, (2) varying intercept and slope by standard soil depth intervals while grouping by their respective soil types, and (3) intercepts that vary at different levels by the influence of soil type and the standard depth intervals. Under a 500 parametric bootstrap simulation, we estimated the fixed effect coefficients and associated profile 95% confidence intervals (CI) for the *Full* model using maximum likelihood estimation. Positive CIs indicated a significant influence on the relationship between SM_g and SM_p, and vice versa. In addition, the range of CIs indicated the reliability of the calibration LME model. The LME model selection was tested by the Akaike information criterion (AIC) (Akaike 2011). We estimated the variance explained by both the fixed and random effect using the conditional R^2 performance metric and the variance explained by only the fixed effects via the marginal R^2 performance metric. By these approaches, we quantified the 'goodness of fit' of our LME model. At all stages of the statistical analysis and graphical representations, we used R software (R Core Team 2017) and R packages $Lsmeans$ (Lenth 2016), $Mumln$ (Barton 2018) and $Lme4$ (Bates et al. 2014). R scripts and full datasets used to draw the conclusions for this study are open-sourced and are accessible on the DOI: 10.5281/zenodo.3662411

5.3 Results

5.3.1 Spatial variability of SM_g in the different soil types at different soil depths

Along the three main topographical units, mean SM_g ranged from 8.11 ± 1.14% (least square mean ± standard error) to 19.67 ± 2.35% with raw SM_g in the range of 0.97 to 42.74% (appendix Table S5.1). Toe slope soils recorded substantially higher SM_g followed by middle to lower slope soils and the upper slope soil. The highest and lowest mean SM_g were both recorded in the toe slope soils. Variability of SM_g was low in middle to lower slope (range = 3%) as compared to the toe slope (range = 12%). In addition, variability of SM_g in all soil types were low except in Siare series, where a change of 11.57% was recorded in mean SM_g (Figure 5.2). The mean of all soil series had their model CI above zero both at the soil type level and at soil depth intervals (appendix Figure S5.1). The lowest CI range (1.7%) was recorded in Lima series as compared to the highest of 9.3% recorded in Siare series.

The highest mean SM_g of 12.6% was recorded at 60 cm depth and the lowest (7.97 ± 0.86%) recorded at 30 cm depth. Mean SM_g along soil depth increased in the order: 30 < 10 < 40 < 20 < 60 cm (appendix Table S5.2). Within the 10–30 cm depth interval across all soil types, variability recorded in mean SM_g was < 6% except in Siare series, where a variability of 22% mean SM_g was recorded. At 40 cm depth, variability in mean SM_g ranged from 5.01% (Dagare series) to 42.74% (Siare series). SM_g in all soil types were responsive to changes in soil depth and increased with increasing depth. These increases in mean SM_g with increasing soil depth was prominent in the > 30 cm depth, particularly in Siare and Changnalili series (Figure 5.2). Mean SM_g in the lower depth (≥ 20 cm) was always higher than the upper depths (≤ 10 cm) except in Dagare series where the reverse was observed. Spatial variability of SM_g across the different topographical units with regards to depth intervals was high in the < 20 cm and low in the > 30 cm depth intervals, but changed conspicuously between the 20–30 cm depth interval (Figure 5.2). The mean SM_g at 60 cm depth was highly variable across the different soil types.

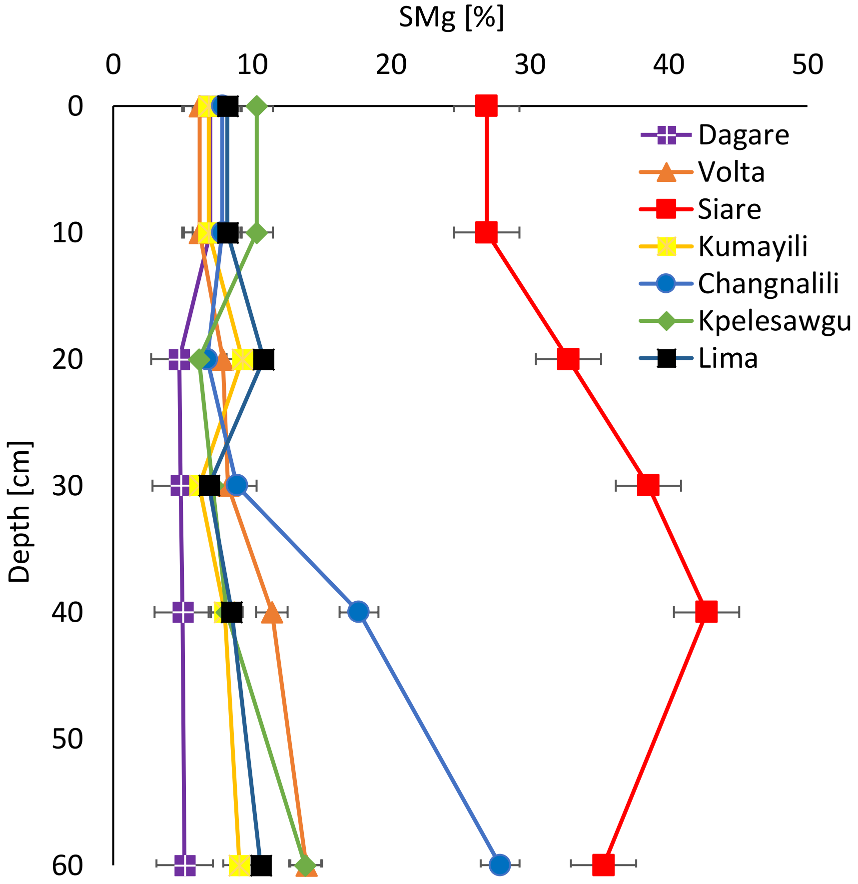


Figure 5.1. Soil profile distribution of mean soil moisture by gravimetric method (SM_g) at standard soil depths for the different soil types. Dagare = Gleyic Fluvisols, Volta = Fluvic Gleysols, Siare = Plinthic Lixisols, Kumayili = Chromic Lixisols, Changnalili = Petric Plinthosols, Kpelesawgu = Eutric Plinthosols and Lima = Gleyic Planosols.

5.3.2 Influence of internal factors on spatial variability of SM_g

5.3.2.1 Clay and silt contents

The particle size distribution of the analyzed soil samples distributed within eight main soil texture classes as illustrated in Figure 5.3. Most of the soil samples fell within the sandy loam, loamy sand and silty loam textural classifications which makes them sandy textured in general. Mean clay content (± standard error) ranged from 7.02 ± 2.07% to 16.19 ± 2.49%, with raw value range of 4 to 44% (appendix Table S5.1). Silt content also ranged from 14.3 ± 5.8% to 36.09 ± 1.27%, with raw value range of 4% to 60% (appendix Table S5.1).

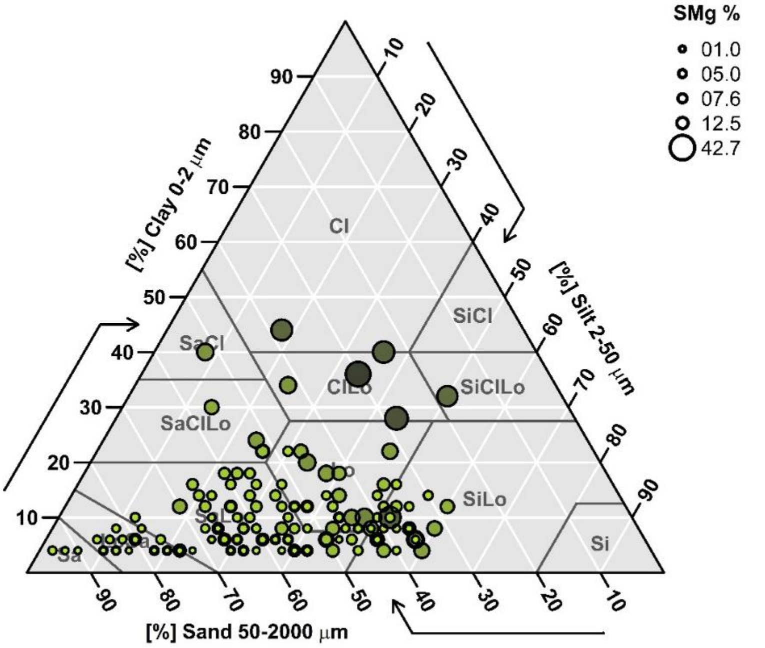


Figure 5.2. Particle size distribution of soil samples of the study area and their textural classes. Size of circles indicates each soil sample's SM$_g$.

Spatial variability of mean clay content was low (range = 9%) as compared to variability of silt content (range = 22%) in the different soil types (Figure 5.4). Compared with the toe slopes, mean clay and silt contents were lower in the middle to lower slopes followed by the upper slope soil. At the different depth intervals, clay and silt contents were inconsistent, as clay content increased with increasing soil depth, the reverse was observed with silt content (Figure 5.4). The highest mean clay content (14.79 ± 0.96%) was recorded at the 60 cm depth and the lowest (7.1 ± 0.88%) recorded at the 10 cm depth (appendix Table S5.2). At 30 cm depth, the highest mean silt content (30.30 ± 2.51%) was recorded and the lowest (27.68 ± 2.96%) recorded at 60 cm depth. Comparing the entire soil profile depth (i.e. 0–60 cm), spatial variability of clay and silt contents were similar (range = 7.7 and 7.4 respectively). Raw silt content values were always higher than raw clay content values at all standard depth intervals. All CIs of clay and silt contents were above zero, with silt content recording the largest CI (range = 13.6) as compared to the smallest in clay content (range = 11.3), particularly at the standard depth intervals (appendix Figure S5.2).

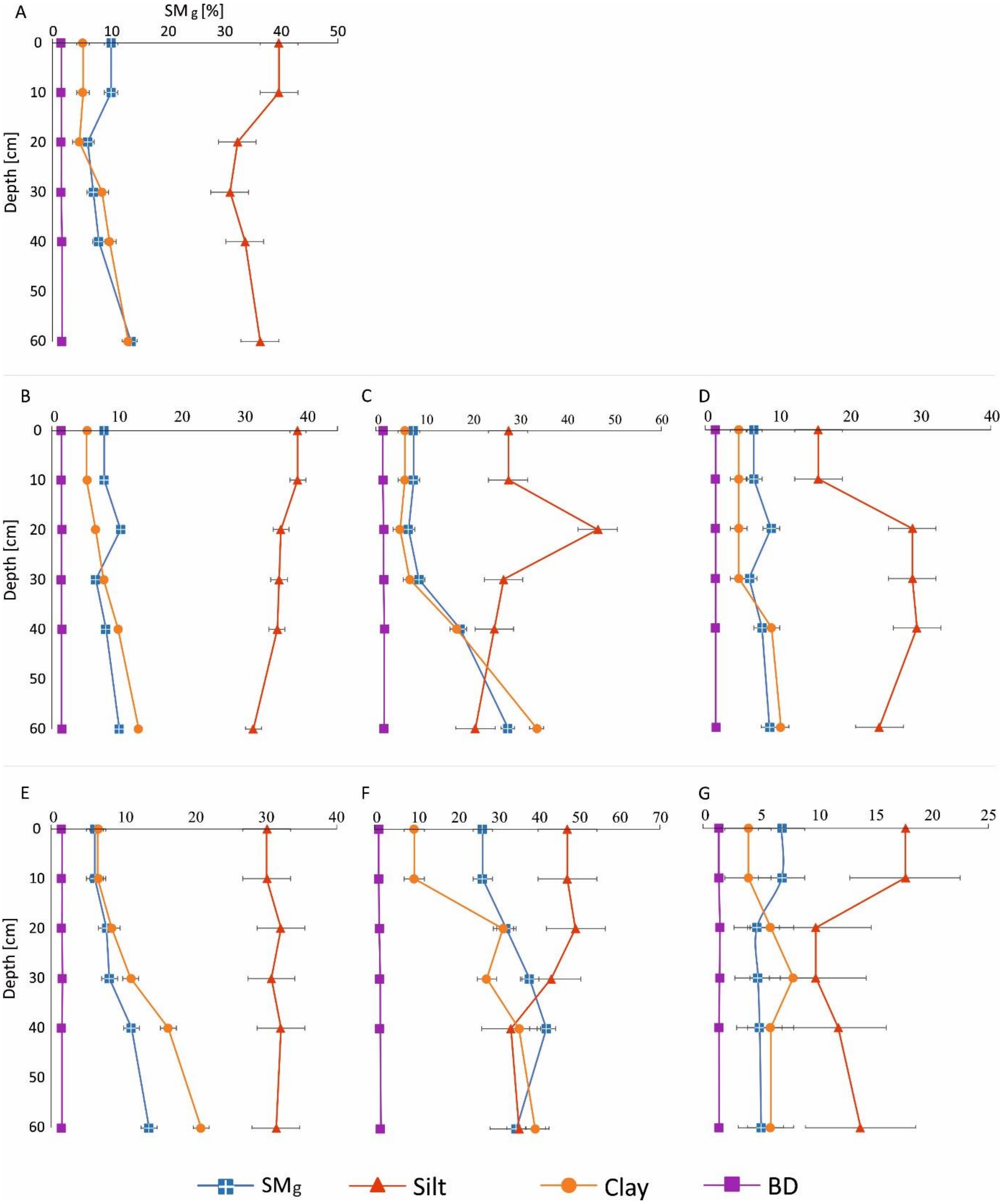


Figure 5.3. Soil profile distribution of mean SM_g, silt content (Silt), clay content (Clay) and bulk density (BD) at soil depths for the different soil types. On the upper slopes is Kpelesawgu series (A), middle to lower slopes includes Lima series (B), Changnalili series (C) and Kumayili series (D) and on the toe slopes also includes Volta series (E), Siare series (F) and Dagare series (G).

In total, clay and silt contents accounted for 91% of the investigated relationships between SM_g and the physical soil properties. Clay content accounted for 68% of the relationship between soil texture and SM_g, while 23% was accounted by silt content. At each change in clay content at a specific depth interval, a corresponding change in SM_g was also recorded (Figure 5.4). Thus, clearly marked consistent variabilities were observed between SM_g and clay content. In addition, at standard depth intervals where clay bulges were observed, a corresponding increase in SM_g was also recorded. SM_g and clay content increased with increasing depth across all soil types in the different topographical units (Figure 5.4).

5.3.2.2 Bulk density

Mean BD (± standard error) ranged from 1.38 ± 0.05 g cm^{-3} to 1.54 ± 0.03 g cm^{-3} (by soil type) and 1.42 ± 0.02 g cm^{-3} to 1.50 ± 0.02 g cm^{-3} (by soil depth) (appendix Tables S3.1 and S3.2 respectively). BD decreased along the topographical units in the order: upper slopes > middle to lower slopes > toe slopes. Spatial variability was high in the toe slope soils (range = 0.12 g cm^{-3}) as compared with the middle to lower slope soils (range = 0.1 g cm^{-3}). Mean BD increased with increasing soil depth. The highest mean BD (1.5 ± 0.02 g cm^{-3}) was recorded at 60 cm depth and the lowest (1.42 ± 0.02 g cm^{-3}) recorded at 10 cm depth (appendix Table S5.2). Spatial variability of BD was uniform from the upper to the lower soil depths of the soil profile across the different soil types (Figure 5.4). However, CI from the estimated means varied slightly with increasing soil depth (appendix Figure S5.2). Eleven percent of the investigated relationships between SM_g and the physical soil properties was accounted by BD.

5.3.3 Influence of LS and TWI on spatial variability of SM_g

Spatial variability was similar for both the LS (range = 1.35°) and the TWI (range = 3.9) (appendix Table S5.3) across the study area. High SM_g corresponded with high TWI across the different soil types of the different topographical units, however, the reverse was observed with LS (Figure 5.5). The highest LS (1.45 ± 0.2 °) was recorded in Kpelesawgu series and the lowest (0.1 °) recorded in Dagare series. Also, the lowest TWI (11.3 ± 0.1) was recorded in Kumayili series with the highest (15.2) recorded in Siare series. Comparing the topographical units, TWI was high in the order of toe slope soils > middle to lower slope soils > upper slope soil but the reverse was observed with LS.

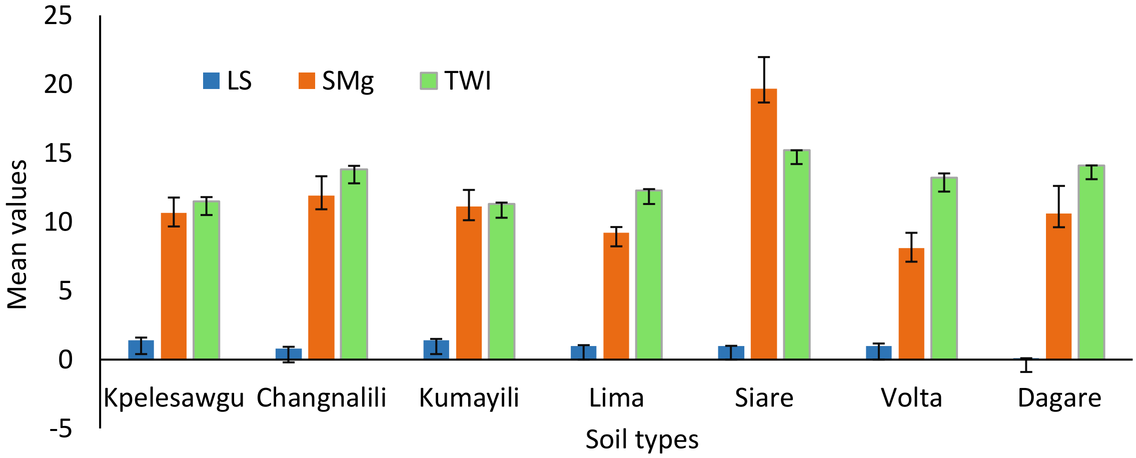


Figure 5.4. Comparing mean value of local slope (LS), topographic wetness index (TWI) and soil moisture (SM_g) in different soil types.

5.3.4 Soil profile characteristics of PAWC

The lowest and highest PAWC were recorded at 20–30 cm and 60–100 cm depth respectively. At the anthesis stage, PAWC ranged from 13.69 mm to 63.77 mm. At the maturity stage, PAWC ranged from 5.38 mm to 12.52 mm (appendix Table S5.4). PAWC exhibited a high-low-high trend from 0–100 cm at the anthesis and maturity stages (Figure 5.6). Decreases observed in PAWC in soils under both the anthesis and the maturity stages mainly occurred between the 20–40 cm depth intervals. Variability of PAWC down the soil profile was low (range = 3.5 mm) at the 0–40 cm depth and high (range = 28.6 mm) at the 40–100 cm depth.

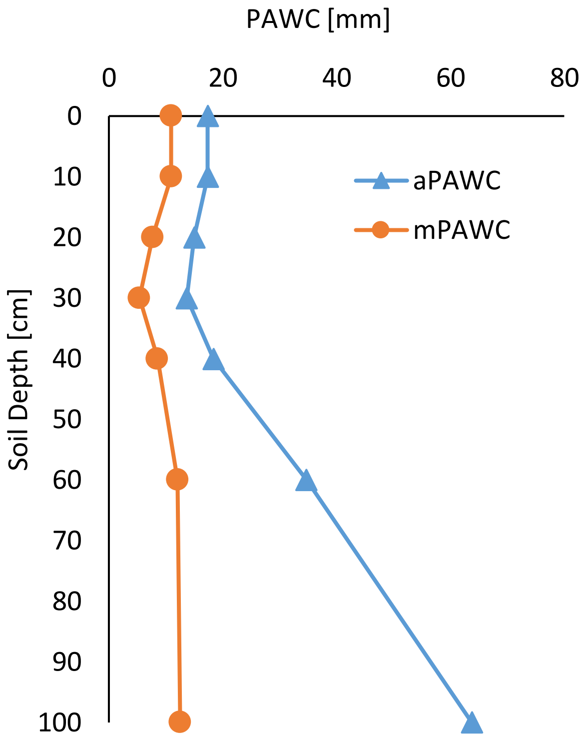


Figure 5.5. Soil profile distribution of PAWC in soils under a 90-day maize at 100 cm soil depth through a single growing season. PAWC was calculated from crop lower limits: 'a' – at anthesis and 'm' – at maturity stages of maize crop.

5.3.5 Relationships between SM_g and SM_p as influenced by soil type and depth

The LME model designed by intercepts varying at different levels by the influence of soil type and depths was selected as the optimal model option to relate SM_g to SM_p (ΔAIC = 5.7; *Full* model and 0.96; *Null* model) (Table 5.2). The *Full* model showed a better fit to the datasets with regards to the model's fixed and random effects (R^2 = 0.51; marginal and 0.74; conditional) as compared to the *Null* model (Table 5.2). The influence of clay and silt content and SM_p on the relationship between SM_g and SM_p indicated that the AIC of the *Full* LME model was 14 units more optimal than the AIC of the *Null* LME model. There were strong trends associated with a positive effect of clay and silt content on the relationship between SM_g and SM_p. Model estimates indicated that the relationship between SM_g and SM_p increased with increase in clay and silt contents (appendix Figure S5.3). Weaker interaction between clay content and SM_p decrease the strength of the relationship between SM_g and SM_p. All model estimates and their CIs were above zero and showed low variability within its fixed effects (range = 0.4). However, the interactive effect of SM_p and clay, recorded a slightly below zero model estimate (-0.003) with lower CI of -0.01 (appendix Figure S5.3). Overall, the selected LME model showed that the relationship between SM_g and SM_p is better explained when there is increase in clay and silt contents but with little or no interaction between SM_p and the fixed effects (in this case, clay and silt contents).

Table 5.2. Goodness of fit for mixed effect model options for models of relationships between SM_g and SM_p

Model	R^2		AIC
	Marginal	**Conditional**	
Null			
Model 1	0.27	0.70	1052.96
Model 2	0.30	0.71	1052.73
Model 3 (*)	0.31	0.70	1052.00
Full			
Model 1	0.43	0.72	1043.50
Model 2	0.50	0.74	1043.42
Model 3 (*)	0.51	0.74	1037.82

R^2 represents the mixed effect model variances explained by the fixed effects (marginal R^2) and model variances explained by both the fixed and the random effects (conditional R^2). *Null* = results of the model fit of the random effect only. *Full* = results of the model fit of the combined fixed and random effects and interactions. AIC = Akaike information criterion (the lower, the better). n = 170 and '*' represent selected LME model option.

5.3.6 Temporal variability of SM

5.3.6.1 Temporal variability of SM of the different soil types at different soil depths

Along the three main topographical units, differences in mean SM_p over the measurement period (February–June 2018) were observed in all soil types and also at the different standard depths (appendix Table S5.5). In the different soil types, Siare series recorded the highest mean SM_p at the different soil depths throughout the measuring period (Figure 5.7). However, the lowest mean SM_p changed with increasing soil depth across the different soil types. In order of increasing soil depths, the lowest mean SM_p was recorded by Kpelesawgu series at 10 cm, followed in the order: Kumayili series (20, 30 and 100 cm) and Dagare series (40 and 60 cm). SM_p increased with increasing soil depth throughout the measurement period.

The lowest mean SM_p (0.05%) was recorded in the second set of measurement at 10 cm depth with the highest (25.13%) recorded in the sixth set of measurement. Compared with the first set of measurement, SM_p at 10 cm was inconsistent with the other sets of measurements and increased in the order of $2^{nd} < 5^{th} < 3^{rd} < 9^{th} < 4^{th} < 8^{th} < 10^{th} < 7^{th}$ and 6^{th} (Figure 5.7A). At 20 cm depth, the lowest mean SM_p (0.13%) was also recorded in the second set of measurement and the highest mean SM_p of 39.48% recorded in the seventh set of measurement (Figure 5.7B). At 20 cm depth, mean SM_p increased in the order: 2nd < 1st < 5th < 3rd < 10th < 4th < 9th < 6th < 8th and 7th. Again, the second set of measurement at 30 cm depth recorded the lowest mean SM_p of 4.76% with the highest (47.1%) recorded in the sixth set of measurement. Mean SM_p increased in the order: 2nd < 1st < 5th < 10th < 9th < 3rd < 4th < 8th < 7th and 6th at the 30 cm depth (Figure 5.7C). Similarly, the lowest mean SM_p was recorded in second and sixth set of measurements respectively, at the 40 cm depth similar to the 30 cm depth. Mean SM_p increased in the order of 2nd < 5th < 1st < 10th < 3rd < 8th < 4th < 9th < 7th and 6th at the 40 cm depth (Figure 5.7D). At 60 cm depth, the trend for both the lowest and highest mean SM_p changed from the second and sixth set of measurements to the third and eighth respectively. The lowest mean SM_p recorded was 12.35% and the highest (67.55%) was recorded at the 60 cm depth. At the 60 cm depth, the order of increasing mean SM_p were: 3rd = 4th < 5th < 2nd < 1st < 9th < 8th < 7th < 6th and 10th (Figure 5.7E). At 100 cm depth interval, the lowest mean SM_p (58.9%) was again recorded in the second set of measurement as observed in the 10, 30 and 40 cm soil depths and the highest (64.9%) recorded in the first set of measurement. Also, the order of increasing mean SM_p at the 100 cm depth were: 2nd < 3rd < 5th < 4th < 8th < 7th < 10th < 9th < 6th and 1st (Figure 5.7F). Compared with the upper soil layers (< 20 cm), temporal variability was high in the lower depths (> 20 cm) throughout the measurement period. Temporal variability (range) was high in the order: 60 cm (55.2%) > 100 cm (48.9%) > 40 cm (43.93%) > 30 cm (42.84%) > 20 cm (39.34%) and 10 cm (25.08%).

Along the three main topographical units, mean SM_p increased with increasing depth, however, in Kpelesawgu and Kumayili series, lower mean SM_p was recorded at the 100 cm soil depth (appendix Table S5.5). The low mean SM_p recorded at the 100 cm depth was prominent from the second to the tenth set of measurements. Across the three main topographical units, the tenth set of measurement recorded the highest mean SM_p (onset of heavy rains), whilst the lowest mean SM_p was recorded in the first set of measurements (peak of dry season). Kpelesawgu series recorded the lowest mean SM_p in its top soils (i.e. 10 cm) throughout the measuring period as compared to the rest of the soil types, however, the reverse was observed in Siare series. Though incomplete set of measurements were obtained for Dagare series, its variability was similar to the rest of the toe slope soils (i.e. Siare and Volta series).

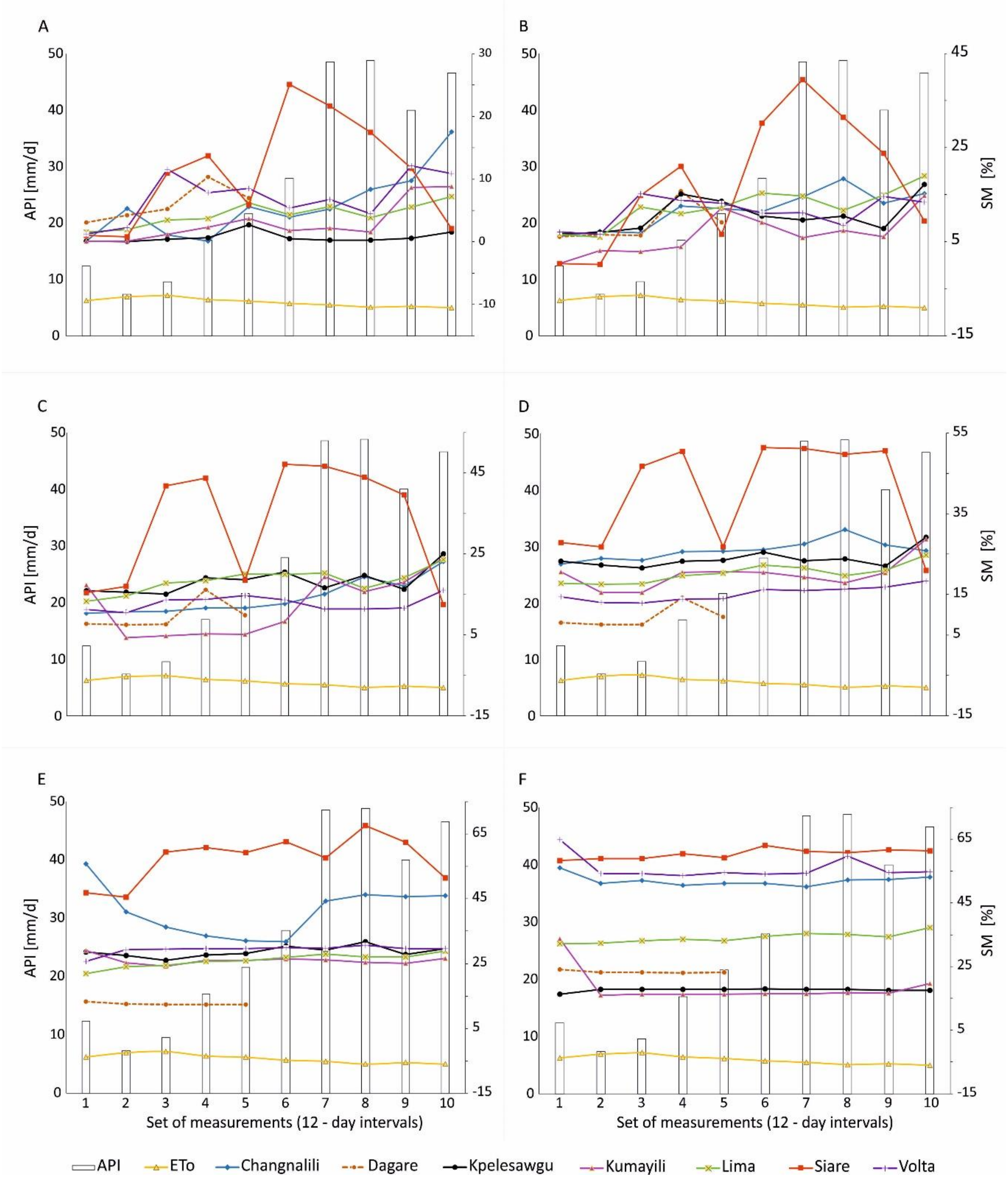


Figure 5.6. Temporal variability of SM_p at standard soil depths of 10 cm (A), 20 cm (B), 30 cm (C), 40 cm (D), 60 cm (E) and 100 cm (F) for the different soil types against antecedent precipitation index (API) and evapotranspiration (ET_o) from February to June 2018. On Dagare series, SM_p measurements were incomplete (i.e. only measurements from 1st to 5th) due to destruction of the installed access tube.

5.3.6.2 Influence of API and ET_o on temporal variability of SM

ET_o was about three times the daily mean precipitation (range = 4.9 mm day^{-1}), particularly during the measurement period of this study (Table 5.3). Daily mean precipitation also ranged from 0 mm day^{-1} to 8.08 mm day^{-1}. The highest API (48.80 mm day^{-1}) was recorded on the eighth set of measurement (15th May) whilst the lowest (7.35 mm day^{-1}) occurred on the second set of measurement (4th March). API peaked, in the order of the set of measurements: 8th > 7th > 10th > 9th > 6th > 5th > 4th > 1st > 3rd and 2nd before decaying on days with no precipitation (Table 5.3). High precipitation corresponded with high API and vice versa. Temporal variability of API was high (range = 41.45 mm day^{-1}) as compared with the temporal variability of ET_o (range = 2.15 mm day^{-1}) during the measurement period.

Table 5.3. Daily mean values of temporal antecedent precipitation index and evapotranspiration from 2010 — 2018 of the study area

Date (2018)	**Set of measurement**	**ET_o**	**Prep**	**API**
	12 – day interval		**mm day^{-1}**	
20th February	1	6.24	5.51	12.38
4th March	2	6.96	-	7.35
16th March	3	7.16	-	9.56
28th March	4	6.43	-	16.98
9th April	5	6.19	0.56	21.61
21st April	6	5.71	1.14	27.89
3rd May	7	5.50	1.20	48.56
15th May	8	5.04	8.08	48.80
27th May	9	5.25	-	40.02
8th June	10	5.01	0.15	46.59

ET_o = evapotranspiration, Prep = precipitation and API = antecedent precipitation index.

Although there were strong linear relationship between mean SM_p and ET_o, and, SM_p and API, variations in mean SM_p in the different soil types at the soil depth intervals were not always consistent with the variations in API and ET_o for our case study areas (Figure 5.7). Mean SM_p of the 10 cm depth of the different soil types highly varied with changes in API and ET_o. With SM_p and ET_o, there was a strong negative relationship ($R^2 = 0.77$) which was explained by the linear regression equation $SM_p = 7.06 – 0.01*ET_o$. With SM_p and API, a strong positive relationship ($R^2 = 0.83$) was exhibited and explained by the linear regression equation $SM_p = 2.9 + 0.12*API$. From the sixth set of measurements, where a substantial increase and decrease were recorded in API and ET_o respectively, mean SM_p across the different soil types also increased, particularly, in the lower depths (≥ 20 cm). With regards to the < 100 cm depth intervals, variability in mean SM_p at 100 cm depth was steady in all soil types, particularly from the second to the tenth set of measurements (Figure 5.7). At 100 cm, SM_p was slightly influenced by the variabilities of API and ET_o.

5.3.6.3 Influence of API and ET_o on temporal variability of SWS

SWS was consistent with the variations of mean SM_p across the different soil types and at the different depth intervals. Inconsistencies reported earlier on the variability of mean SM_p, as influenced by variations of API and ET_o, were different for the temporal variability of SWS. SWS, which is highly influenced by layer thickness, increased with increasing API and decreasing ET_o (appendix Table S5.6). SWS increased with increasing depth across the different soil types. Comparing the three topographical units, SWS was on average highest in the toe slope soils (Siare > Volta > Dagare series), followed by the middle to the lower slope soils (Lima > Changnalili > Kumayili series) and the upper slope soil (Kpelesawgu series). Low and high SWS across the different soil types at different standard depth intervals were consistent with temporal variations of API and ET_o. With regards to SWS, soils in the middle to lower slope followed by the upper slope soil were very responsive to changes in API and ET_o as compared to toe slope soils. In upper slope soil (Kpelesawgu series), SWS ranged from 0.08 mm (early March; peak of dry season) to 110.5 mm (end of April; unset of rains). In the middle to lower slope soils, the highest SWS (222.96 mm; Lima series) was recorded in the early June (10th set of measurement) and the lowest (0.11 mm) recorded in both Kumayili and Changnalili series in the late February and March (1st and 4th set of measurements respectively). In the toe slope soils, the highest SWS of 388.88 mm was recorded in Volta series in late February (1st set of measurement; peak of dry season) and the lowest (1.12 mm) recorded in Siare series in early March (2nd set of measurement). Siare series recorded a substantial increase in SWS within the 10–20 cm depth interval from the 5th to 9th set of measurements. This increase in SWS corresponded with an increase in API from the 5th (April; on set of rains) to 9th (late May) set of measurements (Table 5.3). Temporal variability was always high in the bottom soil layers as compared to the top soil layer of the different soil types. However, low temporal variability was observed with increasing set of measurements throughout the SM_p measurement period.

5.4 Discussion

5.4.1 Spatial and temporal variability of SM within the soil landscape

SM was low only in the sandy upper parts of the soils (≤ 10 cm), while in the lower parts of the soils (≥ 20 cm), clay content increased with increasing soil depth and kept SM for longer periods. Wherever there was clay build ups and bulges within the soil profiles, a corresponding increase in SM was observed. Throughout the dry season, the lowest SM (found at all soil depths) occurred in the beginning of March – peak of dry season, whereas the highest SM occurred in the end of April – unset of rains.

SM is primarily controlled by soil texture (Korres et al. 2015), and thus the sandy upper parts of the soils supported decreased water holding capacity and loss of SM under drought conditions (high ET_o and low API). However, as clay content increased in the lower parts of the soils, SM increases even under similar drought conditions. Baskan et al. (2013) demonstrated similar observations. Despite the variability of SM was highly driven by the dynamics of soil texture, our results showed some general inconsistencies in the variability of SM on Dagare and Siare series. Siare series is developed over colluvio-alluvial sediments in depression bottoms and valleys along the Volta river of the Guinea savannah zone (Adu 1995). The proximity of Siare series to the Volta river is a strong influence on its SM dynamics even in the dry season. It was, therefore, consistent that SM increased in Siare series throughout the dry season irrespective of its soil texture. However, Dagare series is developed over periodically deposited alluvial sediments on the banks of small tributaries of the Volta river (Adu 1995). These tributaries may dry-up or recede during the drought season. Such situations suggest that the influence of the Volta river on the SM dynamics of Dagare series is less as compared to those of the Siare series. Thus, under similar conditions, SM of Dagare series decreased as compared to those of Siare series. This observation makes Dagare series behave like the upper slope soil (i.e. Kpelesawgu series).

Our findings indicate that BD is a critical soil property that contributed to the high SM in the lower parts of the soils. This is because high BD affects soil infiltration rates (Li et al. 2009). Our results show that the ideal BD of soils of the study area facilitate root growth, SM infiltration and percolation and, can also support water uptake by plant roots (Hazelton and Murphy 2016; Weil and Brady 2016). Hence, we suggest that the consistent high SM in the lower soil depth of the different soil types are probably due to the ideal BD observed for our case study areas, making infiltration process driven by gravity drainage easier.

From the field soil sampling campaign, inherent petro-plinthite layers were observed in the lower soil depths of the different soil types at 70–100 cm, as also reported in Asiamah and Dedzoe (1999). Due to this petro-plinthite layer, SM in the above 100 cm depth is cut off from ground water. Consequently, water seepage beyond the 100 cm is inhibited and almost impossible percolation (Asiamah and Dedzoe 1999). In inference, SM stagnates above the 100 cm depth of the different soil profiles. This observation explains why our results always showed consistent high SM in the lower soil depths of the different soil types. Furthermore, Asiamah and Dedzoe (1999) suggested that SM is mainly recharged through rainfall events and river outflow. This observation also corroborates our results, which showed that high SM was always consistent with high precipitation and API.

Our findings show that external factors had a site-specific localized influence on the spatial and temporal variability of SM, particularly along the three main topographical units. Our findings extend those of Baskan et al. (2013), Gutierrez-Lopez et al. (2014) and Wang et al. (2016). The influence of the external factors on SM variability is high in soils with subtle LS and high TWI (i.e. toe slope soils) and vice versa (i.e. upper slope soils). The reason is that lateral SM flow driven by topographic positions influences SM variability, especially, when the different soil types record uniform soil texture at their depth intervals (Kim 2012; Huang et al. 2016). This observation further indicates that the uniform soil texture supports the lateral flow of SM in the study area, similar to what was demonstrated by Kim (2012) and Huang et al. (2016). SM in the toe slope soils are generally expected to be higher than SM in the middle to lower slope and the upper slope soils. Comparing the influence of internal and external factors on the spatial and temporal variability of SM, our findings reveal that localized internal factors such as clay and silt contents, and BD highly influence SM variability than external factors such as LS and TWI (Biswas and Si 2011; Baskan et al. 2013; Gwak and Kim 2017).

In the absence of rainfall, ET_o is one key means by which SM returns to the atmosphere (Daly and Porporato 2005). On purview of this, the 0–10 cm depth interval showed varying results in each SM measurement day through the dry season. The possible explanation is due to the sandy top soils. Thus, at high ET_o and low API, the upper parts of the soils of the study area easily lose SM (Daly and Porporato 2005). In addition, the high variability of SM in the top soils can be implicated by the abundant elephant grass' roots. A slight increase in API promotes root water uptake and might have contributed to these SM fluxes (Seneviratne et al. 2010). In inference, the upper parts of the soils (≤ 10 cm) supports low percolation at high ET_o and low

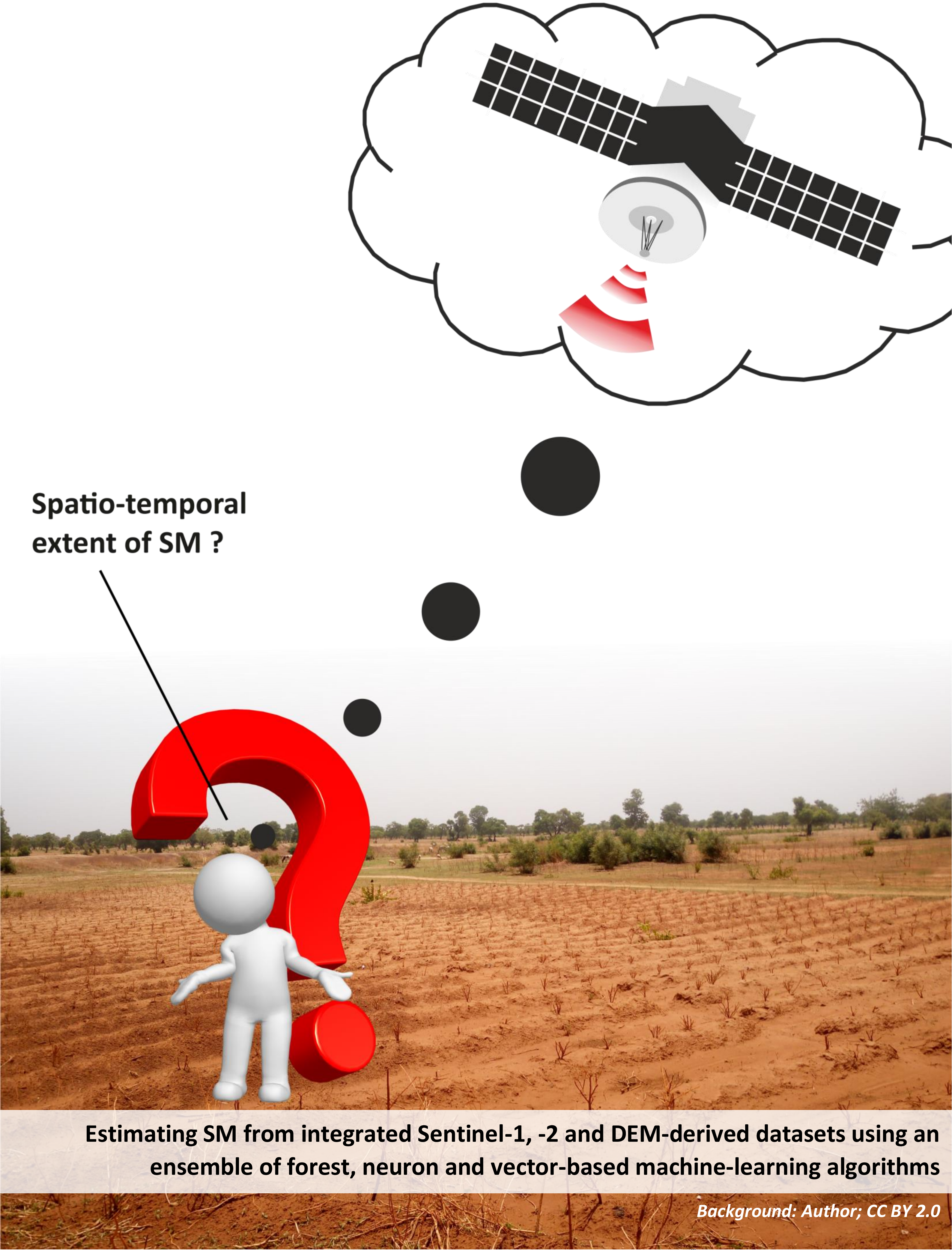
API. In contracts to the top soil layer, SM is highly influenced by only API in the lower parts of the soil profiles (≥ 20 cm). Thus, with continual increase in API coupled with the sandy top soils, SM tends to drain easily by gravity. This reason explains why high SM correlated with high API in the lower parts of the soils throughout the measurement period, especially when gradual increases were observed with API. Furthermore, the upper parts of the soils served as a protection layer (i.e. a mulch) for the lower parts of the soils in order to prevent the loss of SM. This observation promoted infiltration into the far-lower parts of the soils during and after rainfall events (at high API). Similar observations were reported for the landscapes of Watkinsville, United States by Franzluebbers (2002). The low and high SM trends in the upper and lower parts of the soils respectively, of the soil profiles also corroborate this finding. While the sandy upper parts of the soils protected loss of SM and supported infiltration into the lower parts of the soils, clay content build-up in the lower parts of the soils kept SM for longer periods. Our results further reveal that the continual high API increases infiltration and percolation to the very lower depths of the different soil types. This observation shows, particularly, from end of April when the study area started recording some amounts of precipitation.

While we demonstrated the spatial and temporal variability of SM, some observations on the capability of the PR2/60 moisture probe in the SM measurement of the area are important to make. Our findings from investigating the relationship between SM_g and SM_p suggest that accurate SM measurement can be achieved in clay enriched-moist soil types with the PR2/60 moisture probe. Conversely, soil types with high sand content and at drought conditions were less sensitive to the PR2/60 moisture probe measurement. This observation was shown on Kpelesawgu series (dry and sandy textured) and Siare series (moist and clay textured). In addition, loose soil structure, air pockets and coarse materials in the coarse textured soil types, particularly Kpelesawgu and Changnalili series, negatively influenced the performance of the PR2/60 moisture probe. Alternatively, using a combination of the PR2/60 moisture probe and integrating its measurements with the gravimetric method, as demonstrated in this study, provided accurate estimation of SM. In this way, robust spatio-temporal SM monitoring exercise can be achieved to support SM prediction activities.

5.5 Implications of study for dry season farming and conclusion

The main outcome of this work is that almost all soil types in the study area (except for the Kumayili soil series) below 10 cm depth have SWS potentials that match the water requirements of at least some drought-tolerant crops (> 9 mm day^{-1}) (Critchley et al. 1991). Kumayili series is limiting because crop adequate SWS levels can be assessed at > 30 cm depth, which is not ideal planting depth for notable crops grown in the Guinea savannah zone (Sheaffer and Moncada 2012). Although we investigated the possibility to cultivate maize in the study, our findings indicated that the SWS at depth at which maize can be sown is inadequate to support its cultivation (Molatudi and Mariga 2009). However, there are still potentials for some other staple crops such as cowpea, cassava, and potatoes that can be cultivated at soil depths where the crop water requirements can be met (Kumar et al. 2015; Ogundare 2017; Buakum et al. 2018). These crops also have the potential to tap into such SWS rich zones. Amongst the potential crops, cowpea (*Bengpla* cultivar) (Asafo-Adjei and Singh 2005) has the shortest maturity duration (i.e. 52–60 days). In addition, *Bengpla* cultivar has a potential yield of 1.5 t ha^{-1} and susceptible to a variety of disease reported in the Guinea savannah zone (Asafo-Adjei and Singh 2005). From Critchley et al. (1991), legume's crop water requirements average at 6.32 mm day^{-1}, which our results provide the evidence that the lower parts of the soils (≥ 20 cm) of the area have SWS > 9 mm day^{-1} throughout the dry season. Consequently, we make a research proposal for the cultivation of *Bengpla* cultivar particularly from the middle of March to end of April where the rate of ET_o successively decrease with a corresponding increase in API throughout the dry season. Soil fertility improvement is another important trade-off from the proposed legume crop rotation (Kermah et al. 2017). Residual effects from cowpea is important for the successive rainy season food-crop production (Bohlool et al. 1992; Belane and Dakora 2009), particularly for maize (Kermah et al. 2017). For instance, cowpea can fix large amounts of net atmospheric nitrogen to improve the fertility of soils of the area (Yusuf et al. 2009; Schilt-van et al. 2017). In addition, studies have suggested that sequential cropping, for example cassava and cowpea, can improved soil fertility to the point where applications of manure and mineral fertilizer could be reduced up to 50%, with no yield loss (Adjei-Nsiah et al. 2007; Adjei-Nsiah 2008; Ezui et al. 2016). Hence, our research-proposed legume-crop rotation is an important practice to maintain soil fertility of soils for the local communities of the Guinea savannah zone. The cowpea grain can be use as food and nutritional supplements as suggested by Abass (2015). Cowpea residue after crop harvest can also be used to improve mulching or used as feed for livestock.

This study is the first to demonstrate how SM varies throughout the dry season of the Guinea savannah zone of Ghana. SM from 34 soil profiles at six soil depths (0–100 cm) were intensively measured throughout the dry season to investigate its spatial and temporal variability. This paper investigated and demonstrated the multi-factor relationships, as well as direct and indirect interactions between SM and the internal-external factors. Throughout the dry season, SM was low only in the sandy upper parts of the soils (≤ 10 cm). In the lower parts of the soils (≥ 20 cm), clay content increased with soil depth, and SM was kept for longer periods. Localized internal factors such as clay and silt content and BD had stronger influence on the spatial and temporal variability of SM than external factors such as LS, TWI, ET_o and API in the study area. BD observed in this study imply good conditions for root developments, SM infiltration and percolation within the soil profiles. Temporally, ET_o influenced SM only in the sandy top soils while API strongly influenced SM at ≥ 20 cm soil depth. A positive and negative correlation existed between API and SM and, ET_o and SM respectively. Thus, these observation gives an insight to improve soil surface mulching in order to conserve SM in the profiles of the soils for food-crop production, particularly, in the dry season. We conclude that most benchmark soils of the area hold the potential for implementing a dual farming system that can enhance agricultural productivity and increase the income of local farming communities. Future studies are, however, needed to validate our recommendations for the Guinea savannah zone of Ghana.



Estimating SM from integrated Sentinel-1, -2 and DEM-derived datasets using an ensemble of forest, neuron and vector-based machine-learning algorithms

Background: Author; CC BY 2.0

Chapter 6: Estimating SM from integrated Sentinel-1, -2 and DEM -derived datasets using an ensemble of forest, neuron and vector-based MLAs

This manuscript is under review as:
Nketia KA, Erasmi S, Sauer D. Estimating soil moisture from integrated Sentinel-1, -2 and DEM-derived datasets using an ensemble of forest, neuron and vector-based machine-learning algorithms.

Conference presentation at Living Planet Symposium 2019, Milan:
A4.12: HR Soil Moisture Poster Session, Area C - Board 247. https://lps19.esa.int/NikalWebsitePortal/living-planet-symposium-2019/lps19/Agenda/AgendaItemDetail?id=c0a22eed-5857-4f99-a55e-d8204e419ad3

Abstract

Inconsistent algorithm performances and optimal model selection challenges persist in estimating soil moisture (SM) from satellite images with machine-learning algorithms (MLA). In this paper, we propose an innovative ensemble-based model (EBM) that couples random forest regression (RFR), support vector regression (SVR) and artificial neural network (ANN) to estimate SM from integrated Sentinel-1, -2 and auxiliary datasets for the semi-arid landscapes of Ghana. In order to train and validate the SM estimates, we used *in situ* SM (SM_p), measured with a calibrated PR2/60 moisture probe (0–10 cm depth) at 34 locations on key benchmark soils from February to June 2018 (n = 340). Validation locations were stratified with a hybrid methodology that coupled the global weighted principal component analysis and a cost-constraint conditioned Latin hypercube algorithm. At the low to extremely low SM_p levels, Sentinel-2 derived indices showed higher sensitivity to SM_p than Sentinel-1 derived variables alone. While Sentinel-2 indices accounted for specific clusters of measured SM_p, Sentinel-1 variables rather accounted for the entire variability of SM_p and effects due to incidence angle. In addition, terrain-derived variables disentangled ambiguities caused by landscape heterogeneities. With the optimal model input variables, the MLAs estimated SM with *RMSEs* (Vol. SM_p): 0.18% (ANN), 1.14% (SVR) and 0.15% (RFR), while their ensemble estimated SM with an error of 0.12%. One main outcome indicates that ANN estimated “false” SM at minimum thresholds of 0.38% Vol SM_p. Overall, the EBM rather presented the optimum alternative between over- or underestimated SM and “false” estimated SM as compared to the individual MLAs. A close agreement is rather observed between SM_p and estimated SM via the EBM and thus, highlights a viable alternative to estimate SM than using individual MLAs.

Keywords: artificial neural network, auxiliary datasets, ensemble-based machine learning, random forest regression, Sentinel-1, Sentinel-2, soil moisture, support vector regression

6.1 Introduction

Due to the importance of spatio-temporal soil moisture (SM) information in various environmental processes (Srivastava et al. 2016), the Global Climate Observing System through the Climate Change Initiative program has recently identified SM as an Essential Climate Variable and as a High Priority Variable (Wagner et al. 2012). Unfortunately, the estimation of SM through classical point-based measurements for use in environmental process applications is challenging. This is because, not only are point-based measurement stations sparse in nature across the globe, but also, are time and labor intensive, expensive and practically impossible at spatio-temporal global scales (Liujun et al. 2019). These direct-field measurements are moreover impractical due to high spatio-temporal variability of the target variables at these global scales and requires repetitive measurements to estimate temporal dynamics.

An increasing body of studies have recognized and identified the combination of point-based measurements and remote sensing (RS) as an alternative to up-scale SM information on spatio-temporal global scales (Ali et al. 2015; Peng and Loew 2017; Liakos et al. 2018). For this reason, active and passive RS instruments in the microwave domain have frequently been coupled with point-based measurements for SM estimation (Rosich et al. 2003; Entekhabi et al. 2010; Liujun et al. 2019). Despite their high temporal resolution (i.e. 1–3 days), which makes them highly suitable for SM estimation, they rather operate and deliver images at low spatial resolution (~50 km) (Mohanty et al. 2017). Hence, their spatial resolution limits its usage at local-field scales. The estimation of SM at local scales is critical to assess and manage SM resources in semi-arid landscapes, which is essential to support regional, national and global SM monitoring initiatives (Wigneron et al. 2003; Colliander et al. 2017). In addition, one success of SM estimation from RS is largely driven by the quality and the ability to detect small changes at such scales necessary to support agriculture (Gao et al. 2017). In view of this bottleneck, advances made in SM estimation from ESAs-Sentinel-1 and -2 missions offers a balance between spatial and temporal resolutions (Torres et al. 2012). Therefore, Sentinel-1, -2, and other related upcoming satellite sensors, such as the NASA-Indian Space Research Organization Synthetic Aperture Radar (NISAR) mission, opens new perspective for estimating SM at detailed spatio-temporal global scales (Agrawal et al. 2016).

Estimating SM from microwave sensors is challenging and not straightforward (Ali et al. 2015; Liujun et al. 2019). The challenges are mainly due to ambiguities that affect the microwave sensor's signal, the high variability both in time and space of target variables, and the non-linearity and complexity of the estimation (Pasolli et al. 2014). Hence, expected accuracies in SM estimation are hardly achieved (Zribi et al. 1997; Alexakis et al. 2017). For instance, the variability of physic-chemical properties of target variables determine the absorption, transmission, emission and reflection of the microwave and its parameter ambiguity (Ulaby et al. 1982; 2014). Consequently, this results in a complex and non-linear relationship between target variables and RS measurements (Haboudane et al. 2004). These notwithstanding, Ali et al. (2015) and Liakos et al. (2018) pointed out additional ambiguities that affect the microwave signal backscattering. Among others are those caused by site-specific landscape heterogeneities such as SM, surface roughness, vegetation cover and topography, which makes the process of distinguishing backscattering coefficient from SM difficult. It is worth noting that SM estimation is heavily driven by surface roughness, particularly when monitoring changes over global scales, and it is one of the most difficult parameters to estimate (Loew et al. 2006). It is therefore key to consider advanced estimation approaches that can account for these multiple non-linearity and complexities associated in the SM estimation.

Against this backdrop, the use of physical models such as the water cloud model, change detection methods and the advanced integral equation models have proven to be efficient (Graham and Harris 2003; Piles et al. 2009; Choker et al. 2017). In addition, non-parametric and parametric models (such as artificial neural network, random forest and regressions) have also been used extensively to address the aforementioned challenges (Ali et al. 2015; Amazirh et al. 2018). However, due to observational incidence angle and landscape heterogeneities such as SM, surface roughness, soil properties and vegetation cover, these models have a certain range of validity (Fung et al. 1992; Karam et al. 1992; Amazirh et al. 2018). In this paper, we focus on the use of non- and parametric models in SM estimation (referred here as machine learning algorithms (MLA)). With MLAs, some added-on benefits are however achieved: (1) fast to implement without strict data analytical programming (Liakos et al. 2018), (2) the ability to learn from training data and these multiple non-linearity and complexity associated in the SM estimation (Ahmad et al. 2010; Ali et al. 2015; Liakos et al. 2018), and (3) provides improved approximations in non-linear systems (Breiman 2001; Bisgin et al. 2018). Given that SM affects radar backscatter, stability of SM patterns leads to stability in radar backscatter patterns (Wagner et al. 2008). Hence, at low and extremely low SM levels, radar energy penetrates the soil, resulting in low radar backscatter at the satellite system (Ulaby et al. 1982; 2014). It is therefore expected that at such low SM conditions, radar backscatter is also highly influenced by the effects of landscape heterogeneities (Gao et al. 2017). Consequently, the use of MLAs, which is highly recommended for non-linear systems and is also a promising approach for operational SM estimation, would be an effective alternative to learn from the complex remote-sensing landscape interactions (Breiman 2001; Pasolli et al. 2015). Moreover, MLAs can adequately discriminate various SM levels and disentangle many associated ambiguities in the SM estimation (Pasolli et al. 2012; Liakos et al. 2018). These notwithstanding, MLAs can adequately and easily integrate large amounts of data sources necessary to discriminate and estimate SM (Liakos et al. 2018). For example, terrain-derived variables, from digital elevation model (DEM), are essential to disentangle ambiguities due to topography and surface roughness (Pasolli et al. 2012; Pasolli et al. 2014). Whereas soil and water indices from optical images (e.g. Sentinel-2), can also distinguish different SM levels (Torres et al. 2012).

Current research directions indicate rising interest in the use of MLAs, however, very few guidelines are available for their use in SM estimation (Liakos et al. 2018). Thus far, no study has also presented or tested the minimum *in situ* SM thresholds at which the widely used MLAs are efficient to estimate SM, particularly for semi-arid landscapes. In recent years, MLAs that have been frequently used in SM estimation from RS have mainly been compared, where the optimized and the efficient MLA was selected for estimating the target variable (Ali et al. 2015). The most widely used MLAs include artificial neural network (ANN), support vector regression (SVR) and random forest regression (RFR) (Liakos et al. 2018). However, contradictory algorithm performances and model selection challenges have been frequently reported (Pasolli et al. 2015; Holtgrave et al. 2018; Liakos et al. 2018). In addition, only in very few studies, MLAs have been combined, to address their individual inefficiencies in order to optimize SM estimation (Liakos et al. 2018). For example, Kumar et al. (2018) reported conflicting algorithm performances for both RFR and SVR under different agricultural fields when they compared RFR, SVR and ANN in estimating SM from Sentinel-1. In comparing SVR, ANN and multivariate linear regression model, higher performances were also shown by the SVR over the others (Ahmad et al. 2010). Pasolli et al. (2011) suggested a similar conclusion when SVR was compared with ANN in estimating SM. The SVR, using a multiscale active radar images, underscored a strong relationship between predicted and measured SM (R^2 = 0.8–0.9)

(Pasolli et al. 2015). Although ANN was demonstrated as the least performing algorithm in some SM estimation studies (Ahmad et al. 2010; Kumar et al. 2018), Notarnicola et al. (2008) and Paloscia et al. (2013) rather reported a greater compromise and trade-off between SM estimation processing time, stability and estimation accuracy with the ANN. In Paloscia et al. (2013), an accuracy of < 4% Vol. SM was reported with the ANN when SM algorithm and preliminary validation was undertaken for the ESA-GMES Sentinel-1 characteristics. Similarly, a high coefficient of determination ($R^2 = 0.9$) was obtained with the ANN when Alexakis et al. (2017) estimated SM for use in a hydrologic modelling system. Though these individual MLAs have proven to be efficient and robust to estimate SM at high accuracy, even at plot scales (Attarzadeh et al. 2018), it is critical to establish the minimum *in situ* SM thresholds at which these MLAs are efficient to estimate SM. Furthermore, it is important to investigate their combined performance in order to optimize the overall SM estimation. For instance, it was demonstrated that using an ensemble-based model (EBM) rather improve the predictive performance of the model fitting techniques due to the linear combination of the base MLAs (Liakos et al. 2018). Similarly, improved prediction results were demonstrated for sub-Saharan Africa (Hengl et al. 2017b) as compared to earlier predicted soil property maps (Hengl et al. 2015). That notwithstanding, the use of EBM was proven to deal with complex high-dimensional data (Piao et al. 2014), which is similar to the SM estimation.

On purview of the above, this study fills the gaps – we evaluate and validate an EBM to estimate SM from the synergy of Sentinel-1, -2 and other auxiliary datasets. In addition, we test the minimum *in situ* SM thresholds at which the widely used MLAs are efficient to estimate SM. RFR, SVR and ANN are ensembled in this study. In this EBM, we integrate key auxiliary datasets that compounds the ambiguities and complexities in the SM estimation. This builds on the integrated radar-optical-DEM datasets to utilize the high spatio-temporal resolution of Sentinel-1 and -2 missions. We extend, modify and ensemble the methodologies of Paloscia et al. (2013) and Alexakis et al. (2017) (ANN), Pasolli et al. (2014; 2015) (SVR) and Kumar et al. (2018) (RFR). Additionally, we investigate whether the EBM can improve the individual inefficiencies of the base MLAs in order to decrease the overall SM estimation error. We analyze the testing of the minimum *in situ* SM thresholds and the technical implementation of the proposed EBM, its validation and variability, for the semi-arid landscapes of Ghana.

6.2 Methods

6.2.1 Sentinel-1 data acquisition and processing

We obtained Level-1 ground range detected Sentinel-1A image product with 10 m pixel spacing from ESAs Copernicus open access hub (https://scihub.copernicus.eu/dhus/#/home) (Table 6.1). Corresponding Sentinel-1B image products were not available for the study area timely consistent with SM_p measurements. The images used were detected, multi-looked and projected to ground range. For our choice of Sentinel-1 overpass times, we used Interferometric Wide Swath Mode image products in dual polarization (i.e. VH and VV) with incidence angles between 38.3° and 45.7° in ascending orbit.

We processed the image products with ESAs desktop Sentinel Application Platform (SNAP) version 6.0.5 with the use of the Sentinel-1 Toolbox version 6.0.5. Processing was in two main stages: 1) radiometric correction and 2) geometric corrections. Prior to the two-stage image processing, we applied an orbit file with a three-degree polynomial in order to update the orbit state vectors in the abstract metadata of the image products.

Table 6.1. Sentinel-1 and -2 acquisition dates

No.	Sentinel-1A			Sentinel-2A		
	Acquisition dates (mm/dd/yy)	Sensing start time (UTC)	Polarization	Acquisition dates (mm/dd/yy)	Sensing start time (UTC)	Band
1	02/20/2018	6:18:54 PM	VH, VV	02/21/2018	10:20:31 AM	1–12, 8A
2	03/04/2018	6:18:54 PM	VH, VV	03/03/2018	10:20:21 AM	1–12, 8A
3	03/16/2018	6:18:54 PM	VH, VV	03/13/2018	10:20:21 AM	1–12, 8A
4	03/28/2018	6:18:54 PM	VH, VV	03/23/2018	10:20:21 AM	1–12, 8A
5	04/09/2018	6:18:55 PM	VH, VV	04/12/2018	10:20:21 AM	1–12, 8A
6	04/21/2018	6:18:55 PM	VH, VV	04/22/2018	10:20:31 AM	1–12, 8A
7	05/03/2018	6:18:56 PM	VH, VV	05/02/2018	10:20:31 AM	1–12, 8A
8	05/15/2018	6:18:56 PM	VH, VV	05/12/2018	10:20:31 AM	1–12, 8A
9	05/27/2018	6:18:57 PM	VH, VV	05/22/2018	10:20:31 AM	1–12, 8A
10	06/08/2018	6:18:58 PM	VH, VV	06/11/2018	10:20:21 AM	1–12, 8A

No. = Sets of measurements from February to June 2018.

In the radiometric correction (Figure 6.1), we removed thermal noise in each polarization with the use of a bi-linear interpolation algorithm. The image products were then calibrated to the backscatter coefficient σ^0 (appendix Equation S6.1). Prior to geometric correction, we co-registered the images and applied multi-temporal speckle filtering to reduce speckle noise in the images, which are originated by the coherent nature of Sentinel-1 sensor (Quegan and Yu

2001; Ulaby et al. 2014). This filter requires the estimation of the local mean intensity, where various spatial filters can be used (Quegan and Yu 2001; Schlund et al. 2018). We used a Lee Sigma filter with Sigma = 0.9 and a target window size of 3 x 3 for the estimation of the local mean intensity.

In the geometric correction (Figure 6.1), we mainly corrected for distortions in the images due to topographical variations and tilt of topographic slopes with respect to the Sentinel-1 sensor via the Range Doppler Terrain Correction operator of SNAP. We used a bicubic interpolation-resampling algorithm for both the images and the Shuttle Radar Topographic Mission (SRTM) DEM with 1 arc-second resolution.

Radiometric and geometric corrected outputs obtained from the image processing included σ_{VH}^0 and σ_{VV}^0 (in dB; appendix Equation S6.2), DEM and incidence angles (i.e. local and ellipsoid).

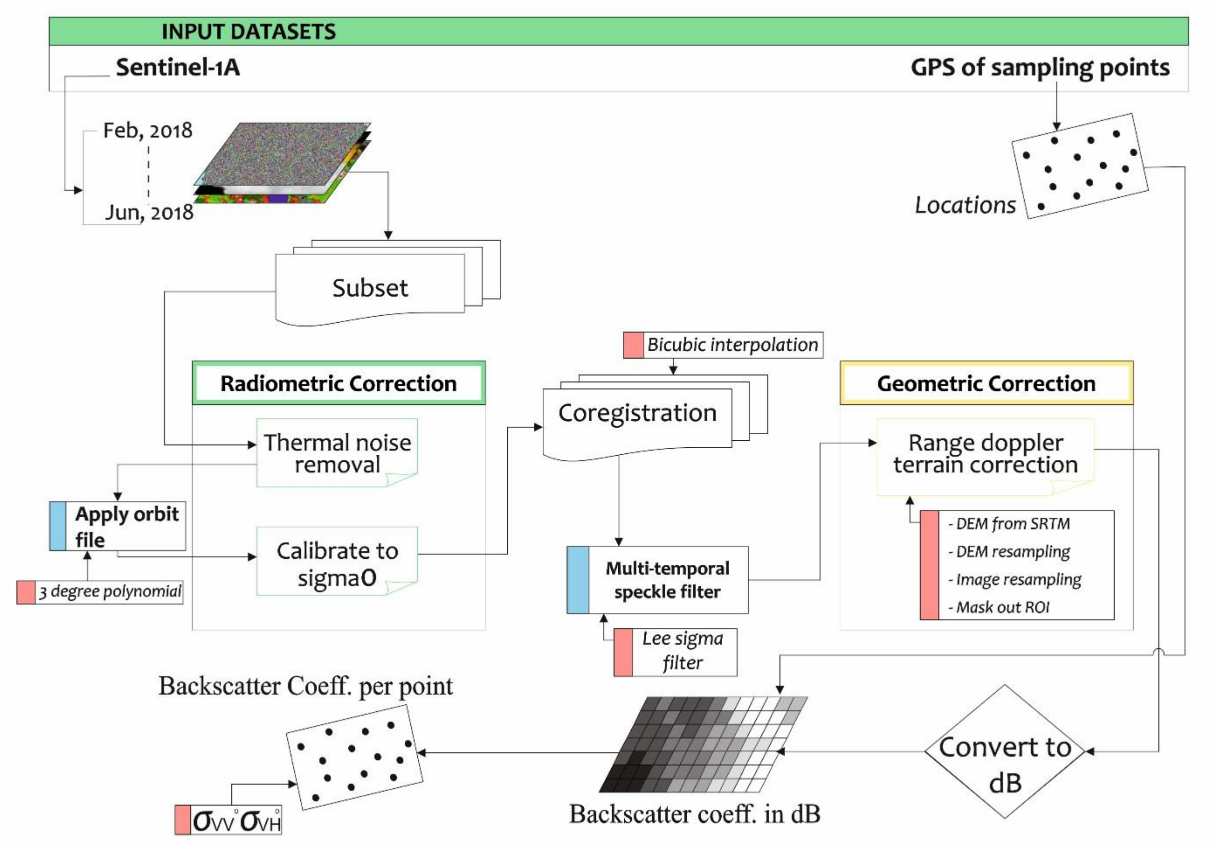


Figure 6.1. Flowchart of the Sentinel-1 image post-processing. dB = decibels.

6.2.2 Sentinel-2 data acquisition and processing

We also obtained Sentinel-2A Level-1C image product from ESAs Copernicus open access hub (https://scihub.copernicus.eu/dhus/#/home) at ± 1–5 days to Sentinel-1 acquisition dates

(Table 6.1). We only used Sentinel-2A image products to ensure commonality with Sentinel-1A images. The images were radiometrically and geometrically processed, including ortho-rectification and spatial registration on a global reference system. We processed the Top-Of-Atmosphere Level 1C image products by applying atmospheric correction via the $Sen2cor$ processor plugin to produce Level-2A Bottom-Of-Atmosphere reflectance image products. While applying atmospheric corrections, terrain correction was also applied with the use of the DEM data to compensate for distortions due to topographic variations of the image scene and tilt of the satellite sensor.

From the processed Sentinel-2 image products, we derived water and soil indices for use in the SM estimation, i.e. the Tasseled Cap Wetness index (TCWI) (Nedkov 2017), the second Normalized Difference Water index (NDWI2) and Color index (CI). We calculated the TCWI (Equation 6.1) with the use of band math operator. NDWI2 (Equation 6.2) and CI (Equation 6.3) were estimated with the use of Sentinel-2 Toolbox water and soil radiometric indices respectively. All stages of water and soil indices estimation were executed in the SNAP desktop software.

$$TCWI = [0.0649 * B1 + 0.1363 * B2 + 0.2802 * B3 + 0.3072 * B4 + 0.5288 * B5 + 0.1379 * B6 - 0.0001 * B7 - 0.0807 * B8 - 0.0302 * B9 + 0.0003 * B10 - 0.4064 * B11 - 0.5602 * B12 - 0.1389B8A] \quad (6.1)$$

where $TCWI$ is the tasseled cap wetness index and B are the various channels of Sentinel-2A image (i.e. 1–12 and 8A).

$$NDWI2 = \frac{(GF * B3 - IRF * B8)}{(GF * B3 + IRF * B8)} \quad (6.2)$$

where $NDWI2$ is the second normalized difference water index, GF is the green factor, B means the various channels of Sentinel-2A image (i.e. 3 – green and 8 – near infra-red) and IRF is the infra-red factor.

$$CI = \frac{(RF * B4 - GF * B3)}{(RF * B4 + GF * B3)} \quad (6.3)$$

where CI is the soil colour index, RF is the red factor, B means the various channels of Sentinel-2A image (i.e. 3 – green and 4 – red) and GF is the green factor.

6.2.3 Auxiliary datasets

More accurate SM estimates could be achieved with the integration of auxiliary datasets (Ali et al. 2015; Klinke et al. 2018). Consequently, we considered key auxiliary datasets that may affect the microwave signal. For bare soils, Sentinel-1 backscatter is strongly influenced by the variability of target variables (in this case SM) and landscape heterogeneities such as surface roughness and local incidence angle (Lievens and Verhoest 2011; Pasolli et al. 2014). As key auxiliary datasets, we considered an extensive stack of spatial layers, which were used as proxies to explain the variabilities of SM and the associated landscape heterogeneities (Table 6.2). The stack of auxiliary datasets were classified as process-based, dynamic and static features. Process-based features included $SAGA$ topographic wetness index and other DEM-terrain variables. We calculated the DEM-terrain variables with the use of $SAGA$ morphometry and hydrology functions (Conrad et al. 2015) in R software (R Core Team 2017). Static features also included physical soil properties such as clay, silt and sand contents, bulk density, soil type and buffer-distances. Following the approach of Hengl et al. (2018), buffer-distances were derived via the geographical locations of the validation locations, which were used to account for spatial relations and closeness between a validation location and the others (Hengl et al. 2018). In addition, we accounted for the spatio-temporal autocorrelation in the SM estimation using estimated cumulative days (CUM) and day of the year (DOY) of the *in situ* measurement dates (Hengl et al. 2018). Both CUM and DOY also represented the temporal-distances of *in situ* measurements at the validation locations in the time domain, which moreover addressed site-specific effects due to seasonality, time of day and long-term trends that occurred at the validation locations (Hengl et al. 2018). The CUM and DOY forms part of the list of features classified as dynamic features, and also included precipitation and evapotranspiration. All auxiliary datasets were resampled via a bicubic-interpolation algorithm to match the Sentinel-1 image's spatial resolution in order to ensure image compatibility.

Table 6.2. List of datasets used in the study

Parameters	Abbreviation	Description	Units
Sentinel-1 derived variables			
Backscatter coefficients	σ	Calibrated backscatter coefficients represented as VH () and VV () polarizations.	dB
Local incidence angle	LIA	The local angle between the incident Sentinel-1 radar beam and the normal line to the soil surface.	degree
Sentinel-2 water and soil derived indices			
Tasseled cap wetness index	TCWI	Tasseled cap wetness developed from Tasseled cap coefficients for Sentinel-2 bands (1 – 12 and 8A).	-
The second Normalized Difference Water Index	NDWI2	The second normalized difference water index used to detect surface water and extent.	
Color index	CI	Color index to differentiate soil types of an area.	
SRTM-DEM derived variables			
Topographic wetness index	TWI	SAGA topographic wetness developed from Shuttle Radar Topographic Mission (SRTM)-DEM.	-
Local slope	SLP	Slope indicating steepness of a location on the soil surface.	degree
Aspect	ASP	Aspect indicating the direction in which a location's slope faces.	
Surface roughness	SurfR	Surface roughness indicating the soil surface texture which influences the radar backscattering.	-
In situ measurements from validation locations			
Soil moisture	SM_p	*In situ* volumetric SM by calibrated PR2/60 capacitance moisture probe.	%
Clay content	Clay	Percentage clay content of soil samples.	
Silt content	Silt	Percentage silt content of soil samples.	
Sand content	Sand	Percentage sand content of soil samples.	
Bulk density	BD	Bulk density of soil samples.	g cm^{-3}
Soil type	Soil_type	Mapping units at series level.	-
Other derived variables			
Cumulative days	CUM	Cumulative day of the date of in situ SM measurement.	-
Day of the year	DOY	Day of the year for each in situ SM measurement	-

6.2.4 Influence of integrated datasets in SM estimation

6.2.4.1 Feature selection for use in SM estimation

We used the random-forest recursive feature elimination (RF-RFE) model to subset the stack of auxiliary datasets (Kuhn 2018). This enrichment step was critical to reduce dimensionality and, remove redundant and irrelevant features in the stack of auxiliary datasets (Table 6.2). In the RF-RFE, the *Gini* index was used to rank the stack of auxiliary datasets. Here, with expert knowledge as documented in the literature for SM estimation and the RF-RFE model, auxiliary datasets that were highly correlated were eliminated while selecting those with high importance. In this way, we improved computational efficiency, increased MLAs learning accuracy and reduced the generalization error of the MLAs used as inputs in the EBM (Hounkpatin et al. 2018). Choosing the optimal set of feature combination was moreover relevant to ensure same model input datasets for all base MLAs. In the stack of auxiliary datasets, we only considered σ_{VV}^{0} polarimetric backscatter from Sentinel-1. This was because, for the case of our study areas, σ_{VH}^{0} polarimetric backscatter values were all below the Noise Equivalent Sigma Zero of Sentinel-1 data (Torres et al. 2012). In addition, σ_{VV}^{0} has also been demonstrated to improve SM estimation than σ_{VH}^{0} (Amazirh et al. 2018) or the use of both σ_{VV}^{0} and σ_{VH}^{0} (El Hajj et al. 2017; Kumar et al. 2018). The optimal feature combination by the RF-RFE model was selected based on the cumulative R^2. We assumed that the selected features represented the optimal regression matrix, used as inputs, for the base MLAs. By evaluating sets of RF-RFE model fitting parameter configurations, we implemented the iterative RF-RFE model with a repeated cross validation (CV). In the optimal RF-RFE model parameter configuration, RF-RFE model was implemented under 16 resampling iterations and at 5 CV-folds. We executed all stages of the RF-RFE in R software (R Core Team 2017) using $caret\ R$ package (Kuhn 2018).

6.2.4.2 Relationship between SM_p and optimal feature combination

As adequately demonstrated in the literature, Sentinel-1 backscatter values are sensitive to medium and high SM values under bare or vegetated fields (Pasolli et al. 2015; Alexakis et al. 2017; Holtgrave et al. 2018; Klinke et al. 2018). However, very little is reported on low and extremely low SM conditions and the behavior of Sentinel-1 backscatter and other auxiliary datasets under similar landscape conditions (Paloscia et al. 2013; Urban et al. 2018). Consequently, we investigated the direct linear relationship between SM_p and integrated

Sentinel-1, -2, and other auxiliary variables at the validation locations of the study area. In this way, we assessed the relationship between σ^0_{VV} with SM_p and how other integrated auxiliary datasets optimized the SM estimation. These notwithstanding, there was the need to identify which auxiliary dataset accounted for which range of SM_p values in order to address their respective ambiguities and influences in the SM estimation (Pasolli et al. 2012). To help interpret the results, we standardized the values of all auxiliary datasets (due to their wide different ranges) in order to obtain a $z - score$ unit (Mohamad and Usman 2013). $z - scores$ were calculated based on their mean and standard deviations via the $scale$ function in R software (R Core Team 2017). Subsequently, we assessed their variability in discriminating SM_p levels in the SM estimation. From expert knowledge, SM_p values were cluster along three main SM ratings, i.e. $SM_p > 10\%$ Vol (medium), $5 \leq SM_p \geq 10\%$ Vol (low) and $SM_p < 5\%$ Vol (extremely low) (Grote et al. 2010; Cooper 2016).

6.2.5 SM estimation

6.2.5.1 Deriving EBM input dataset

In each of the base MLAs, the SM estimation was mainly executed under a training and a validation phase. The MLAs train the relationships between the independent variables, which is represented by Sentinel-1 and -2 derived parameters and selected auxiliary datasets, and the dependent variable, represented by SM_p (Table 6.2). We partitioned input datasets into 80% (training) and 20% (validation) sub-samples (Kuhn 2018). By applying k-fold CV, on the one hand, we used $k - 1$ sub-samples, iteratively, to train the SM estimation MLAs, and on the other hand, used the remaining sub-samples for MLAs validation. Through the -fold CV, we averaged the performance of the validation sub-samples to ensure an improved accuracy and robustness of an optimized MLAs for use in the EBM. The k-fold CV considered reference sample for training SM predictors and tuning its free parameters (Kuhn 2018). In this way, all dataset observations in the integrated Sentinel-1, -2 and other auxiliary datasets were considered in the training and validation of the base MLAs.

With the ANN algorithm, the best compromise between computational time and SM estimation accuracy can be achieved (Paloscia et al. 2013). The ANN consists of an input, a hidden and an output layer. Through the interconnected hidden neurons of the ANN, input datasets were converted to output datasets (Paloscia et al. 2013). In the ANN model (Figure 6.2A), we varied the number of neurons in the hidden layer, trained the ANN with the training sub-samples and tested the trained outputs with the validation sub-samples. The ANN model optimization was achieved with the use of a Bayesian Regularization algorithm. Through the ANN model fine-tuning, we additionally minimized the ANN input-output $RMSE$ in order to obtain an optimal ANN architecture. Optimized ANN training parameter configuration was achieved with 10 neurons at 1e+03 epochs. The SM estimation with ANN was implemented with the $brnn - R$ package (Pérez and Gianola 2018).

SVR is a supervised learning model that non-linearly generalizes unseen data adequately – mainly developed in the framework of classification and regression problems using kernel-based learning systems (Keerthi and Lin 2003; Ahmad et al. 2010). In the SM estimation with SVR, similar training and validation processes, as implemented for the ANN was undertaken. Here, we used the Radial Basic Function, which has been demonstrated to outperform other kernel functions in SVR model (Keerthi and Lin 2003; Pasolli et al. 2011). We used a grid search in designing the parameter configuration for an optimized SVR model (Figure 6.2A). Optimized SVR training parameter configuration was achieved with an epsilon regression with 0.1 ε (insensitive-loss function), a hyper-parameter of 0.05 σ (RBF kernel width) and a 10 cost of constraints violation (C; regularization parameter) to obtain an optimal SVR model. The SVR was implemented via the $kernlab - R$ package (Karatzoglou et al. 2018).

Random forest is a bootstrap sampler, random feature selector, out-of-bag error estimator and a full-depth decision tree grower (Breiman 2001). In random forest, prominence is given to the iterative training of its algorithm in order to produce a strong learning result instead of formulating the statistical model (Hengl et al. 2018). Also, the RFR model followed similar training and validation processes as in ANN and SVR. In the RFR model, random sub-samples were drawn from the training sub-sample with replacements. Each regression result was generated from a sub-decision tree constructed by a sub-sample. Consequently, the outputs of individually generated decision trees from the training sub-samples were combined to optimize the RFR model (Breiman 2001). SM was estimated by averaging the estimation outputs of the full-depth individual decision trees in the RFR model (Pal 2005). We configured the RFR via the $caret - R$ wrapper package (Kuhn 2018). Optimized model parameter configuration was obtained with a grid search training control. Resampling of the training sub-samples across the RFR tuning parameters was achieved with a $10 - fold$ CV that was repeated 10-times (Figure 6.2A).

Further details are reported in Haykin (2011) for the ANN, Smola and Schölkopf (2004) for the SVR and Breiman (2001) for the RFR algorithms. All stages of the SM estimation were implemented in R software (R Core Team 2017).

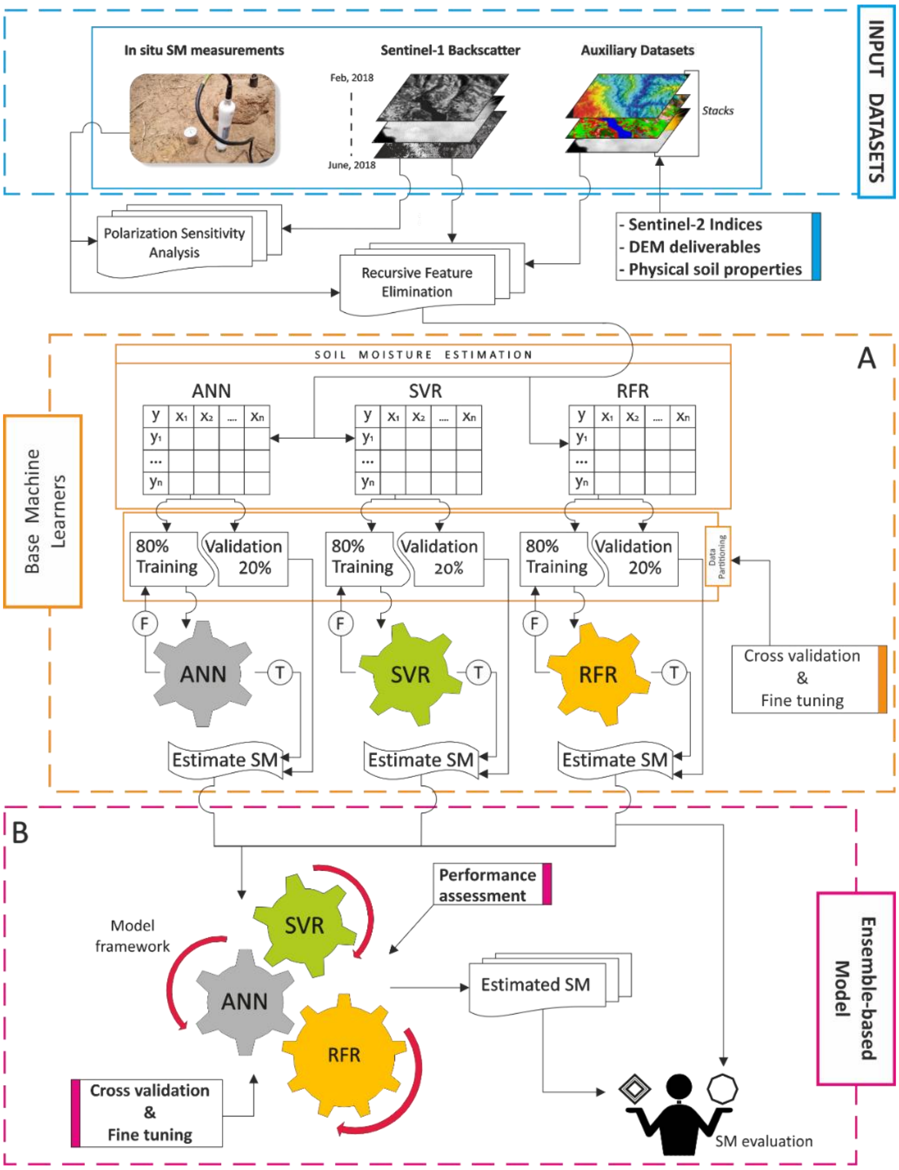


Figure 6.2. Methodological diagram of the SM estimation process using inputs from in situ SM, Sentinel-1 and -2 images and other auxiliary datasets. The diagram summarizes the ensemble modelling framework with the use of base MLAs, i.e. Artificial Neural Network (ANN; A), Support Vector Regression (SVR; A) and Random Forest Regression (RFR; A), and a linearly combined EBM (B).

6.2.5.2 EBM

We linearly combined the outputs of the base MLAs in the EBM in order to develop a SM estimate, which held prior estimated SM values as compared to the independent SM estimates of the base MLAs (Figure 6.2B). In this way, we synthesized the outputs of the individual MLAs into a single score in order to find a balance between over- or underestimated SM and also reduce the overall estimation error (Klinke et al. 2018). Following the approach of Reid and Grudic (2009), we applied an Elastic-Net Generalized Linear model on the inputs to ensemble the estimates of the base MLAs. Again, an 80-20 training-validation data split was executed (Kuhn 2018). Optimized parameter configurations used for the EBM were obtained with a grid search and tune parameters with a $gaussian$ response model type, 0.1 α (elastic-net mixing parameter) and 0 λ (lambda sequence). The EBM was implemented in R software (R Core Team 2017) using $caret$ (Kuhn 2018) and $glmnet$ (Friedman et al. 2017) R packages.

6.2.5.3 Accuracy assessment of the SM estimation

The performance of the base MLAs, in estimating SM was assessed by using a -fold CV. This was because, Holtgrave et al. (2018) demonstrated that MLAs results varied per iterations due to data partitioning. Consequently, following the approach of Kuhn (2018), we re-fitted each MLA 10-times using 80% (test sub-samples) of the datasets and estimated with the remaining 20% (validation sub-samples). In this approach, model stability and reliability of model outputs were ensured by the estimated standard deviations. For each estimated SM (SM_e), the variations explained by the models (R^2) and the average error of the models (*RMSE*) were derived (appendix Equations S6.3 and S6.4 respectively). In addition, we also considered the percent bias (*Pbias*; appendix Equations S6.5) to assess the over- or underestimation of SM by the base MLAs and the EBM. Optimal *PBias* is zero, with negative values indicating model-underestimating bias, whereas positive values indicated the reverse. We implemented accuracy assessments in R software (R Core Team 2017). Prior to SM estimation and model performance assessment, SM_p values were log-transformed due to its skew distribution. Consequently, we report model performances on a log scale.

6.2.5.4 Estimation framework

All R processes were implemented on the GWDG's high performance computing clusters of the University of Göttingen. We provide detailed workflow of the SM estimation on a publicly accessible repository (DOI: 10.5281/zenodo.3662411). Publicly available resources include R scripts, metadata and readme files, which also promotes open and reproducible science. Full datasets used for the present study are also available on the same DOI.

6.3. Results and discussion

6.3.1 Variability of integrated datasets used for estimating SM

Figure 6.3 shows the temporal patterns of $\mathrm{SM_p}$, σ_{VV}^0, TCWI, NDWI2, and CI along soil toposequences of the study area. $\mathrm{SM_p}$ was extremely low (0.99 ± 0.26% Vol; mean ± standard error) and medium (10.77 ± 1.05% Vol) at February 20, 2018 (peak of the dry season) and at June 8, 2018 (onset of heavy rains) respectively. In total, 54.12% of $\mathrm{SM_p}$ distributed in the extremely low SM cluster (n = 184), whilst 25.29% (n = 86) and 20.59% (n = 70) distributed in the low and medium SM clusters respectively. Temporally, $\mathrm{SM_p}$ showed a high-low-high trend along all soil toposequences (Figure 6.3A–C). Variability of $\mathrm{SM_p}$ (range) was slightly high (9.78%) in the middle to lower slope as compared to the toe slope (8.90%) and the upper slope (5.58%).

The lowest (-16.98 ± 1.83 dB) and highest (-13.30 ± 1.64 dB) σ_{VV}^0 were both recorded in the toe slope at March 28th, and June 8, 2018 respectively, which corresponded with variability of $\mathrm{SM_p}$ (Figure 6.3). Similarly, the temporal variability of σ_{VV}^0 along the soil toposequences was generally consistent with the trend of $\mathrm{SM_p}$. Though the number of validation locations were not equal in each soil toposequence, σ_{VV}^0 showed similar temporal variability across the landscape, i.e. range = -3.52 dB (upper slope), -3.49 dB (middle to lower slope) and -3.68 dB (toe slope).

Throughout the measuring period, water (TCWI and NDWI2) and soil (CI) indices from Sentinel-2 images generally bears an inverse resemblance (Figure 6.3). Low TCWI and NDWI2 indices were recorded in the toe slope followed by the middle to lower slope and the upper slope. The reverse condition was observed with CI along the soil toposequences. Some consistencies were however observed between the Sentinel-2 water and soil indices at May 27th, and June 6, 2018 respectively (Figure 6.3).

In the study area, wet areas are associated with toe slopes, whereas dry areas are associated with upper slopes. Corroborated by this observation, the topographical positions and similar soil texture as observed by soil types of the validation locations mainly drove the amount of

SM. This consequently influenced the radar backscatter along the soil toposequences (Franceschetti and Lanari 2018). On the one hand, the toe slope soils with low elevation and high clay content, tends to have higher SM and hold SM for longer periods than the other toposequences. On the other hand, the medium to low SM_p values of soil types of the different soil toposequences determined the reflectivity and intensity of radar backscatter (observed here by the low) (Petropoulos 2013). In terms of the derived water indices from Sentinel-2 image products, low TCWI and NDWI2 is consistent with medium SM_p, similar to the observations of Sánchez-Ruiz et al. (2014) and Sánchez-Ruiz et al. (2015). The higher discrepancies observed with CI index is probably due to the strong and bright soil colors and the well to moderately-well drained soils in the upper and middle to lower slopes of the study area (Adu 1995). Another possible reason for the inconsistencies observed with soil and water indices, between May 27 and June 6, 2018 (Figure 6.3), was implicated by the increased amounts of precipitation in the study area from June 8, 2018.

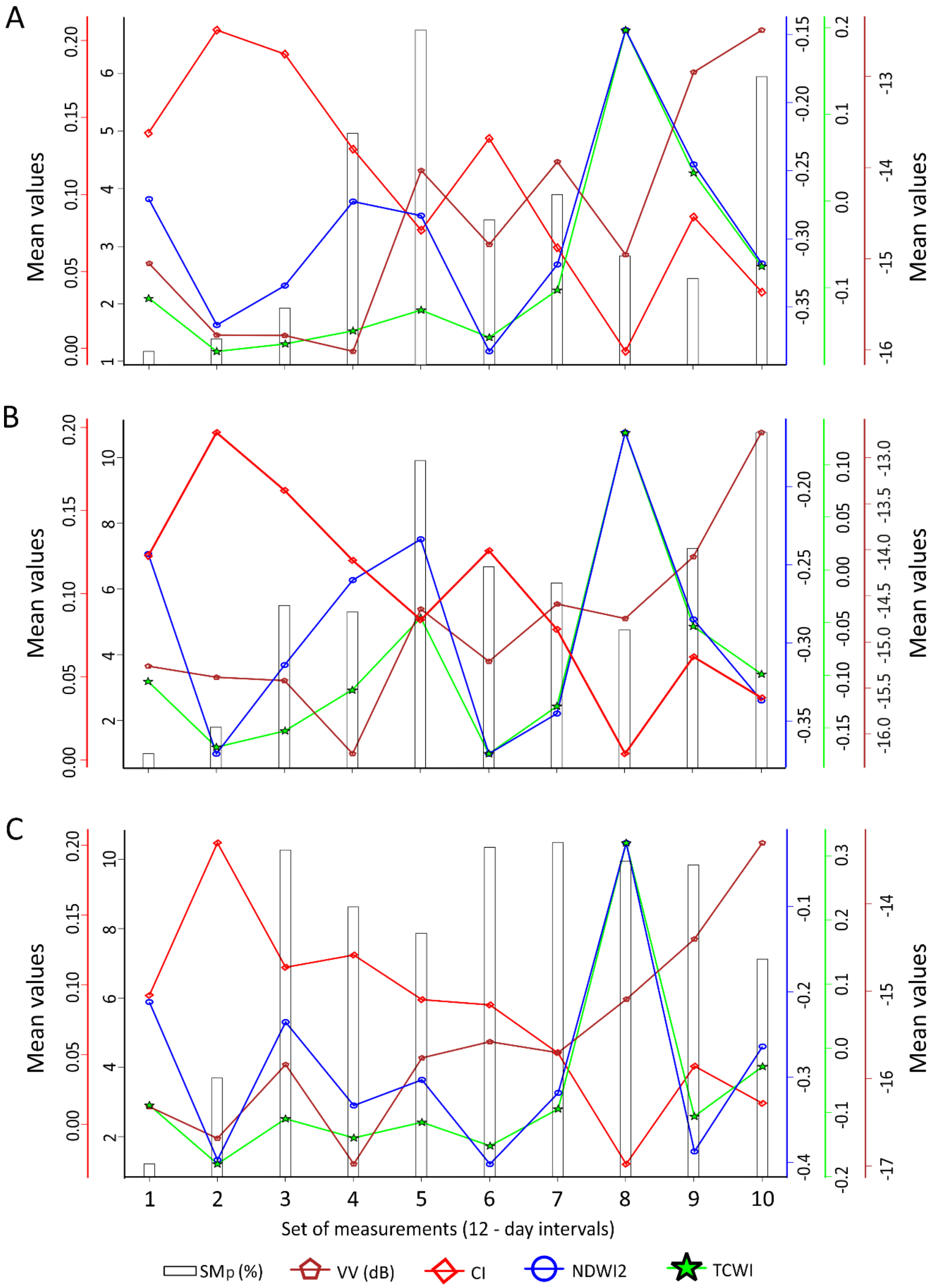


Figure 6.3. Temporal variability of in situ SM (SM$_p$), Sentinel-1 σ^0_{VV} (VV) and Sentinel-2 derived soil (CI) and water (NDWI2, TCWI) indices along topographical units. CI = soil Color Index, NDWI = the second Normalized Difference Water Index and TCWI = Tasseled Cap Wetness Index. Along the topographical units are; A = upper slope, B = middle to lower slope and C = toe slope.

6.3.2 Influence of integrated datasets in the regression matrix of the EBM

6.3.2.1 Optimal model input features

Optimal model input parameter combination and their importance is presented in Figure 6.4. Ten most important features were selected as optimal predictors for estimating SM for our case study areas (R^2 = 0.98) (Figure 6.4). These optimal features follow those documented in the literature for bare soils of other geographical settings (Lievens and Verhoest 2011; Pasolli et al. 2014). Physical soil properties were considered less important via the RF-RFE model. Generally, all selected parameters cumulatively showed strong importance in the list of model inputs. However, on the one hand, there was a negligible decrease in variations explained by the RF-RFE when LIA was initially included (Figure 6.4). On the other hand, no marked differences (ΔR^2 = +0.01) were observed when DEM-derived variables were further included (particularly TWI; Figure 6.4). Despite the slightly marked differences in the estimated R^2, our results suggest that the main contributing variables, cumulatively, were from Sentinel-1 and -2 derived variables. This observation indicates that though the DEM-derived variables contributed minimally, they however accounted for some effects due to inherent site-specific landscape heterogeneities such as those from surface roughness and topography at the validation locations. One other possible explanation is the coarse resolution of the DEM (i.e. 1 arc-second). It can be assumed that the higher the resolution, the higher the estimation accuracy of intended soil properties with a DEM-derived variable (Hengl et al. 2017b). In inference, the enrichment step of selecting optimal-model input feature combination is critical in SM estimation than using as many auxiliary datasets available (Holtgrave et al. 2018). Thus, over-usage of auxiliary datasets could be detrimental to accurate SM estimation from RS data.

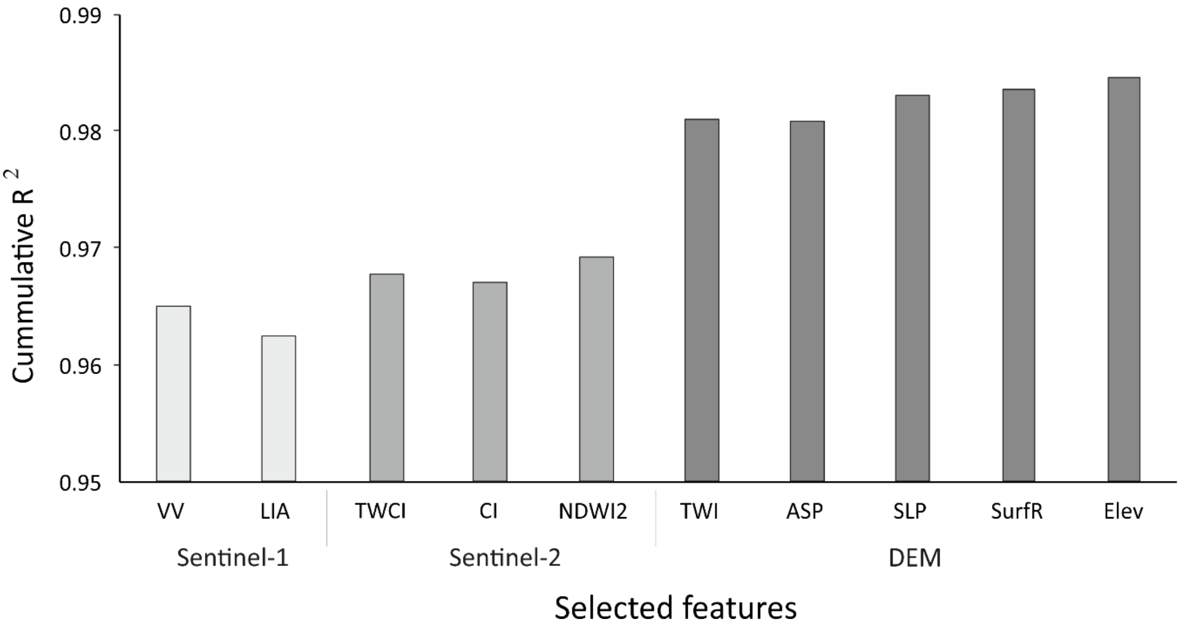


Figure 6.4. Optimal set of input features and its importance (high R^2) in estimating SM from integrated radar-optical-DEM-derived variables.

6.3.2.2 Sentinel-1, -2 and DEM-derived variables

Radar backscatter is mainly affected by the dielectric properties of soils that is driven by the amounts of inherent SM (Dobson et al. 1985; Barrett et al. 2009). Consequently, at low SM levels, radar backscatter is also highly affected by landscape heterogeneities such as surface roughness, topography and others (Gao et al. 2017). The generally measured extremely low and low SM_p levels suggest that not only inherent SM accounted for the radar backscatter for our case study areas, but also, other site-specific landscape heterogeneities (Gao et al. 2017). Higher variation of predictors (Figure 6.3) show a capability in discriminating the target variable (in this case SM_p) (Omar et al. 2017). From the integrated radar-optical-DEM dataset, σ^0_{VV}, LIA, and DEM-derived variables (TWI, ASP, SLP, SurfR and ELev) showed higher capability in discriminating the entire range of SM_p levels on the one hand (Figure 6.5) (Holtgrave et al. 2018). Sentinel-2 derived variables, i.e. TCWI, CI and NDWI2, specifically discriminated different ranges of measured SM_p, on the other hand. At the extremely low SM_p, SM_p was mainly discriminated by TCWI whilst CI and NDWI2 discriminated the medium, and low to medium SM_p respectively (Figure 6.5). This shows that our input parameters allowed a wider range of variability in order to discriminate and potentially estimate SM within the entire range of SM_p of the validation locations (Urban et al. 2018).

Concerning radar backscatter and SM, strong linear relationships ($R^2 = 0.9$) have been reported for moist and wet soil conditions (~25% Vol to 90% Vol), particularly for Sentinel-1 and other related SAR images (Amazirh et al. 2018; Holtgrave et al. 2018). However, very little has been reported on dry and slightly moist soil conditions (El Hajj et al. 2017). In El Hajj et al. (2017), weak linear relationship ($R^2 = 0.3$) was reported between σ_{VV}^{0} and low to medium SM levels (~7% to 36.3% Vol). This observation suggests that very weak linear relationships between SM and σ_{VV}^{0} are expected at low and extremely low SM conditions. Corroborated by these observations, we observed a similar trend and a very weak linear relationship was recorded between SM_p and σ_{VV}^{0} ($R^2 = 0.1$; Figure 6.5). Similarly, all other integrated variables used for the SM estimation also recorded weak relationships with SM_p ($R^2 = 0.1$–0.2; $\rho < 0.05$). Our results thus suggest that the synergetic use of Sentinel-1, -2 and DEM-derived variables rather presents an improved approach, which strengthens the relationship between input parameters and SM_p. Supporting this point, a cumulative R^2 of 0.66 was observed (Figure 6.5). In addition, the radar-optical-DEM data integration strengthened the SM levels' discrimination capability (Holtgrave et al. 2018). It is worth noting that this observation is only valid for validation locations of the installed access tubes, which are characterized by bare arable soils with no vegetation cover.

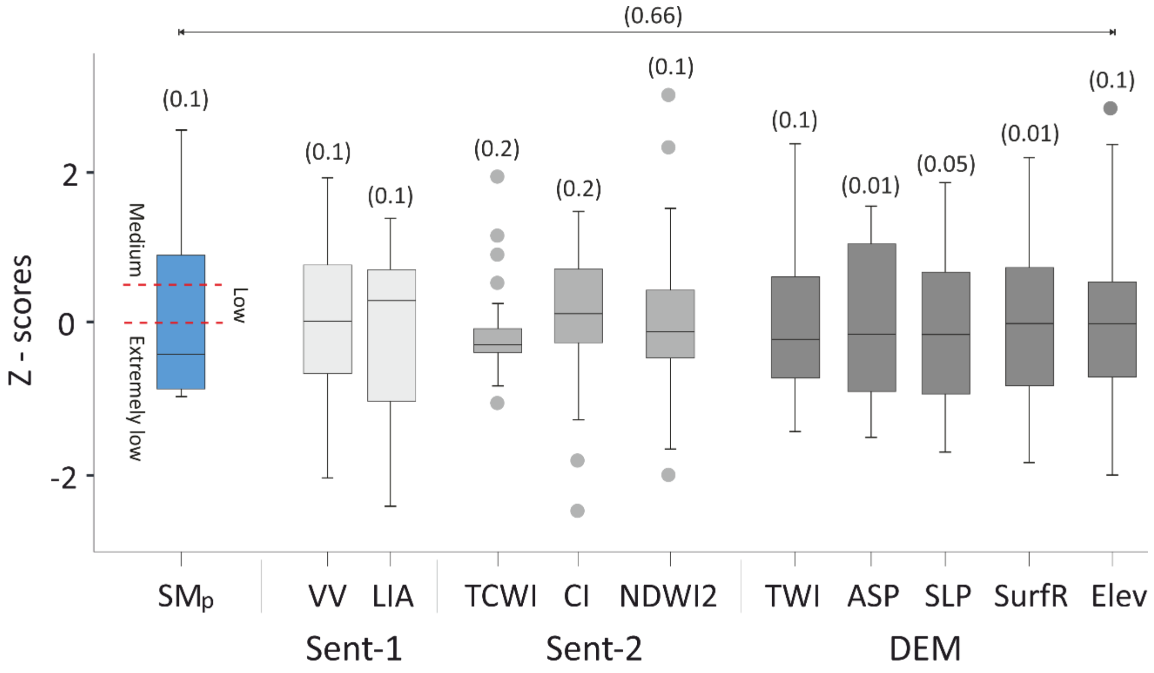


Figure 6.5. Distribution of SM estimating variables showing their capability in discriminating the range of *in situ* SM levels (SM). VV = Sentinel-1 σ_{VV}^{0} and LIA = local incident angle from Sentinel-1. Sentinel-2 derived variables include CI (soil Colour Index), NDWI (the second

Normalized Difference Water Index) and TCWI (Tasselled Cap Wetness Index). DEM-derived variable also includes TWI (SAGA topographic wetness index), ASP (aspects), SLP (slope), SurfR (surface roughness) and Elev (elevation). Coefficient of determination (R^2; in parenthesis) were all significant at $\rho < 0.05$.

6.3.3 Model validation and performance

6.3.3.1 SM estimated via the base MLAs

From the optimal input feature combination, SM estimated by the ANN algorithm recorded an estimation error (*RMSE*) of 0.18% Vol SM_p (R^2 = 0.92) and is represented by the linear regression equation $SM_e = 0.09 + 0.93*SM_p$ (Figure 6.6A). SM estimated via the SVR algorithm resulted in an estimation error of 1.14% Vol SM_p (R^2 = 0.94; Figure 6.6B) and is also represented by the linear regression equation $SM_e = 0.18 + 0.88*SM_p$. From Figure 6.6C, an estimation error of 0.15% Vol SM_p (R^2 = 0.93) was achieved via the RFR. The linear regression equation $SM_e = 0.30 + 0.8*SM_p$ represents the relationship between SM_e and SM_p via the RFR algorithm. Though the highest proportion of variance of the base MLAs was explained by SVR, RFR recorded the lowest SM estimation error. Concerning SM estimation error, SVR was the least performing algorithm (Figure 6.6). However, all MLAs showed a strong generalization ability in estimating SM along the entire range of SM_p. The margin of model performance (*ΔRMSE*) between base MLAs were all < 1% Vol SM_p. Corroborated by their performance metrics shown in Figure 6.6, ANN and RFR performed similarly as compared to the SVR algorithm.

Our findings indicate that the base MLAs exploited the range of variability of the integrated radar-optical-DEM model-input datasets in order to estimate SM (Pasolli et al. 2015; Alexakis et al. 2017; Kumar et al. 2018). In addition, the MLAs exhibited a strong robustness to noise or ambiguities that were associated with the model input datasets (Pasolli et al. 2011). Alexakis et al. (2017), Pasolli et al. (2014; 2015) and Kumar et al. (2018) demonstrated similar observations for the ANN, SVR and RFR respectively. However, in Figure 6.6B, there is a systematic artefact effect with the SVR, which is mainly implicated by outliers such as image saturation and noise from Sentinel-1 and -2 images, and the extremely-low inflated SM_p (Ulaby et al. 2014; Holtgrave et al. 2018). The increased number of support vectors in the SVR explains this observation and thus, indicates the need for more training points (Keerthi and Lin 2003; Smola and Schölkopf 2004). This observation appears to be one of the drawbacks of limited-SVR training datasets in image analyses. It is worth noting that sometimes increased number of iterations tends to overestimate data and thus, might have also contributed to the observed systematic artefacts shown in Figure 6.6B (Hauck and Mühll 2003).

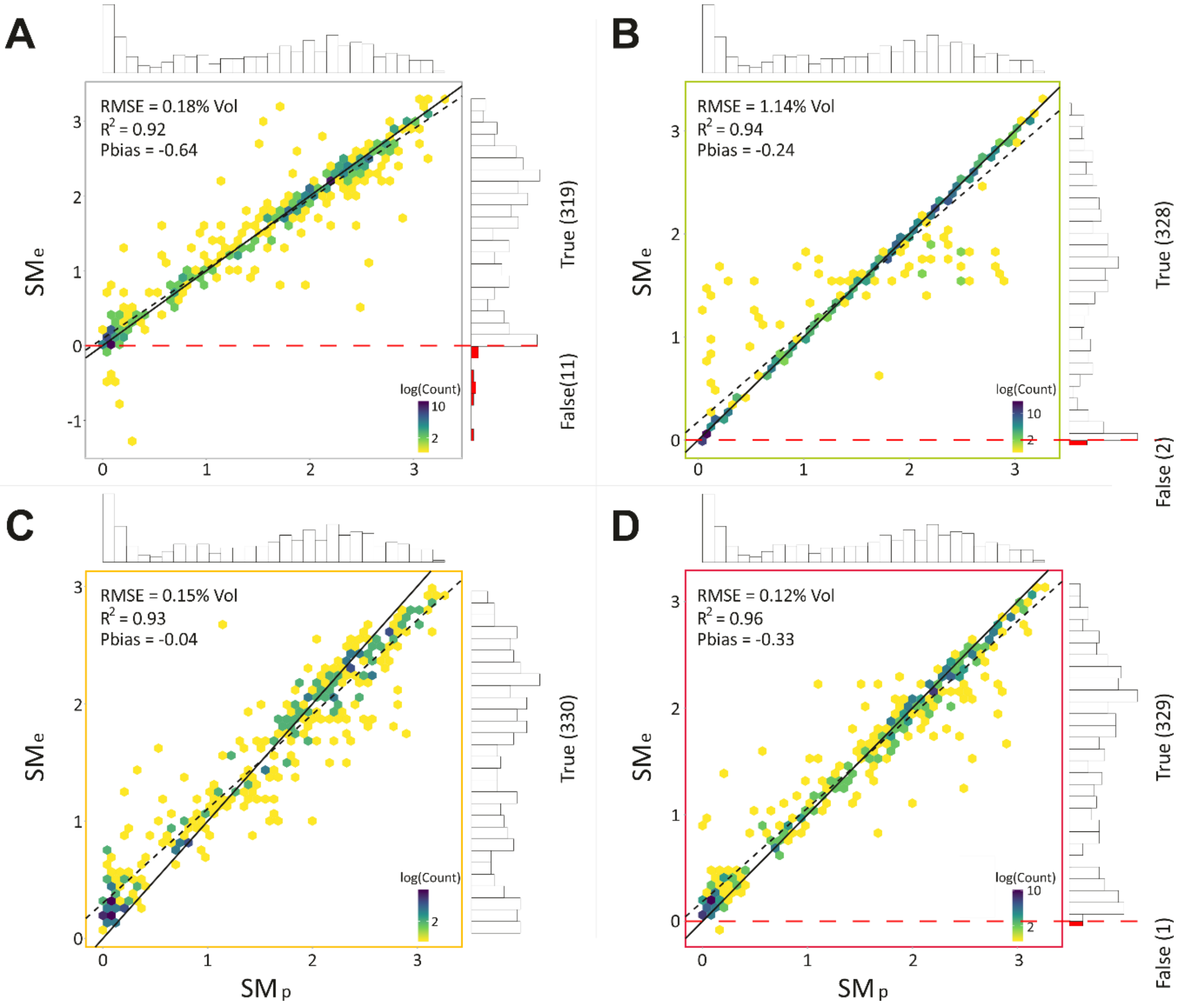


Figure 6.6. SM$_p$ versus SM$_e$ by the optimized base MLAs and the EBM. For A) ANN algorithm, B) SVR algorithm, C) RFR algorithm and D) EBM.

Comparing the base MLAs, the order of model stability and reliability (i.e. robustness to extremely-low SM$_p$ values) were RFR (0.81; standard deviation) > SVR (0.89) > ANN (0.95). In the study of Kumar et al. (2018) where RFR, SVR and ANN were compared, similar model stability was demonstrated for RFR. Notarnicola et al. (2008) and Paloscia et al. (2013) also reported similar model stability for the ANN. Although SVR was the least performing algorithm with regards to SM estimation error, it showed similar model stability as the RFR. The possible explanation for this observation is due to the strong generalization ability of SVR than just reducing the SM estimation error (Tomar et al. 2011; Vapnik 2013). In addition, SVR allowed greater tolerance to noise and converged on a global minimum, as reported by Bisgin et al.

(2018). Hence, SVR closely showed a strong relationship between SM_p and SM_e, than RFR and ANN.

Seemingly, Sentinel-2 derived variables contributed to an increase in the base MLAs performance compared to using only Sentinel-1 derived variables – as depicted by the cumulative R^2 of the RF-RFE model (Figure 6.4). The dominant low and extremely low SM conditions at the validation locations confirms these findings and are corroborated by those reported in the literatures for dry-bare soils (Paloscia et al. 2013; Pasolli et al. 2015; El Hajj et al. 2017). Another critical contribution to the low SM estimation errors was the inclusion of the DEM-derived variables. This is because, it has been demonstrated that by including DEM-derived variables in SM estimation from RS data, ambiguities that results from landscape heterogeneities such as surface roughness and topography are disentangled (Pasolli et al. 2014). The aforementioned findings therefore suggest that at extremely low to medium SM conditions, Sentinel-1 σ_{VV}^{0} alone, without the inclusion of Sentinel-2 and other DEM-derived variables, is not a viable way to estimate SM (Banson et al. 2016).

Slight discrepancies observed as compared to the baseline methodologies, from which this study extends and modifies, can mainly be attributed to the different characteristics of the validation locations. In this study, we focused on bare arable lands with very dry soil conditions whilst the others validated SM in vegetated moist to wet soil conditions.

6.3.3.2 SM estimated via the EBM

By combining the base MLAs, we recorded the highest model performance and the lowest estimation error, i.e., 0.96 (R^2) and 0.12% Vol SM_p (RMSE) respectively (Figure 6.6D). A linear regression equation $SM_e = 0.19 + 0.88*SM_p$ represents the EBM. Due to the linear synthesis of the outputs of the base MLAs into a single score, similar systematic artefacts observed with the SVR was reflected in the EBM. However, marked improvements were achieved with regards to the systematic artefacts via the EBM (Figure 6.6D). Proportion of variations explained by the EBM, between SM_e and SM_p, was highest than the individual base MLAs.

One key objective of an EBM is to bag "weak" MLA into a "strong" algorithm (Kotu and Deshpande 2018). In the present study, the EBM achieved this objective. Supporting this finding is the fairly symmetrical 1:1 plot shown in Figure 6.6D. Although an enhanced SM estimates were achieved via the EBM, the observed systematic artefacts with the EBM is one of the drawbacks of coupling weak and strong performing MLAs. We thus argue that the performance of the EBM is worth its usage due to the clearly marked different performance metrics of the base MLAs, which shows competing quality of SM estimation technique from the same model input datasets. Generally, the EBM reproduced the whole range of variability of the complex high-dimensional input datasets as compared with the individual MLAs (Table 6.2 and appendix Table S6.1) and represented a balance of the model performances of the base MLAs in the SM estimation (Klinke et al. 2018; Liakos et al. 2018).

6.3.3.3 Effectiveness of SM estimation models to SM_p thresholds

SM_p showed slightly wider variability than estimated SM via the estimation models (Figure 6.7A). Conversely, variability of SM_e was wider than SM_p when models were evaluated along each soil toposequence (Figure 6.7B–D). In inference, there is wider generalization of the estimation models, particularly, when validation locations were evaluated in its entirety than being evaluated along a segmentation criterion. Thus, at wider model representation, the MLAs and the EBM tends to obscured these competing qualities, whereas on a segmentation presentation, such competing qualities are clearly exhibited. Holtgrave et al. (2018) reported similar observations for the north-east of Germany.

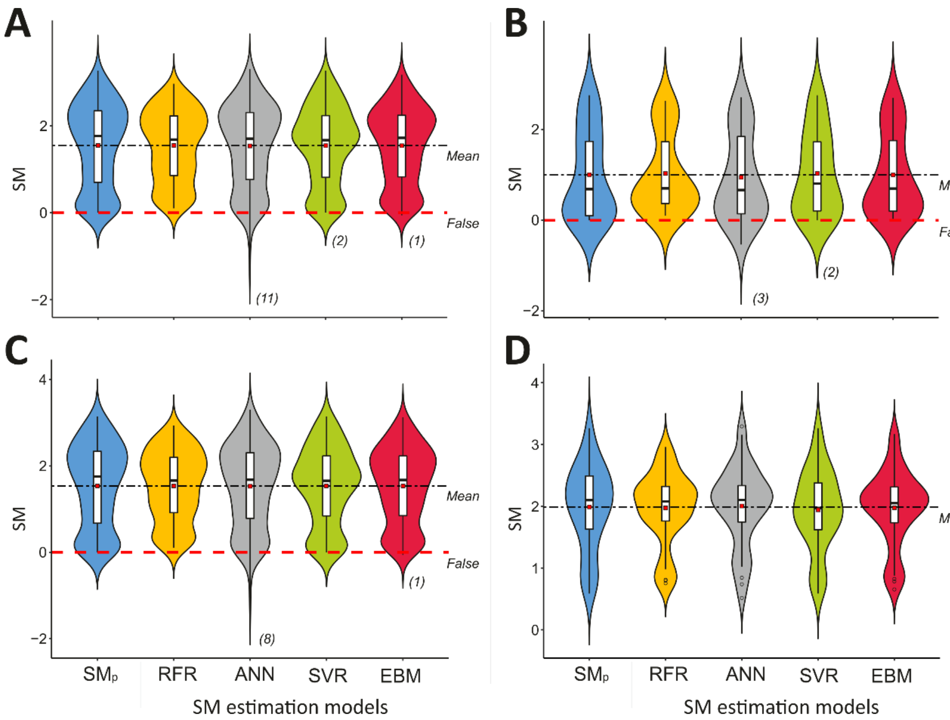


Figure 6.7. Comparing the effectiveness of base MLA and the EBM for estimating SM from integrated radar-optical-DEM dataset. A = Overall performance of base MLAs and the EBM. For performance of base MLAs and the EBM along topographical units, B). upper slope, (C). middle to lower slope, and (D). toe slope.

Concerning the entire variability of SM_p, all models underestimated SM with the highest occurring with the ANN (-0.64%; PBias) and the lowest via the RFR (-0.04%) (Table 6.3 and Figure 6.7A). Over- or underestimated SM corresponded with the extremely low and low SM_p values across all models, particularly, as those depicted by the upper slope toposequence (Table 6.3). Our observations follow those of Holtgrave et al. (2018). Similarly, Ahmad et al. (2010) demonstrated such model overestimation at low SM_p, particularly with the SVR. Also, the SVR showed marked differences in terms of over- or underestimation across soil toposequences. Although it has been documented that SVR falsely consider low SM_p values as outliers (Holtgrave et al. 2018), improved SM_e were however achieved with SVR than the ANN in the upper slope (Figure 6.7). Generally, low numbers of over- or underestimated SM_p were observed at either the entire validation locations or along topographic unit model-segmentation (except for the upper slope; Table 6.3). Meanwhile, inconsistencies (over- or underestimated SM_e) observed with the EBM is implicated by the performance of the base MLAs along the soil toposequence segmentation. Comparably, in the middle to lower and toe slopes, all models again underestimated SM_p (range = 2.36% SM_p; Table 6.3) except for the upper slope toposequence, where an over-underestimation trend was observed. The possible reason for this observation, particularly in the upper slope, is due to the extremely-low inflated SM_p (Vorontsov 2009; Tomar et al. 2011). In addition, these trends could also be attributed to the sparse nature of our *in situ* SM measurement network and the long abrupt transition periods between *in situ* measurement (i.e. 12-day interval) (Ramcharan et al. 2018). In Baskan et al. (2013), SM is reported to change at small distance and time intervals. With decreased time-lags, gradual changes can be registered to enhance model estimation performances, especially in the case of low-inflated SM_p.

Table 6.3. SM that was over- or underestimated by the base MLAs and the EBM

Level of bias estimate	**RFR**	**ANN**	**SVR**	**EBM**
	% Vol.			
By all models	-0.04	-0.64	-0.24	-0.33
Along soil toposequences				
Upper slope	2.59	-5.93	2.62	-0.59
Middle to lower slope	-0.07	-0.52	-0.03	-0.23
Toe slope	-0.89	-0.73	-2.39	-0.74

Positive values indicate overestimating bias, whereas negative values indicate underestimating bias.

Although improved SM_e can be achieved via the estimation models, there is a < 0% Vol SM estimation effect (represented here as "false" SM_e) with the ANN at the extremely low SM_p cluster for the case of our study areas (Figure 6.6 and 6.7). At eleven *in situ* measurement locations, ANN recorded false estimated SM with SM_p threshold of 0.375% Vol (minimum). Although such effect was observed with the SVR and EBM, they are however insignificant (Figure 6.7). The findings of Ahmad et al. (2010), which demonstrated that ANN was unable to capture low SM_p, corroborates our observations. For the case of our study areas, at ≤ 0.375% Vol SM_p, ANN is ineffective and under-performs, which caused the EBM to behave similarly. This finding is corroborated by results obtained from the toe slope where consistent higher SM_p was recorded at all stages of the dry season, and no "false" SM was estimated (Figure 6.7D). Temporal antecedent precipitation of the study area well correlate with these findings. Thus, medium and extremely low SM_p in toe and upper slope respectively, and low and high precipitation at February 20, 2018 (peak of the dry season) and at June 8, 2018 (onset of heavy rains) respectively. Clearly, over- or underestimation and "false" estimated SM results rather show a significant optimization via the EBM. Figure 6.8 confirms the higher correspondence between SM_e and SM_p. These observed phenomena, as shown in Figure 6.8, buttress our call for rather using a coupling approach instead of individual MLAs in SM estimation from RS data (Ali et al. 2015). It is important to note that we did not cover higher ranges of SM in soils of the study area at the time of SM_p measurements. Hence, we cannot also recommend the maximum SM_p thresholds at which these MLAs are also inefficient. Further studies with extremely higher SM values could verify these MLAs performance thresholds.

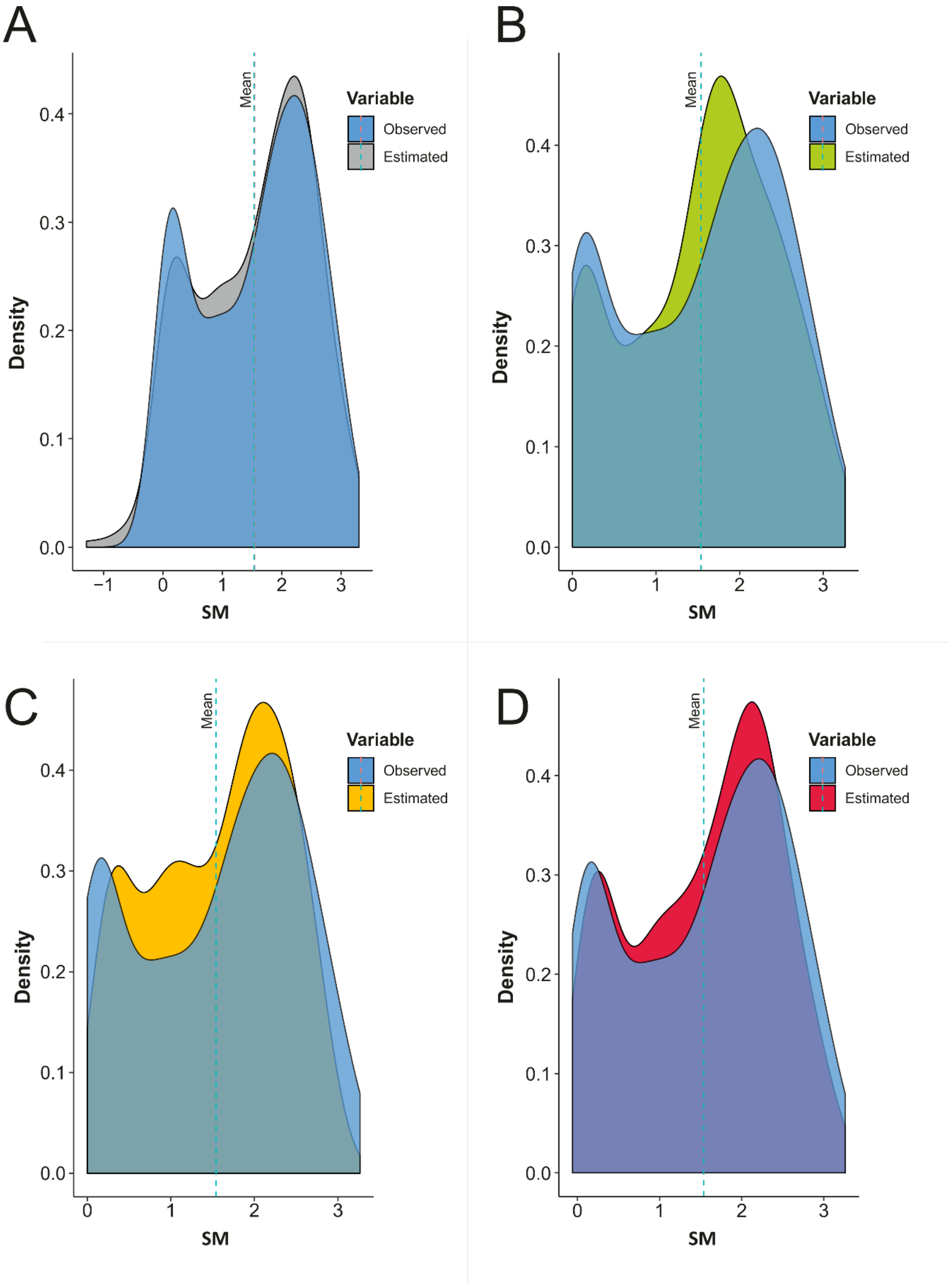


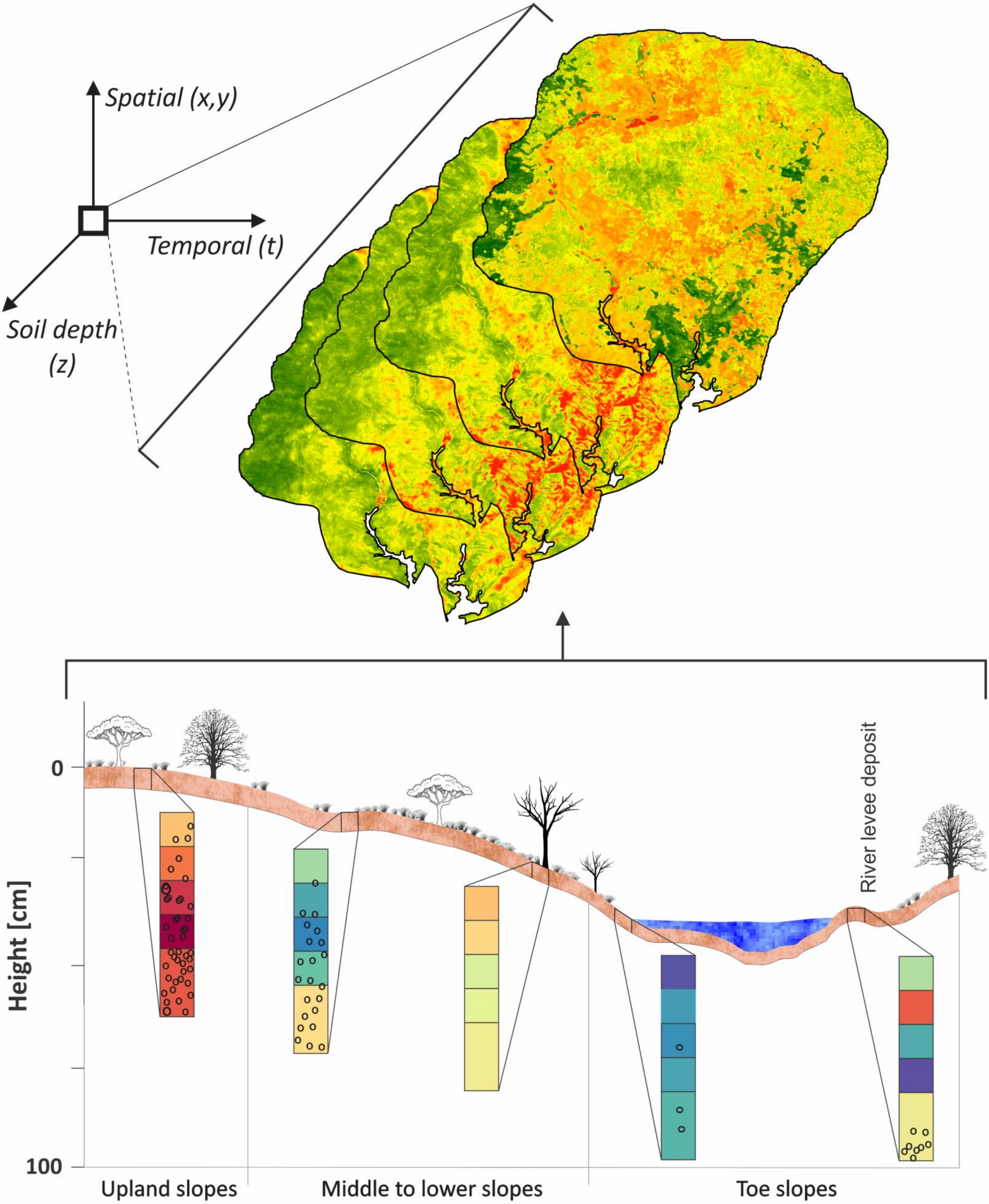
Figure 6.8. Scaled probability distribution of measured (SM_p) and estimated (SM_e) SM for A) the ANN algorithm, B) the RFR algorithm, C) the SVR algorithm and D) the EBM.

6.4. Conclusion and outlook

In this paper, the feasibility of using an ensemble of RFR, ANN and SVR to estimate SM from integrated Sentinel-1, -2 and DEM-derived datasets was investigated and validated for the Guinea savannah zone of Ghana. The SM estimation was validated by comparing SM_p to SM_e. By linearly combining the outputs of the base MLAs in the EBM, the overall SM estimation errors and model performances were improved. Hence, the EBM points out an effective way to estimate SM from RS data with an improved estimation accuracy. In general, the SM estimation approach is data-driven, hence, selecting optimal input parameter combination was key to improve the estimation accuracy. From the generally low SM_p levels, Sentinel-2 derived variables highly contributed to an increase in the base MLAs performance compared to using only Sentinel-1 derived variables. Nevertheless, by coupling the robustness of the base MLAs and the use of the integrated radar-optical-DEM input dataset, improved estimation results are achieved. Despite competing performances from the base MLAs, the EBM presented a greater data generalization ability and robustness to ambiguities in the model input datasets. All models (except RFR) estimated “false” SM as certain minimum SM_p thresholds. At SM threshold of $\leq 0.375\%$ Vol SM_p, ANN is ineffective and under-performs for the case of our study areas. However, the EBM showed a significant optimization and presented the best compromise between the over- or underestimation and “false” estimated SM of the base MLAs. One novel aspect of this study was the determination of the minimum SM thresholds at which these widely used MLAs under-performed. Future work will investigate the maximum SM thresholds at which these MLAs are also inefficient in order to provide full model performance recommendations.

The outcome of this study suggests that at certain SM regimes and soil conditions, the EBM is an effective tool for SM estimation from ESA’s Sentinel-1 and -2 missions. Hence, new perspectives are opened to improve SM inputs into crop simulation models and also map surface SM at detailed spatio-temporal global scales.

Spatio-temporal mapping of soil water storage in a semi-arid landscape of Northern Ghana – A multi-tasked machine-learning approach



Chapter 7: Spatio-temporal mapping of SWS in a semi-arid landscape of Northern Ghana – A multi-tasked machine-learning approach

This manuscript is under review as:
Nketia KA, Ramcharan A, Asabere SB, Herbold S, Erasmi S, Sauer D. Spatio-temporal mapping of soil water storage in a semi-arid landscape of Northern Ghana – A multi-tasked machine-learning approach.

Conference presentation at AGU Fall Meeting 2019, San Francisco: H134, The Role of Soil Moisture in Land–Atmosphere Interactions. https://agu.confex.com/agu/fm19/webprogrampreliminary/Paper535503.html

Abstract

Root-zone soil water storage (SWS) illustrates the water storage adequacy of soil horizons and its water reservoir capabilities from which crops can draw upon during transient water deficit periods. However, such a critical variable is not available at spatio-temporal models for sub-Saharan Africa, where the agricultural sector is the mainstay of many national economies. For this reason, we developed the first spatio-temporally explicit modelling framework to map the four-dimensional SWS at a 100 m spatial resolution (12-day temporal resolution). We validate the modelling framework with time-series *in situ* SM data (n = 1,920) from 36 locations on key benchmark soils in the Guinea savannah landscape of Ghana. The multi-tasked modelling framework was implemented with an ensemble of a parallelized random forest for spatial data and the extreme gradient boosting algorithms to predict maps at six standard soil depths (after the *GlobalSoilMap* specifications). As model-input datasets, we used an optimal subset from *a priori* stack of over 100 key parameters that drives the variability of SWS and follows the *scorpan* concept for soil mapping. The multi-tasked model framework was optimized and evaluated via a $5-fold$ Leave-Location-Time-Out cross validation algorithm, and its performance assessed by its fitting and validation accuracy, model bias and stability. Our results showed that the multi-tasked ensemble model is a viable alternative to predict the four-dimensional SWS. An accuracy ($RMSE$) of 0.22 mm ($0.84 \leq CCC \leq 0.86$) was recorded for predicted SWS via the ensemble approach, which was an improved performance over the base machine-learning algorithms. Comparing the predicted SWS to its corresponding *in situ* SWS, similar results were recorded (mean; 3.54 ± 0.03 mm and 3.56 ± 0.04 mm respectively). Predicted SWS in the shallow soil layers were highly variable, unstable and consistently dry (high temporal instability) compared to the bottom soil layers, which showed time-stable wet cluster of locations due to increase SM and clay buildup that promoted stability. The main implication of this study is that time-stable locations where crop water requirements can be met during crop growing periods can be explicitly identified for improved crop production. Our outcomes form a core support system necessary to guide the implementation of drought-adaptation measures and complement existing digital soil mapping initiatives. Further possible management options from the predicted maps include, e.g., to either prolong the existing major single farming window or identify locations and durations where additional crop-specific farming is applicable in these semi-arid landscapes. The automated modelling framework is open-sourced and robust, and can easily be updated with new *in situ* measurements.

Keywords: digital soil mapping, Guinea savannah zone, root zone, soil moisture, soil water storage, semi-arid, spatio-temporal variability, sub-Saharan Africa

7.1 Introduction

Despite crop production limitations and climate change projections in sub-Saharan Africa (SSA), smallholder farmers can meet their food demands through an adapted agricultural practice, particularly, in the Guinea savannah zone where a vast prospect for agricultural production exist (Morris et al. 2009; Xie et al. 2018). Studies have suggested that because of the vast Guinea savannah zone of SSA, it could become a major food producing area not only for the region, but also, for the rest of the world (Bourne Jr 2014; Guilpart et al. 2017). The degree to which such a suggestion holds is dependent on the sustainable implementation of ongoing adaptations. Within SSA, West Africa is of utmost concern regarding agricultural production limitations and land mass expansion (van Ittersum et al. 2016). This is because, on the one hand, West Africa has mainly accounted for ~60% of the total value of agricultural outputs in Africa for about two decades (OECD 2016). On the other hand, West Africa has a vast prospect for commercial agriculture in SSA (Morris et al. 2009). Hence, it is critical for ongoing agriculture adaptations to adequately improve production limitations so as to boost food-crop production. Historically, yield increase in SSA is partly attributed to land extensification rather than crop intensification (Brink and Eva 2009; van Ittersum et al. 2016). In this regard, van Ittersum et al. (2016) suggested that reduction in crop land extensification and yield gap closure can only be achieved by abrupt acceleration in the rate of crop intensification efforts. It is therefore key to enhance the options that are geared towards crop intensification. Such intensification options are essential due to the dwindling suitable arable lands in SSA and land-demand pressures due to increasing population (Vanlauwe et al. 2014).

Among such adaptation options has been the call for digital soil mapping (DSM) that is needed to guide sustainable agriculture, crop intensification, modelling agricultural systems and site-specific recommendations (Montanarella and Vargas 2012; Shepherd et al. 2015; Hengl et al. 2017a). Despite the proliferation of DSM initiatives to address the requirements of soil data and information around the world (Sanchez et al. 2009; Grunwald et al. 2011; Hengl et al. 2017b), SSA still has large terra incognita ahead of its DSM initiatives (Vågen et al. 2010; Guilpart et al. 2017; Leenaars et al. 2018). For instance, critical spatio-temporal SWS information of the root zone of soil that is necessary to support the ~80% population of smallholder farmers involved in rainfed agriculture is non-existent (Guilpart et al. 2017; Leenaars et al. 2018). Thus far, no study has functionally mapped root zone SWS. In addition, very little is known about the spatio-temporal variability of SWS and its potential to boost food-crop production, particularly in the semi-arid landscape of SSA (Guilpart et al. 2017). Furthermore, the missing SWS information reveals the paucity of data within the root zone of soils of SSA that can also drive dry season farming and drought adaptation interventions for notable crops of the area.

In this regard, the flagship projects of SSA, i.e. the Africa Soil Information Service (AfSIS) project (http://africasoils.net/) and the Global Yield Gap Atlas (GYGA) project (http://www.yieldgap.org/) were implemented. However, they are both limited to address the information gap on root zone SWS. This is mainly because, the AfSIS project focused on improving data availability, accuracy and resolution of soil fertility parameters (Vågen et al. 2010), whilst the GYGA project focused on gap filling for spatially explicit soil maps and root zone plant-available water holding capacity (Leenaars et al. 2018). Other such DSM initiatives in SSA and around the world includes those of Dewitte et al. (2013), McBratney et al. (2003), Forkuor et al. (2017), Hengl et al. (2015; 2017a; 2017b) and Padarian et al. (2019a), however, none reported on mapping root zone SWS of soils. In this study, we focus on filling the missing knowledge gap on the spatio-temporal SWS with explicit four-dimensional maps for the root zone of soils. Such critical contribution to existing DSM initiatives can address concerns such as (1) where (spatial location), (2) when (temporal), (3) quantity (requirements) and (4) stability of SWS available to support crops through their growing season (Chivenge et al. 2015; Mei et al. 2018). In addition, information on SWS would be useful for optimizing soil water management and the design of soil water modelling agricultural systems.

The concept of SWS is not recent, however, it has currently been acknowledged as one of the critical state variables necessary for plant growth and the sustainability of rainfed agriculture, especially in areas with soil water deficits (Zhao et al. 2017). SWS is mainly recharged by infiltration from rainfall events or irrigation and or upward groundwater movement. After field capacity is achieved, SWS reflects the size of water reservoir of soil horizons and its water storage adequacy from which crops can draw upon between rainfall events or irrigations (Guilpart et al. 2017). Within the crop roots, SWS indicates the potential of soils to buffer crops against transient water deficits that are critical at specific growth stages (Hall et al. 1982). At water deficit periods, the quantification of SWS shows the criticalness of crop productivity when evapotranspiration demands are not met by rainfall (Passioura and Angus 2010). However, due to the heavy influence of soil-landscape factors such as soil texture, evapotranspiration, precipitation, vegetation and others on SWS, water is easily lost from the soil system (Baskan et al. 2013; Biswas and Zhang 2018; Mei et al. 2019). These input-output phenomenon leads to different SWS adequacy at different soil depths and at different times. Around the world, such vertical soil water fluxes form a key component for the survival of rainfed agriculture, especially in water deficit landscapes. Due to the importance of SWS in the agricultural systems, it has been broadly applied in diverse climatic zones such as semi-arid (Mei et al. 2019; Ren et al. 2019), tropical (Adeboye et al. 2017) and temperate (Huang et al. 2018) settings, on different hillslopes ranging from upper to toe slopes (Mei et al. 2018; He et al. 2019; Mei et al. 2019), and over other land uses such as grassland (Huang et al. 2018; Mei et al. 2018), cereals (Jia et al. 2013), legumes (Adeboye et al. 2017) and forest (Zhang and Shangguan 2016; He et al. 2019). Other studies have also investigated how farm management affects SWS dynamics (Adeboye et al. 2017) and its spatio-temporal variability along hillslopes (Duan et al. 2016; He et al. 2019; Ren et al. 2019). In terms of spatio-temporal variability of SWS, topography (i.e. slope, elevation and slope positions) (Yang et al. 2015; Mei et al. 2019), soil properties (i.e. texture, depth, organic matter and bulk density) (Baskan et al. 2013; Gutierrez-Lopez et al. 2014), vegetation cover (Fang et al. 2016), and precipitation amounts and intensity are mainly the integrated site-specific driving factors. In Yang et al. (2015), topographic factors mainly drove SWS in surface and bottom soil layers of the semi-arid hillslopes of the Loess Plateau of China. Similar observations were demonstrated by Mei et al. (2018) also for the same area during wet periods. However, for dry periods, soil properties mainly account for SWS spatio-temporal variability (Grayson et al. 2002). To our knowledge, only few studies have studied the spatio-temporal variability of SWS along an entire soil profile and how its spatio-temporal distribution is influenced by multiple-environmental factors simultaneously (Zhang and Shangguan 2016; Zhao et al. 2017; Mei et al. 2018). It is important to state that these studies were mostly conducted at a traditional point-to-point scale, and thus, does not reflect the spatio-temporal distribution of root zone SWS along an entire soilscape. Thus, it is imperative that a robust spatio-temporal modelling investigation is conducted to adequately address concerns such as spatial location, temporal variability and the stability of SWS quantities simultaneously.

One thing that is common among the aforementioned DSM studies, whose approaches can be applied to model the spatio-temporal distribution of SWS, is the use of machine-learning algorithms. Acknowledging the plethora of extensive investigations and evaluations of machine-learning algorithms such as random forest, gradient boosting, support vector machine, cubist and sophisticated neural networks, which has proven to be very useful for DSM, we thus mainly focus on applying these proven approaches to model the root zone SWS. Here, we apply a multi-task modelling approach similar to the modelling concept of Padarian et al. (2019) with the widely used random forest and extreme gradient boosting algorithms. In this way, we account for the spatial structures such as patterns, proximity and time-domain dependences of the *in situ* and optimal model-input datasets. We therefore aimed in this paper to contribute to the existing DSM initiatives, by reporting on results of a complimenting DSM modelling framework for the semi-arid landscapes of Ghana. We show findings from integrated remote sensing data and time-series *in situ* SM measurements, which were used to map the four-dimensional root zone SWS. We validate the developed maps at a local scale in a 150 x 80 km agricultural catchment of Ghana. Due to large amounts of small farm sizes in the study area, which is reminiscent of the general farm size of SSA (Lowder et al. 2016), we use high spatial resolution multi-temporal radar and optical remote sensing images to infer modelling covariates. We analyze the spatio-temporal SWS changes and its stability to support recommendations for additional applicable cropping window, which is necessary to boost crop production capacities and potentials. Further, we examine the extent to which integrated site-specific driving factors influence SWS variability in the semi-arid landscape of Ghana. By these objectives, we therefore enhance existing DSM initiatives to ensure an improved adaptation option critical to support sustainable agriculture. Hence, smallholder farmer's adaptive capacity can adequately consider fundamental soil information necessary to boost crop productivity.

7.2 Methods

7.2.1 Static and dynamic covariates

In order to produce a predicted root zone SWS map for our case study areas, we followed the *scorpan* concept, which is core of DSM and has been widely used in studies with similar objectives (McBratney et al. 2003). Thus, we obtained an extensive stack of covariates for our root zone SWS modelling, which were classified as static, dynamic and process-based model input datasets. Here, from expert knowledge and literature, we considered these covariates to be relevant to soil formation and represented factors that have a causal relationship and drives the variability of SWS (Baskan et al. 2013; Mei et al. 2019). Among others were attributes that represented climate, vegetation, topography, geology or parent material and the space-time factor. Static covariates included: (1) gridded distance from rivers and soil physical properties (such as soil types, clay and silt contents, and bulk density), (2) gridded thickness of soil, regolith and sedimentary deposit layers and (3) predicted soil properties (such as organic matter and available soil water capacity until wilting point) from AfSIS data repository (Vågen et al. 2010). Dynamic covariates included precipitation, evapotrainspiration, cumulative days, day of the year and MODIS-land surface temperature (day and night). As gridded process-based covariates, we selected features derived from the Shuttle Radar Topographic Mission-DEM with spatial resolution of 1 arc-second and other spectral reflectance of remote sensing images. Among such process-based features were $SAGA$ - topographic wetness index, flow accumulation, Multi-resolution valley bottom flatness, modified soil adjusted vegetation index and profile curvature (Conrad et al. 2015).

See appendix Table S7.1 for further descriptions. Covariates were either obtained at 100 m spatial resolution or resampled to 100 m spatial resolution. For example, covariates at > 100 m spatial resolution were resampled via a cubic spline algorithm and with a bilinear algorithm for those < 100 m spatial resolution. Covariates were processed and derived in R software (R Core Team 2017) using $GSIF$ (Hengl et al. 2014), $GDAL$ (Mitchell and Developers 2014) and $raster$ (Hijmans et al. 2015) R packages with libraries and functions from $SAGA–GIS$ software (Conrad et al. 2015).

7.2.2 Root zone SWS estimation framework

7.2.2.1 Root zone SWS as a function of soil depth and time

To estimate the root zone SWS at each *in situ* data point, we used measured SM at each time-step (t) and depth (d) to recursively profile SWS as a function of the effective soil thickness and the degree of soil compaction that accounted for the effect of soil depth on the *in situ* SM measurements (Figure 7.1) (Mei et al. 2019). We assumed SWS for the subsequent depth (d), at each temporal domain (t) with temporal resolution of 12-days, as a reservoir for the preceding soil depth $(d-1)$ only. Here, ground water movement was not considered in profiling SWS, because in the study of Asiamah and Dedzoe (1999), it was reported that inherent plinthic and petro-plinthic layers that occurs at 70–100 cm in the study area cut off the movement of ground water into the upper soil depths. Hence, SM was mainly recharged from the top soil layers into the bottom layers. For these reasons, the observed changes in SWS at the ith location between depth $(d-1)$ and depth (d) at a tth time-step is proportional to the change in measured SM of the soil depths (i.e. i_{d-1} and i_d). For a time-step (t), which defines the temporal resolution at the ith location for depth (d), SWS is expressed by:

$$SWS_{itd} = 0.1 * f(SM_{itd}, BD_{id}, h_{itd})_n + K_n * f(SM_{itd}, BD_{id}, h_{itd})_{n-1} \quad (7.1)$$

where input variables for f were SM_{itd}, which is the *in situ* SM measurement (%Vol) at the ith location at the tth time-step (i.e. *in situ* measurement date) and at the dth soil depth, BD_{id}, the bulk density at ith location (g cm^{-3}) at the dth soil depth. Variable h_{itd} represents the soil layer thickness at the ith location at the tth time-step of the dth soil depth interval (cm). SWS_{itd} (mm; Equation 5.1) was estimated at a constant factor of 0.1 (i.e. the density of water). Time-steps for SWS_{itd} are in units of days, i.e. 12-day interval. Following the approach of Sure and Dikshit (2019), K_n is estimated by the function that defines the rate of gain or loss in SM_{itd} at the ith location at the tth time-step and at the dth soil depth between the initial and preceding SM measurement. K_n varies from 0 (= low gain or loss) to ± 1 (high gain or loss).

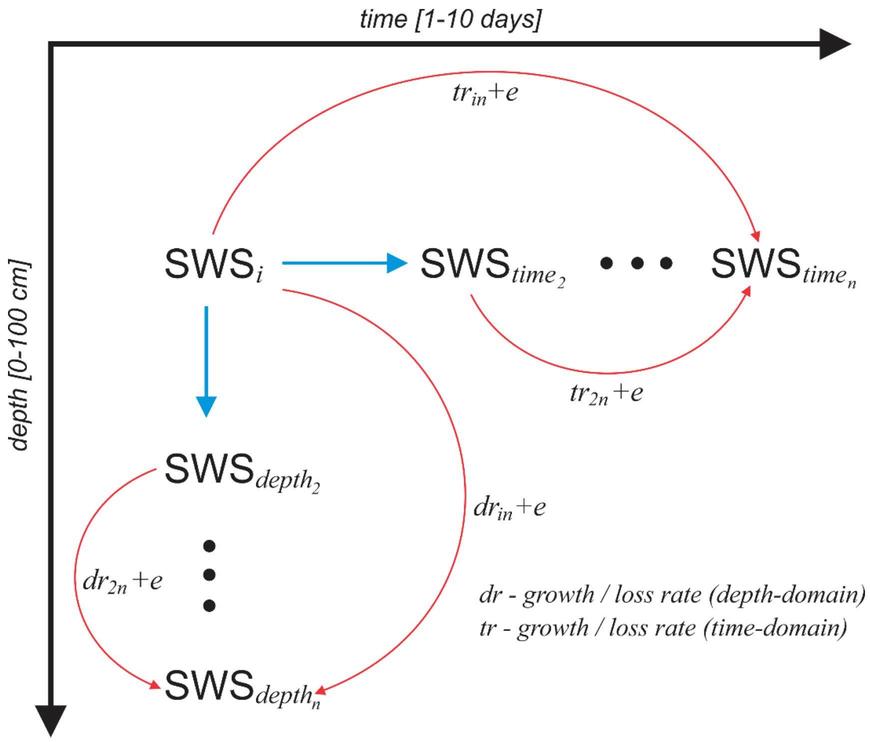


Figure 7.1. Spatio-temporal SWS as a function of soil depth at the ith location. 'tr' and 'dr' indicates the rate of growth or loss in SWS in time- and depth-domain respectively.

Furthermore, we developed two main functions (Equation 7.2–7.4) as extra model input dataset to account for the spatio-temporal autocorrelation in estimating SWS. We first used geographical covariates, estimated as *spatial-distances*, to account for proximity and spatial relations between a measurement location and the other locations (Hengl et al. 2018). Then, we calculated the *Cumulative Days* (from the start of *in situ* SM measurements) and *Day of* *the Year* to represent the *temporal-distances* in the time domain of SM measurements (Hengl et al. 2018). By this approach, we accounted for time of day impact, seasonality and long-term trends that occurred in the *in situ* SM measurements. Hengl et al. (2018) reported details on this approach.

$$\begin{bmatrix} Y_i \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_i \\ \vdots \\ X_n \end{bmatrix} * f_{std} \begin{bmatrix} (X_{gi1}, X_{ti1}) & \cdots & (X_{gin}, X_{tin}) \\ \vdots & \vdots & \vdots \\ (X_{gn1}, X_{tn1}) & \cdots & (X_{gnn}, X_{tnn}) \end{bmatrix} \tag{7.2}$$

where Y_i is the influence due to the space-time domain of SM measurement (X_i) at the ith location. Function f_{std} defines (X_{gi1}, X_{ti1}), which represents model input variables that accounted for the *spatial-distances* (X_g) and *temporal-distances* (X_t) respectively at the ith location. X_g and X_t are expressed by Equation 7.3 and 7.4 respectively:

$$X_g = (d_i, d_{i+1}, \cdots, d_n) \tag{7.3}$$

$$X_t = ([CD_i, DOY_i], [CD_{i+1}, DOY_{i+1}] \cdots, [CD_n, DOY_n]) \tag{7.4}$$

where d_i is the *spatial-distance* between a SM measurement location and others. CD_i and DOY_i are the *Cumulative Days* and *Day of the Year* respectively, used as *temporal-distance* between SM measurements X_{i1} and X_{i2}.

In this study, an effective soil depth of 100 cm was used to predict maps of the SWS at six standard soil depths (5 cm, 15 cm, 30 cm, 40 cm, 60 cm and 100 cm). Here, we vertically discretized the calculated SWS per soil depth according to *GlobalSoilMap* specifications (Arrouays et al. 2014) (Figure 7.2). For example, the trapezoidal rule (Equation 7.5) was applied to derive the weighted average of predicted SWS at depth intervals 0–5 cm or 60–100 cm.

$$\frac{1}{d_L - d_U} \int_{d_U}^{d_L} h(x)dx \approx \frac{1}{(d_L - d_U)^2} \frac{1}{2} \sum_1^{n-1} (x_n - x_{n-1})(h(x_{n-1}) + h(x_n)) \tag{7.5}$$

where $d_L - d_U$ is estimated depth interval from lower depth (d_L) to upper depth (d_U), $h(x_n)$ is SWS at the dth depth. An extract of SWS depth curves for the standard depth intervals are shown in Figure 7.2.

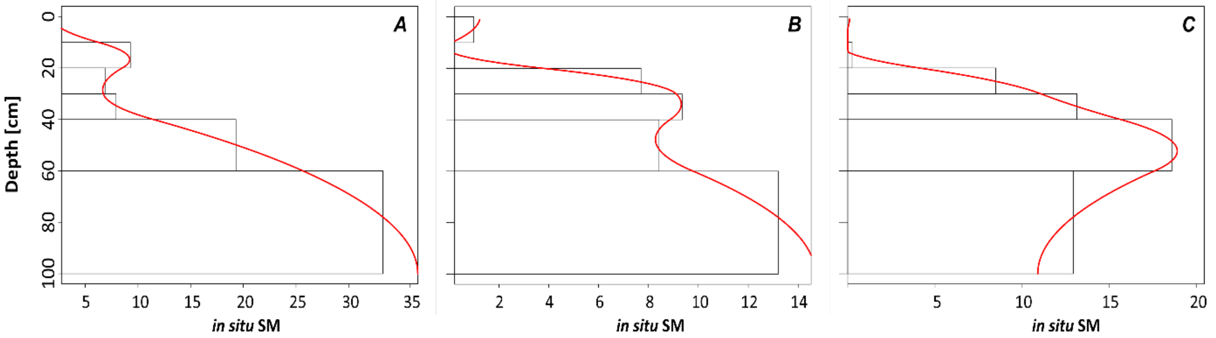


Figure 7.2. Examples of depth curves (A–C) for calculated SWS (horizontal bins) vs full soil profile estimated SWS (red lines). Calculated SWS was derived from specific in situ soil depths whilst the estimated SWS covered 0–100 cm soil depth interval.

7.2.2.2 Spatio-temporal stability and changes of SWS

We estimated the spatio-temporal stability of the SWS (TS_{itd}) at the ith location for each depth interval (d) at the tth time-step by Equation 7.6 in order to access the characteristics and persistence of spatial structures within the soil root zone (Vachaud et al. 1985; Kachanoski and Jong 1988). Following the mean-relative difference method proposed by Vachaud et al. (1985), we calculated the spatio-temporal stability. Also, such a mean-relative difference method was useful for us to account for the time-invariant correlation between *in situ* SM (driven by site-specific factors) and its spatial locations. In addition, estimated TS_{itd} allowed us to identify soil layers and areas that were consistently-wet (positive values) or consistently-dry (negative values) than the temporal SWS mean of the study area. We classified the consistently-wet or -dry areas into five interval ratings as shown in Table 7.1. TS_{itd} is defined as:

$$TS_{itd} = \frac{SWS_{itd} - \overline{SWS_{td}}}{\overline{SWS_{td}}} \tag{7.6}$$

where $(\overline{SWS_{td}})$ is the estimated mean SWS expressed as:

$$\overline{SWS_{td}} = \frac{1}{i}\sum_{i=1}^{i} SWS_{itd} \tag{7.7}$$

In addition to TS_{itd}, we investigated the temporal change of SWS across the *time-depth* domain throughout the measurement period by selecting the peak of the dry season (February 20, 2018) as the initial time-step:

$$\Delta time_{SWS(i)} = SWS_{it(n)} - SWS_{it(n-1)} \quad (\textit{time-domain}) \tag{7.8}$$

$$\Delta depth_{SWS(i)} = SWS_{id(n)} - SWS_{id(n-1)} \quad (\textit{depth-domain}) \tag{7.9}$$

Table 7.1. Interval ratings indicating temporal stability classification schemes for SWS values

Interval	Interpretation	Ranking
> 0.5	Pixel values that are extremely higher than the temporal mean pixel value. Classified as consistently-wet for the corresponding time-step.	Wet
> 0.0 – 0.5	Pixel values that are higher than the temporal mean pixel value. Classified as a gradient of neutral and consistently-wet for the corresponding time-step.	Slightly wet
0.0	Pixel values that are equal to the temporal mean pixel value. Classified as unchanged.	Neutral
< 0.0 – 0.5_n	Pixel values that are lower than the temporal mean pixel value. Classified as a gradient of neutral and consistently-dry for the corresponding time-step.	Slightly dry
< 0.5_n	Pixel values that are extremely lower than the temporal mean pixel value. Classified as consistently-dry for the corresponding time-step.	Dry

Negative values are assigned with the subscript ‘neg’.

7.2.3 Mapping spatio-temporal SWS

7.2.3.1 Recursive feature elimination

To obtain the optimal set of predictors for mapping root zone SWS for our case study areas, we used random-forest recursive feature elimination (RF-RFE) algorithm to subset the extensive stack of covariates (Kuhn 2018). By this way, productive features were also selected for the target-oriented cross validation of SWS predictions. Iteratively, with our expert knowledge and the RF-RFE algorithm, we eliminated features considered least important from the a *priori* model that included all covariates. Here, covariates were ranked using the *Gini* index. On the one hand, highly correlated covariates with highest correlations were eliminated whilst on the other hand, the RF-RFE algorithm selected covariates ranked with high importance that constituted the optimal prediction set with the SWS values (Kuhn 2018). This enrichment step ensured reduced dimensionality in the stack of covariates and the usage of only relevant features in the final model predictions. The optimal RF-RFE model was evaluated using $10 - fold$ cross validation, which was repeated 5-times. Model performance metrics used to assess selected covariates were root mean square error (RMSE), R-square (R^2) and mean absolute error (MAE). RF-RFE was implemented in R software (R Core Team 2017) with the $caret$ package (Kuhn 2018).

7.2.3.2 Ensemble prediction model

We used a hybrid modelling approach that combined two tree-based algorithms to predict root zone SWS: the Random Forest for spatial data (RFsp) (Hengl et al. 2018) and the extreme Gradient Boosting (XgBoost) (Chen and Guestrin 2016) algorithms. In study 3, we found that RF algorithm performed poorly and overestimated SM at low and extremely low *in situ* SM values in the study area. This finding was partly driven by the sparse nature of the *in situ* SM measurement network, and thus we included the XgBoost to address this limitation. This is because, in sparse data prediction systems similar to our measurement network, DSM benchmarks found that XgBoost is among the best performing algorithms in sparse data systems (Chen and Guestrin 2016; Ramcharan et al. 2018). Other reasons for combining RFsp and XgBoost were due to their high prediction accuracies in complex interactions among covariates (Hengl et al. 2017b) and their effective parallelization for large number of spatio-temporal pixels (Ramcharan et al. 2018).

Random forest algorithm in itself uses an ensemble of nested decision tree predictors to estimate a target variable (Breiman 2001). It iteratively draws random sub-samples from the training sub-sample and generates an average of individually constructed decision trees for the final prediction. This random selection allows a robust out-of-bag error estimate necessary to assess the model fit. It is important to note that only the optimal predictors drawn from the random sub-samples were used to partition the training sub-sample at the respective splits (Meyer et al. 2018). The working principle of the RFsp system (Hengl et al. 2018) follows the traditional random forest algorithm, except that the traditional approach used for DSM calibrates a target variable as a function of the intersection between the soil property and covariates that define its occurrence. Such an approach only reflects a point-to-point relationship and neglect the spatial structures such as patterns, proximity and time-domain dependence especially in the temporal domain (Hengl et al. 2018; Padarian et al. 2019a). However, the RFsp system used here leverages on the spatial structures in the model input datasets to address this limitation of the traditional random forest algorithm (Hengl et al. 2018). See Breiman (2001) and Hengl et al. (2018) for details on random forest and RFsp algorithms respectively.

XgBoost's scalable tree boosting approach allows weighted quantile sketches to handle weights in the approximation of learning trees (Chen and Guestrin 2016). Here, an ensemble of trees uses additive functions to sum up scores in the corresponding leaves of the sub-decision tress for the final prediction. In an additive way, the XgBoost model uses learned patterns to train itself in order to refine its parameters. Also, overfitting in the XgBoost model is prevented by shrinkage of newly introduced weights after each tree boosting and feature sub-sampling as implemented in traditional random forest. However, an optimized model is driven by, among others, the parameters of tree shrinkage, number of iterations, trees branches, influence of each iterative tree, and percentage of sub-samples to be used for sampling. See Chen and Guestrin (2016) for further details on XgBoost algorithm.

7.2.3.3 Multi-task model implementation

A parallelized multi-task ensemble modelling approach was adopted in this study. This was because, in a multi-task modelling approach, learned model representations of the predicted property of interest, during the model training phase, is shared in the final prediction framework (Padarian et al. 2019a). In other machine learning systems where such multi-task approaches were used, it was demonstrated that the multi-task approach reduced model prediction time and overfitting, and computational resources, which consequently reduced model prediction errors (Ruder 2017; Padarian et al. 2019b). We simultaneously predicted SWS at the six standard depth intervals and at the ten time-steps (Figure 7.3). The multi-task ensemble modelling was executed in R software (R Core Team 2017) on the University of Göttingen's GWDG high performance computing clusters. The base RFsp and XgBoost algorithms were implemented with R packages: $ranger$ (Wright and Ziegler 2015) and $xgboost$ (Chen and Guestrin 2016) respectively. In the ensemble model, we applied a weighted average where each base algorithm was weighted by their respective performance error. By this approach, we linearly combined both RFsp and XgBoost to derive the final SWS prediction and thus, address the limitation of equal contribution by RFsp and XgBoost in the ensemble model.

In order to optimize model implementation, we partitioned the stack of optimal model-input covariates into stacks of spatial tiles that covered a spatial extent of 10 x 10 km per block with a 10% overlap. Prior to the multi-task ensemble modelling, all open-surface water covers such as irrigation dams, large river bodies and ponds were masked out. Subsequently, we parallelized the multi-task ensemble model with the generated spatial tiles to predict SWS at each time-step and soil depth interval. Spatial tiles of predicted SWS at each time-step and soil depth interval were later mosaicked into a single raster file (Figure 7.3). In both algorithms, optimized model training configuration was achieved via a grid search, which was resampled with a $10 - fold$ cross validation and repeated 5-times. To obtain an optimal balance between over- or under-fitted model by the RFsp, we used a *mtry* of 45 from a sequence of *mtrys* (i.e. from 1 to 150 by 5 units) for hyperparameter tuning. For the XgBoost model, we controlled the learning rate and the maximum depth of a tree with an *eta* of 0.4 and *max_depth* of 4 respectively. The other tuning parameter used in the XgBoost model were *min_child_weight* (instances set in terminal node = 1), *gamma* (minimum loss reduction for leaf node partition = 0) and *colsample_bytree* (columns to sample in tree construction = 0.8).

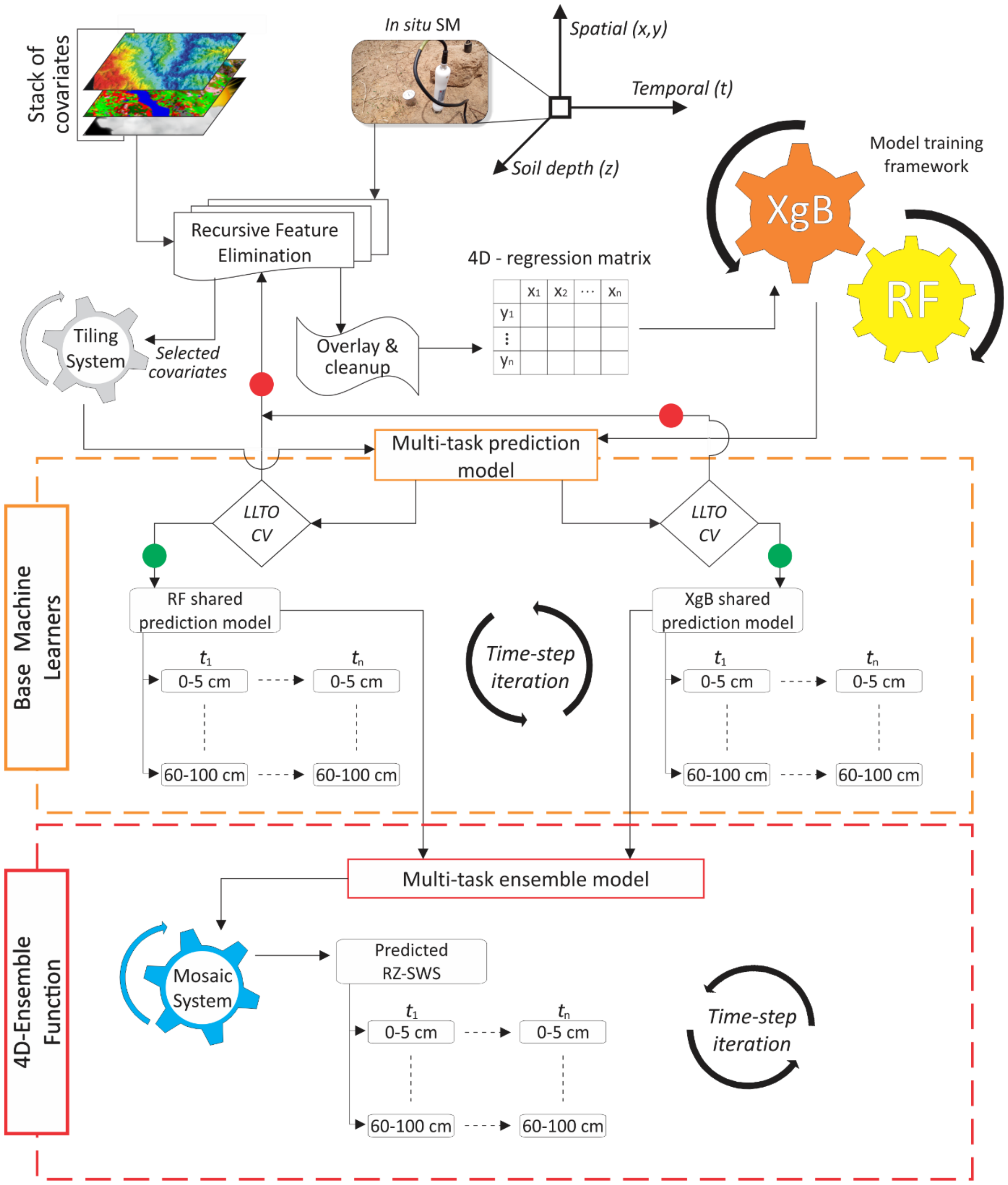


Figure 7.3. Methodological flow of the multi-task ensemble modelling. t_1 represents the first time-step whilst t_n indicates the nth time-step. Section of chart labelled 'Base Machine Learners' describe the parallelised iterative prediction of SWS at each soil depth interval by RFsp (RF) and XgBoost (XgB) algorithms.

7.2.3.4 Target-oriented accuracy assessment

We tested the SWS model performance (i.e. fitting and validation accuracy, model bias and stability) using a *target-oriented* cross validation (CV) as proposed by Meyer et al. (2018). In the *target-oriented* CV, we used a Leave-Location-and-Time-Out CV (LLTO-CV). Data partitioning were stratified into $n - folds$ using the ith location's identifier and its DOY_i in order to ensure equal representation of target variable in each fold. In training the LLTO-CV model, all locations and time-steps (i.e. folds) except one location and a time-step (a fold) were used whilst the held-out fold (i.e. a location and a time-step) was then used as a testing sub-sample to assess the LLTO-CV performance. This step was repeated for all other folds in which a held-out fold in each iteration was later used for model testing. Calculated SWS were log-transformed prior to its use in the ensemble modelling due to its skew distribution. Consequently, we used a log scale to report on the LLTO-CV accuracy statistics. However, predicted maps were back-transformed to their original values. Performance measures used to assess the SWS predictions were: R^2 of the LLTO-CV residuals and the *RMSE*. In addition, the concordance correlation coefficient (CCC) at 95% confidence interval was used to assess how the estimated SWS deviated from the 1:1 line of perfect concordance after the LLTO-CV (Steichen and Cox 2002) (appendix Equation S7.1, S7.2 and S7.3 respectively).

7.2.4 Multi-task modelling framework

As part of open and reproducible science, we contribute by providing detailed workflow including procedures used for the SWS modelling framework on an open-access repository (DOI: 10.5281/zenodo.3662411). Other such documents available for use includes R scripts, metadata and readme files. Full datasets are publicly available on the same online repository.

7.3 Results and discussion

7.3.1 Ranking the importance of covariates for SWS modelling

Figure 7.4 shows the top 12 model input features for our case study areas. Across the considered RF-RFE selection levels (i.e. Figure 7.4A–C), soil depth was ranked consistently highest amongst the features used for predicting SWS. This result suggests that soil depth is a crucial factor that influences the spatio-temporal patterns of SWS in the study area. Our results are consistent with the observations of Mei et al. (2018), who found similar results in the semi-arid Loess regions of China. In addition, it is an indication of strong correlation between *in situ* measuring soil depths and the predicted SWS, as depicted in appendix Table S7.2. Similar strong correlation and feature importance with regards to soil depth were demonstrated for other soil properties by Ramcharan et al. (2018). On the first level (Figure 7.4A), soil depth, available water content, bulk density and sand content were considered the most important features. However, from the perspective of soil survey, this observation is not entirely supported by expert knowledge as documented in the literature for SM (Jost et al. 2005; Gao and Shao 2012; Zhao et al. 2017; He et al. 2019). The possible explanation for this observation lies in the *target-oriented* errors which are hardly considered by the RF-RFE algorithm (Meyer et al. 2018). In Ramcharan et al. (2018) and Leenaars et al. (2018), it was also reported that feature selection importance metrics could be limiting due to feature correlation, and thus, impact less important features (e.g. at the 24th *in situ* location) from receiving boosted importance. For instance, covariate 'Bdst24' was ranked among the top 4 features in the stack of covariates. This geographical distance covariate (at the 24th location) was derived from a toe slope soil location with notably high SWS value. Hence, such strongly correlated variables are interchangeably used in the RF-RFE (Strobl et al. 2007). This observation appears to be one of the drawbacks of RF-RFE as it mainly considers the internal importance (Meyer et al. 2018).

As highly correlated covariates such as geographical distance could impact the feature selection process, in addition to its computational intensiveness during its development, it would be practical for such covariates to be derived for clusters of target variable (Hengl et al. 2018). For this purpose, the feature selection process would consider the general spatial structures and apply boosted importance equally. Although the ranked features on the first level are not fully supported by expert knowledge, they showed continuous consistency of importance at the different RF-RFE selection levels (Figure 7.4). Cumulatively, including additional features based on expert knowledge (subsequently ranked important variables) such as clay and silt contents, surface roughness, day of year, soil types and landform classes (Figure 7.4A and 7.4B), the RF-RFE showed marked differences in R^2 and RMSEs. The downward change of feature importance from depth to the next important variable, i.e. available water content, was abrupt but changed gradually across the rest of the stack of covariates (Figure 7.4). Generally, soil physical properties such as clay, silt and sand contents showed higher importance after soil depth as compared to the process-based and dynamic covariates. This observation agrees with the generic *scorpan* concept of optimal features used for DSM (McBratney et al. 2003).

From a soil-landscape standpoint, although RFE is time consuming, it is critical that model feature selection is guided by expert knowledge of factors that drives the variability of the target variable of interest, while considering the model selection performance errors. Therefore, the *Gini* importance cannot be the sole feature selector in this case, as such, we limited the use of RF-RFE results for selecting model input features but rather coupled the selection process with expert knowledge of the study area.

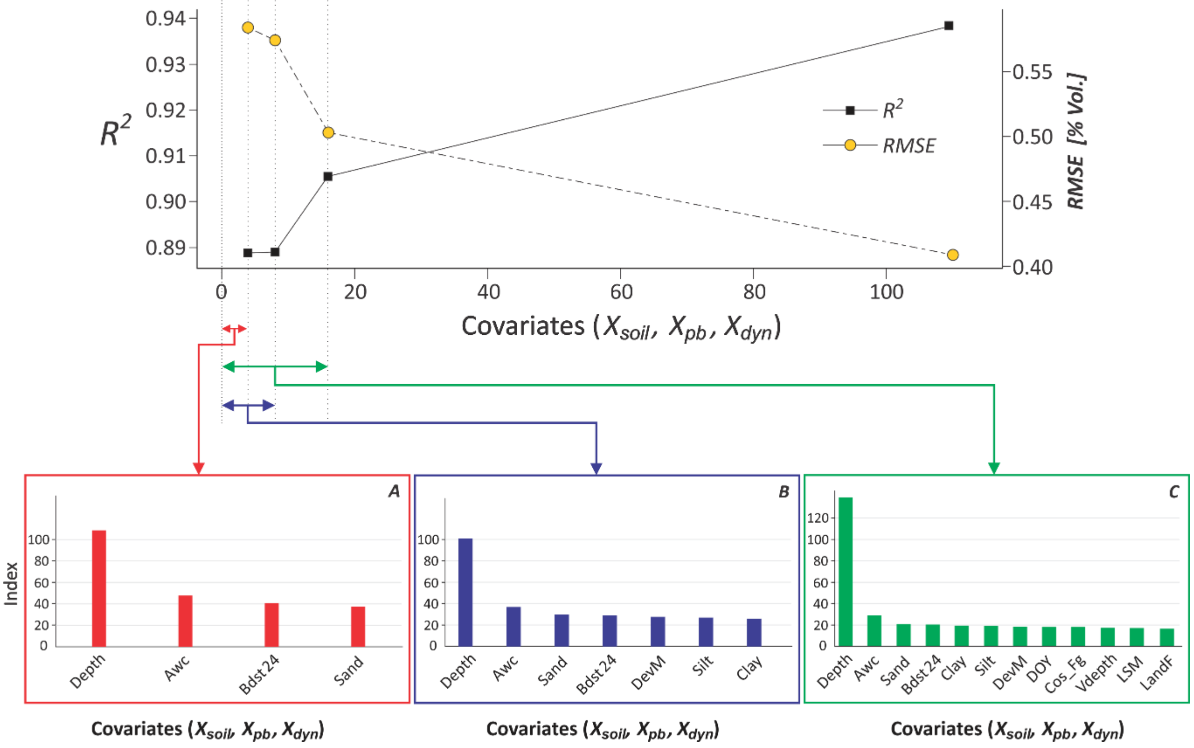


Figure 7.4. Selected optimal features based on the RF-RFE algorithm and expert knowledge used for the SWS prediction model. Appendix Table S7.1 presents further description on the codes used to represent the covariates. A–C shows top three levels of feature importance with their corresponding optimization measures (R^2 and RMSE). Section 'A' represents top 4 selected covariates that explain the variability of SWS in the study area (RMSE = 0.58 mm, R^2 = 0.89% and MAE = 0.43 mm). Section 'B' presents top 7 selected covariates at model performance of RMSE = 0.57 mm, R^2 = 0.89% and MAE = 0.42 mm, whilst 'C' shows marked model performance with errors RMSE = 0.50 mm, R^2 = 0.91% and MAE = 0.35 mm for top 12 optimal covariates. X_{soil} represents gridded physical and chemical soil properties. X_{pb} represents process-based covariates derived from DEM whilst X_{dyn} represents dynamic covariates.

7.3.2 Multi-task model performance

The performance measures used to assess the multi-tasked modelling approach of SWS are of two-fold: (1) the internal model fitting performance statistics during the training phase and (2) the *target-oriented* LLTO-CV performance statistics between calculated and predicted SWS (Table 7.2).

During the model fitting phase, with hyperparameter tuning and an optimal model configuration, both the RFsp and XgBoost showed comparable performance measures (Table 7.2). R^2 and RMSE for RFsp were 0.96 (based on out-of-bag) and 0.32 respectively, and 0.96 (based on repeated-CV) and 0.29 for XgBoost respectively. Although we sought to address the limitation of the RF algorithm for sparse data systems, the RFsp rather showed a generic robustness to locality similar to the performance of the XgBoost algorithm. In addition, RFsp was resistant to the low-inflated SWS values for most measurement locations (appendix Table S7.3). This suggest that in sparse measurement networks, RFsp and XgBoost can equally be used to achieve satisfactory results. In the present study, using only the RFsp may be suitable. Another possible reason for this finding was the inclusion of geographical distances as covariates in the model input datasets. In Hengl et al. (2018), it was suggested that such model input data structure distinctly smooths out predictions and thus, result in satisfactory performances. Furthermore, the observed localized variations characterized by the model input dataset complements the spatial distribution of SWS and, therefore, might be better represented by the RFsp and XgBoost (Hengl et al. 2018; Chen and Guestrin 2016). Although the performance measures of RFsp bears a close resemblance of the XgBoost, it is important to note that the XgBoost showed a slight improvement over the RFsp.

Table 7.2. Detailed model regression statistics via the multi-task modelling framework

Model	**ME**	**MAE**	**RMSE**	***R*²**	**CCC**	**CCCL**	**CCCU**
Internal							
RFsp	-	0.199	0.323	0.956	-	-	-
XgBoost	-	0.196	0.294	0.964	-	-	-
Target-oriented LLTO-CV							
RFsp	0.023	0.136	0.221	0.747	0.841	0.829	0.853
XgBoost	0.022	0.142	0.227	0.730	0.846	0.833	0.858
Ensemble	0.023	0.136	0.218	0.752	0.850	0.838	0.861

Model performance metrics were mean error (ME), mean absolute error (MAE), root mean square error (RMSE), R-square (R^2) and Lin's concordance correlation coefficient at 95% confidence interval (CCC) with its lower and upper limits. We show the internal-model fitting performance statistics during training phase (Internal) and the target-oriented LLTO-CV performance statistics between calculated and predicted values of SWS (*Target-oriented* LLTO-CV).

While the internal model fitting performance showed a somewhat “perfect-fit” of the training data, the *target-oriented* LLTO-CV rather showed a slight reduction in model performance (i.e. their respective R^2). Downward ΔR^2 was 0.21 for RFsp and 0.23 for XgBoost (Table 7.2). This is an indication of less robustness of the model beyond its training data. The example by Meyer et al. (2018) where such performance decreases were demonstrated with the LLTO-CV approach corroborates these findings. Although the essence of LLTO-CV is to test the robustness and efficiency of a prediction model beyond its training data locations in space and time (Meyer et al. 2018), it is also important to draw evaluation conclusions with regards to the adequate representation of the target variable also in space and time. Particularly, with the use of model’s temporal error (i.e. RSME), which has a high effect on the reliability of the prediction due to complete elimination of locations and time-steps. Comparing the *target-oriented* LLTO-CV performances for both RFsp and XgBoost, they also showed similar model performance measures (Table 7.2 and Figure 7.5). This similarities for both RFsp and XgBoost shows, in its *space-* and or *time-domain*, high ability to predict SWS at an unknown location or time with certainty. Temporal errors for RFsp was 0.22 mm ($0.83 \leq \text{CCC} \leq 0.85$) and 0.23 mm ($0.83 \leq \text{CCC} \leq 0.86$) for XgBoost. The CCC used here can describes the overall variance of the predicted SWS that is explained by the multi-task modelling approach. Supporting this result, from the ME of the LLTO-CV (Table 7.2), neither the RFsp nor XgBoost showed clearly any under- or overestimation of SWS (ME nearly zero). Overall, the models accurately predicted the variability of the reference calculated SWS (Table 7.2 and Figure 7.5).

One key reason of an ensemble model is to bag “weak learning” algorithms into a “strong learning” algorithm (Kotu and Deshpande 2018). In the present study, the ensemble of the base machine learners showed similar performance as the base machine learners in the prediction of SWS: RMSE was 0.22 mm ($0.84 \leq \text{CCC} \leq 0.86$). This is also visible from the fairly symmetrical correlation plot presented in Figure 7.5.

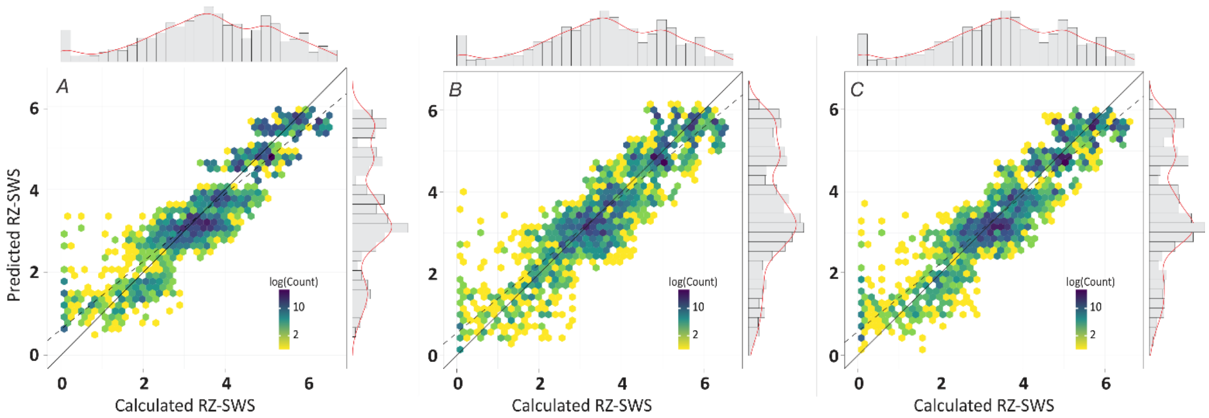


Figure 7.5. Correlation plots from target-oriented LLTO-CV between calculated and predicted values of SWS with goodness-of-fit (dashed line) and 1:1 line of perfect concordance (solid line). Table 7.2 presents further description on performance measure statistics. 'A' shows the relationships based on RFsp model, 'B' shows that for XgBoost and 'C' for the ensemble model. See appendix Figure S7.1 for further details on the model variability and mean performance differences.

Mean predicted SWS (± standard error; on a log-scale) were equal for all base and ensemble models (3.54 ± 0.03) but negligibly lower compared to the calculated SWS values (3.56 ± 0.04) (appendix Figure S7.1). However, variability of predicted SWS (range) was higher in the XgBoost model (6.61) compared to the ensemble (5.89) and RFsp (5.44) models (in order of reducing variability). These results indicate that despite similar performances from the prediction models, RFsp and the ensemble model predicted slightly lower SWS values compared to the calculated SWS (6.60). Generally, the RFsp and XgBoost show outcomes that agree with the calculated SWS. One possible reason for these findings is that tree-based machine learners are able to represent adequately the local soil-landscape interactions, which often results in low performance errors (Minasny and McBratney 2010; Hengl et al. 2018; Ramcharan et al. 2018). Such local soil-interactions were initially accounted by the sampling design adopted in this study (Nketia et al. 2019). Here, the local spatial structures and heterogeneities were defined and accounted by the global weighted principal component analysis, while suppressing areas that were less influencing and representative of SM by the cost-constrained conditioned Latin hypercube algorithm. Although ensemble models can achieve higher model improvements (≤ 30%), it does not always translate as improved accuracies in DSM (Michailidis 2017), supporting this point is Figure 7.5 and appendix Figure S7.1. It is important to compare base learners to assess their efficiency and risk for the target variable of interest prior to combining such base algorithms.

7.3.3 Spatio-temporal SWS as a function of soil depth

SWS vary strongly both with depth and time. Appendix Table S7.3 describes in details the statistics of spatio-temporal SWS at each standard depth interval. The vertically discretized *in situ* SWS used to develop standard soil depth intervals showed no differences (mean ± standard error) between 0–5 and 5–15 cm, probably due to the short thickness of the *in situ* measurement depths (i.e. at 0–10 and 10–20 cm). However, from 15–100 cm there were clear increases in temporal mean SWS with increasing soil depth. Within the study period, at 0–15 cm depth, the lowest temporal mean SWS was recorded with values ranging from 1.59 mm (peak of the dry season) to 7.16 mm (onset of heavy rains) whilst the highest was recorded at 60–100 cm (from 117.53 mm to 131.86 mm). This observation was inversely proportional to the coefficient of variation observed over time with increasing soil depth, and thus, showed a high spatial variability at 0–5 cm as compared to the other soil depth intervals (appendix Table S7.3).

Generally, the lowest SWS (range) at all standard soil depths were recorded at February 20, 2018 (~9 mm; peak of the dry season), whilst all other time-steps showed relative increases from preceding time-steps – the highest occurring at the onset of heavy rains (~176 mm; June 6, 2018). Meanwhile, regardless of time-step, specific cluster of locations showed consistently high SWS with increasing soil depth; thus, the toe slope soils (i.e. on Siare, Volta and Dagare soil types). Specifically, locations with low spatial distances from river bodies were much affected. One possible reason for these cluster of locations is due to their morphological characteristics and geographical location. Siare and Volta series are colluvio-alluvial sediments along the major Volta river and its large tributaries in the study area (Adu 1995). Their proximity to the Volta river influences its consistent high SWS despite the period of the year. Whereas, Dagare series are periodically deposited alluvial sediments on small river banks. Consequently, receded rivers during the dry season impacts its storage capacity and vice versa. This finding further indicates that despite the toe slope soils recorded consistently high SWS, Siare and Volta series' SWS is highly influenced by river networks than the Dagare series, which lose most of its water in prolonged periods of drought.

From Figure 7.6, the extent of time-stable cluster of locations increased with increasing soil depth interval with preceding time-steps. This result extends the findings of He et al. (2019) in that increasing soil depth leads to evenly distributed SWS, which weakens the degree of fragmentation and enhance spatial continuity of SWS hotspots. Similarly, Gao and Shao (2012) demonstrated such soil-depth drift of SWS variabilities, which further complement our findings. Extracts of spatio-temporal gridded SWS maps based on the ensemble model for all standard soil depth intervals are presented in Figure 7.6. All maps (n = 60) are publicly available on the project's open-access repository. The observed simultaneous spatio-temporal trends for SWS at the various soil depth intervals follows studies around the world, supporting this are the studies of Gao and Shao (2012) and Mei et al. (2019) for the 0–100 cm depth interval. However, these observations are site-specific, scale-dependent and not universal as it is in contrast to the findings of Hu et al. (2010) for similar soil depth intervals. In the study area, precipitation is the sole source of SWS recharge (Adu 1995). This is because of the sandy textured top soils, high evapotranspiration and inhibition of upward groundwater movement due to inherent petro-plinthic soil layers (Asiamah and Dedzoe 1999).

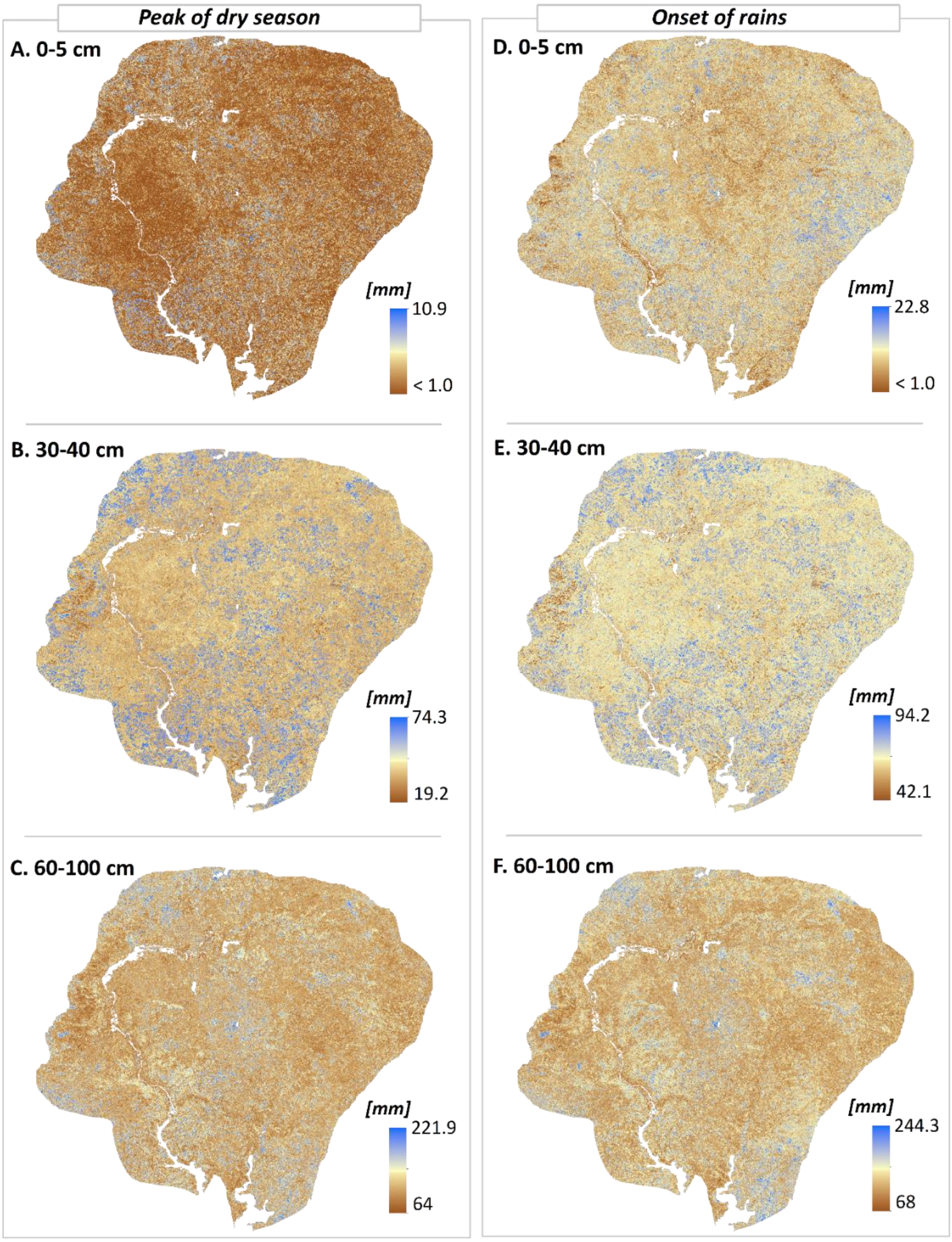


Figure 7.6. Example of gridded spatio-temporal SWS maps based on the ensemble model for the standard soil depth intervals. A–C shows those for the peak of the dry season (February 20, 2018) whereas D–F presents those for the onset of heavy rains (June 6, 2018). Complete set of maps for this study are accessible on the GitHub repository.

In the study area, two main possible explanations could be given for the observed phenomenon in the bottom and surface soil layers. First, in the bottom soil layers are the derived pedogenetically variations and the inherent plinthic and petro-plinthic layer that occurs at ≥ 70 cm in the soils of the study area (Asiamah and Dedzoe 1999). These soil layers and internal processes tend to inhibit downward movement of soil water, suggesting that stagnation occurs at ≥ 70 cm. The inherent plinthic and petro-plinthic layers further makes it difficult for SM to be replenished from groundwater. Consequently, increasing the size of water reservoir of the bottom horizons and its water storage adequacy. Second, in the surface layers, evapotranspiration demands are unmet by precipitation. Intensive evapotranspiration and low amounts of precipitation decreases SWS, particularly in surface soil layers of < 40 cm soil depth (Greenland et al. 2018). For instance, Oki and Kanae (2006) found evapotranspiration to utilize ~60% of rainfall input in surface soil layers. From Study 2, we also demonstrated that the intense variability of SM in the surface layers of the study area was primary driven by evapotranspiration, soil texture, precipitation and other topographical features. Furthermore, daily evapotranspiration in the Guinea savannah is reported to be ~3-times the daily average precipitation, suggesting high loss of SM. All these observations coupled with the sandy 0–15 cm soil depth contributes to the recorded high variability of the surface layers. Meanwhile, the high evapotranspiration makes it difficult for water to percolate into the sub-surface soil layers (within 5–30 cm), which consequently affected its storage capacity. Other studies have found evapotranspiration, precipitation and soil texture to highly influence the variability of SM in shallow soil layers of other landscapes (Seneviratne et al. 2010; Huang et al. 2016; Biswas and Zhang 2018). Land cover has also been reported to largely affect SWS patterns in soil-plant systems (Zhao et al. 2017), however, due to the different characteristics and interspersed occurrence of *Khaya senegalensis*, *Borassus aethiopum* and *Vitellaria paradoxa* in the study area, effects due to land cover is less evident in the bottom soil layers. In contrast to the surface soil layer, the abundant roots of elephant grass (*Pennisetum purpureum*) might have contributed to its high variability due to root water uptake, in agreement with Seneviratne et al. (2010).

Over the study period, SWS at all soil depth intervals showed positive correlations between time-steps, i.e. from 0.3% to 1.0% (Supplementary Table S7.2). The strongest relationships were recorded at 60–100 cm depth interval, whilst the weakest were recorded at 0–5 cm depth interval. This observation corresponded with the high-low temporal means of SWS. However, the strength of the relationships degrades over time which indicates a *time-domain* drift with increasing soil depth. For instance, at 0–5 cm, the correlation between time-steps decayed from 0.90% (strong; February 20, 2018) to 0.32% (weak; June 8, 2018). This result is exemplified in all other soil depth intervals with increasing time-steps (Supplementary Table

S7.2). These findings suggest that factors such as evapotranspiration, precipitation and plants roots, that highly drove the variability of SWS in the shallow soil depths were less influencing in the deeper soil depths. Significant correlations were reported for soil texture, particularly clay content, and SM for soils of the study area (Study 2), which further suggest that the increase in clay content in the bottom layers makes water holding capacity high for the bottom soil layers. For these reasons, regardless of vegetation cover and other soil-landscape factors, SM is the main dependent condition that drives the high dependencies in the bottom layers. Thus, SM mainly determines the spatial heterogeneities of SWS with increasing soil depth (Gao and Shao 2012). Again, supplementary Table S7.2 further shows that the explanatory power of SM and increasing soil depth, which determines the variability of SWS, was high for the bottom soil layers. Overall, in the surface soil layers, SWS was highly affected by soil-landscape factors; in contracts, SWS in bottom soil layers was mainly controlled by SM and soil texture.

7.3.4 Spatio-temporal change in SWS as a function of soil depth

The temporal changes in SWS observed in the soils of the study area reflect soil desiccation and depletion of infiltrated rainfall with soil depth (Figure 7.7). From February to June 2018, an increasing trend in SWS was observed. The peak of the dry season (February 20, 2018) was selected as the baseline for determining the temporal changes in SWS. At 0–40 cm depth interval, consistently extremely-low to low SWS (< 10 mm) across the study area were observed, whereas, at > 40 cm relatively higher SWS (> 30 mm) were recorded. The severity of the extremely-low to low SWS decayed with increasing time-steps. June recorded the highest SWS across the *depth-time* domain (Figure 7.7). The temporal change in SWS may be attributed to antecedent SWS, which is mainly driven by antecedent precipitation and precipitation of the study area (Study 2). In addition, the variability of other environmental influencing factors such as topography, vegetation and soil texture, might also be the cause of the large temporal changes in SWS. These observations suggest that SWS improves with increasing time-steps, and thus, the bottom soil layers has potentials to support food crop production – as such moisture reserves can be tapped to support crop growth.

The tipping point in the temporal change in SWS, where marked increases were recorded, was in April (onset of rains) and peaked in June (onset of heavy rains). This observation is consistent with increasing soil depth and time-steps, suggesting an increase in storage adequacy. Supporting this point, is the underlying plinthic and petro-plinthic layer that occurs at ≥ 70 cm in the soils of the area and the interspersed occurrence of trees to promote root water uptake (Asiamah and Dedzoe 1999). However, another possible reason for the severity of changes in temporal SWS, particularly in the surface soil layers (≤ 30 cm), is due to the heavy occurrence of elephant grass. Root structure and vegetation significantly change and controls spatio-temporal pattern of SWS (Jia et al. 2013). The roots of elephant grass are heavily distributed within the 0–50 cm depth (WCSP 2014). The predominance combined effect of evapotranspiration demands being unmet by precipitation, the sandy top soils and water uptake by the abundant elephant-grass roots might be the explanation for the recorded extremely-low to low changes in the surface soil layers. Meanwhile, the infiltrability of the surface soil layers were high, which promoted the storage adequacy of the bottom soil layers. In contrast, changes in the bottom soil layers is mainly controlled by SM and clay build-ups due to the low effect of other environmental influencing factors reported for the surface soil layers. Thus, SWS accumulated over time and increased with increasing soil depth, in agreement with the findings of Mei et al. (2019) and Gao and Shao (2012).

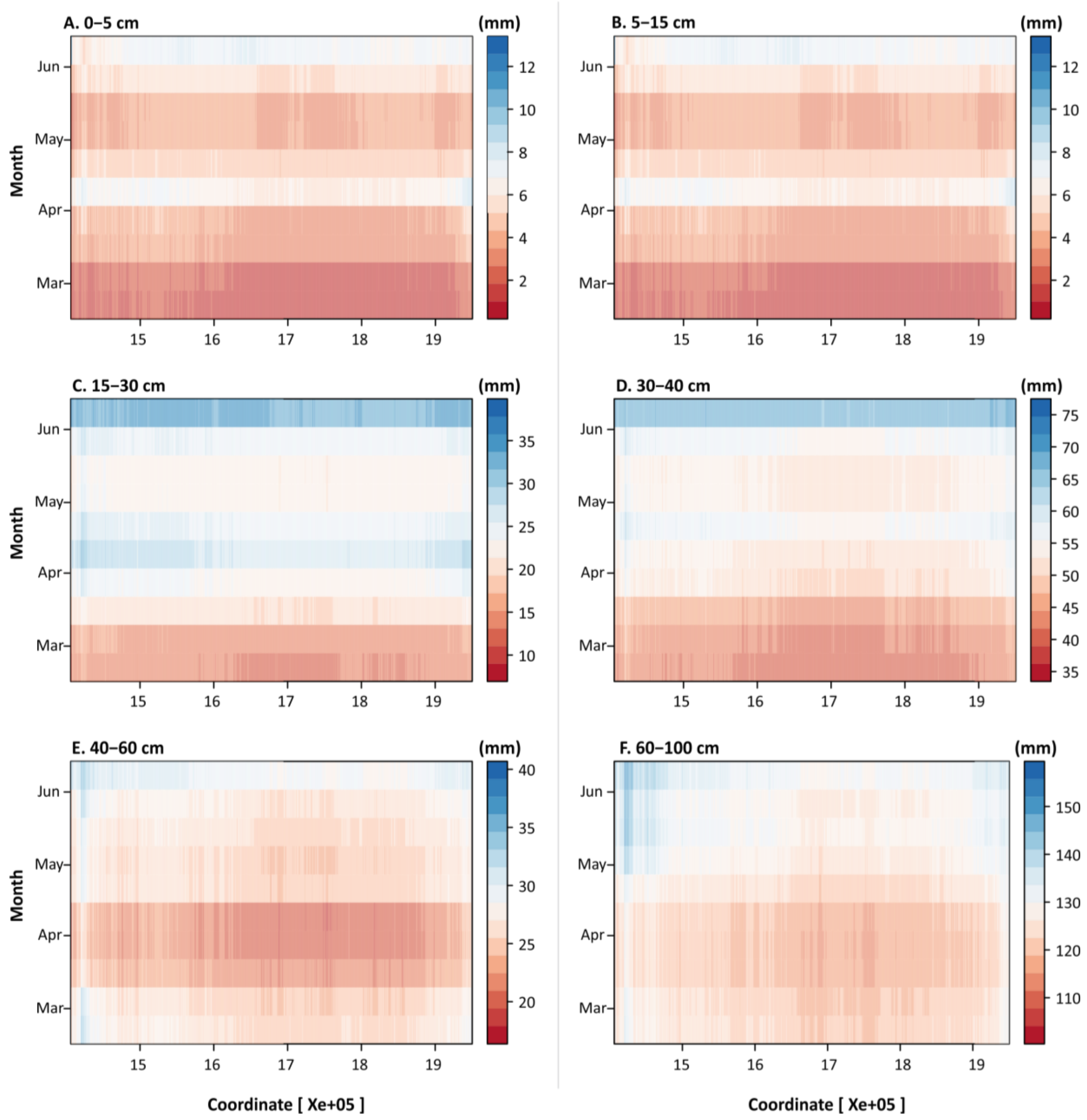


Figure 7.7. Spatio-temporal changes in SWS at each gridded location (i.e. pixel values) for each standard depth intervals from February 20th to June 8, 2018. 'X' values are in a measure based on cartesian coordinates.

7.3.5 Spatio-temporal stability of SWS as a function of soil depth

Under the five interval ratings (; Table 7.1), we found consistent time-stable SWS locations across the various standard depth intervals (Figure 7.8). Supporting this is the strong correlations between time-steps with increasing soil depth (appendix Table S7.2). These results are consistent with those of Duan et al. (2017) in other geographical setting. Overall, the temporal stability across the depth intervals showed a *time-domain* drift, particularly in the ≤ 15 cm depth interval. However, the spatial heterogeneities of these time-stable locations showed satisfactory characteristics of time stability before and after the tipping point of SWS temporal trends (in April; onset of rains), especially at ≥ 30 cm. In the surface layers, temporal stability were patchier (greater instability) compared to the bottom layers, suggesting more temporally stable areas – in agreement with the explicit results of He et al. (2019), Duan et al. (2017) and Jia et al. (2013). The decrease in patchiness at ≥ 30 cm, before and after the tipping point, can be explained by the less influence of environmental factors on SWS in the bottom soil layers, but rather was mainly influenced by SM and soil texture. Our findings follow those of Jia et al. (2013). The greater instability seen in the < 30 cm may be attributed to root water uptake by the abundant elephant-grass and higher variability of rainfall, evapotranspiration and the sandy top soils (He et al. 2019).

We found abrupt transition periods between time-steps and soil depth intervals, which showed an enhanced instability across the study area (Figure 7.8). This enhanced instability is likely due to the temporal resolution of the sampling strategy (i.e. 12-day interval). It has been found that SM changes rapidly at small distance and small-time intervals (Baskan et al. 2013), hence, it would be necessary to decrease the time-lag between measurement networks in order to have a better understanding of the transition periods between time-steps and soil depth intervals. Meanwhile, we found clear increases in temporal stability with increasing soil depth, consistent with decreasing patchiness with increasing soil depth. Other similar studies that assessed the temporal stability of SWS around the world complements our findings (Martínez-Fernández 2003; Gao and Shao 2012; He et al. 2019). The possible explanations for these increasing temporal stability with increasing depth are: (1) the reduced dependence on root water uptake by the abundant elephant-grass and, other evapotranspiration and water processes that affected the variability of SWS; thus, SWS remained stable over time (Zhao et al. 2010), (2) the pronounced clay content in the bottom soil layers, which improves the ability of soils to keep moisture for long periods (Baskan et al. 2013; Huang et al. 2016) and (3) the plinthic and petro-plinthic layer, which promotes stagnation of SM in the bottom soil layers (Asiamah and Dedzoe 1999). In addition, time-stable locations (particularly, in the toe slope soils), as documented in Study 2 to be consistently wet due to their pedological formation, might have also introduce more stability for the bottom soil layers. Thus, SWS tended to be more stable with time (Figure 7.8). Some studies suggested that landscapes with less undulating topography exhibited good characteristics of temporal stability due to low or no effect of topography, which is similar to the setting of the present study (Zhao et al. 2017). A number of studies have debated whether temporal stability is greater in wet or dry periods (Hupet and Vanclooster 2002; Martínez-Fernández 2003), in the present study, we found a clear relation with increasing wetness.

All depth intervals across the study area (except 40–60 cm) spatially showed consistent dry distribution until April, whereas, other time-steps ranged from slightly wet to wet. However, due to the easy influence of the intertwin environmental factors at ≤ 30 cm depth, they rather showed a wet-dry-wet trend until the onset of heavy rains in June. This wet-dry-wet phenomenon suggest that small amounts of precipitation is adequate to recharge the SWS of the surface soil layers. Within the 0–15 cm, there were no marked differences in temporal stability of SWS. From Figure 7.8, we can imply that neither a time-stable dry nor a time-stable wet location can represent the temporal stability of SWS for the various standard depths, rather an adequate extent of locations of an area. In essence, a cluster of extended consistent time-stable locations makes information on SWS applicable and useful for food crop production.

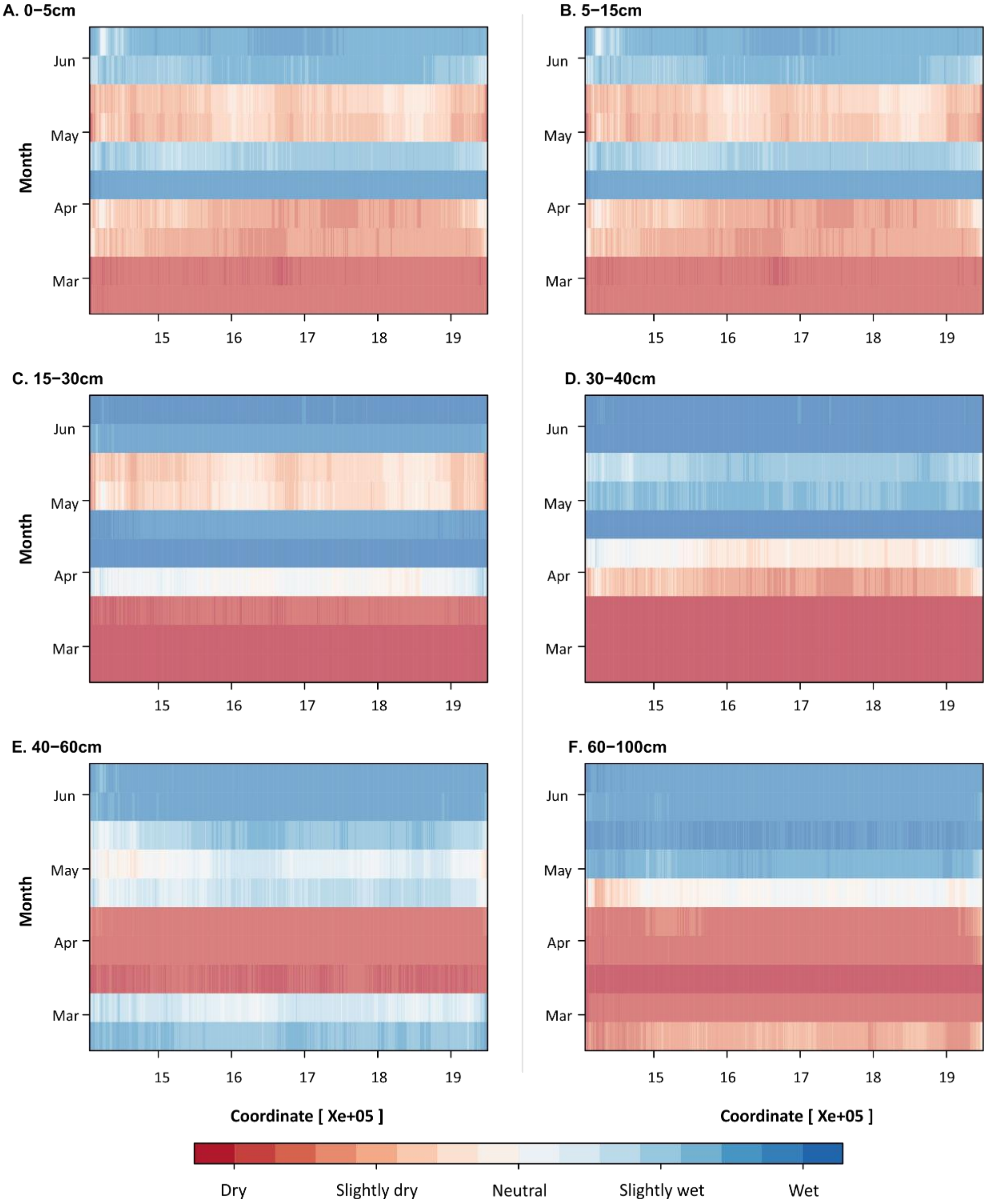


Figure 7.8. Spatio-temporal stability of SWS at each gridded location (i.e. pixel values) for each standard depth intervals from February 20th to June 8, 2018. 'X' values are in a measure based on cartesian coordinates.

7.4 Implication of four-dimensional SWS information for food-crop production

For example, in the study area, a single major cropping window is applicable, similar to other water deficit landscapes. One of the main challenges to implementing a dual cropping window in the Guinea savannah zone is water availability. Consequently, the livelihood options of subsistence farmers are limited, contributing to food insecurity. However, cultivating some drought tolerant crops in the lean season to complement the major cropping season may be an essential step to boost food-crop productivity, and increase rural livelihood options. Thus far, there are a number of drought tolerant crop varieties available for adoption in SSA (Fisher et al. 2015; AGRA 2017), but the SWS information needed to support and guide these recommendations are missing (Chivenge et al. 2015). The outcomes of the present study then form a core support system that is necessary to guide the implementation of such drought-adaptation measures, which also extends the findings of Study 2. For instance, from our results, it is now possible to explicitly address concerns such as (1) where (spatial location), (2) when (temporal), (3) at what sowing depths, (4) quantity (requirements) and (5) stability, with regards to SWS that are critical for the cultivation of notable crops grown in these water deficit landscapes. Also, the gridded SWS information can be useful for optimizing soil water management schemes and in the designing of soil water modelling systems.

One other adaption measure that was recently introduced in the study area to boost crop production limitations, due to the single cropping window, was the “Planting for Food and Jobs Program” by the Government of Ghana. However, such an enormous step is still limited as the program can hardly be extended beyond the rainy season. In this study, we found time-stable locations that can be used to grow, for example, short maturity crops or crops with deep rooting system. Our results therefore become a bridge that can extend some of these governmental programs by identifying time-stable locations where crop water requirements can be met in order for plants to survive throughout their growing period. Thus, either prolong the farming season or cultivate other notable crops that can tap these SWS reserves at the different soil depths. Furthermore, one key implication of this study is that on-going DSM initiatives across SSA and the world at large, can be complemented with our modelling framework in order to present an enhanced DSM package for improved agricultural productivity. Key explanation for this observation is that our modeling framework follows the *scorpan* concept, hence, covariates used here are also reported for existing DSM approaches, suggesting easy transferability of our approach into other jurisdictions. With regards to the optimal set of covariates used in the final predictions, *in situ* measurements and the RF-RFE guided by expert knowledge will always drive the optimal subset of the intended study area.

7.5 Conclusion

In this study, we produced the first spatio-temporally explicit modelling framework for mapping the 4-dimensional distribution of SWS. We found similar accuracy performances of the respective machine-learning algorithms and their ensembled model. Even with high spatio-temporal variability of the soil-landscape, the prediction accuracies of our SWS model outputs were high in all algorithms, and thus, they may be useful for similar modelling approaches. Future modelling endeavors may benefit from an increased intensity of network of *in situ* SM measurements at shorter time lags and soil thicknesses that also covers an entire year, as compared to those of this study. This will reduce patchiness of temporal stability and inconsistencies that were reported with the increasing soil depth as in this study. Also, an increased *in situ* SM measurements may increase the sensitivity of the multi-task modelling framework to slight changes in SWS variability, as SM is the main determining factor across the whole soil profile. Beyond these modelling findings, our study also found that the shallow soil layers showed high variability and was consistently dry, whereas the bottom soil layers showed time-stable wet cluster of locations. Also, although the temporal stability of SWS for the various standard depths is of importance for plant growth, our results showed that neither a time-stable dry nor a time-stable wet location can represent the temporal stability of SWS for the various standard depths.

The findings of this study demonstrate the potentials of key benchmark soils of the Guinea savannah zone, which can support crop productivity in parts of the dry season due to their soil water adequacy. The output maps from this study may be useful to inform the placement of suitable crops on the landscape that would make use of water storage in locations that can support its growth. Similarly, areas that may have serious drawdown of water resources (i.e. limited SWS) that could affect the main growing season crop may also be avoided with the help of our output maps.

Data on seasonal soil moisture and physical soil properties from the breadbasket of Ghana

Photo: Author

Chapter 8: Data on seasonal SM and physical soil properties from the "breadbasket" of Ghana

This manuscript is under review as:
Nketia KA, Sauer D. Data on seasonal soil moisture and physical soil properties from the breadbasket of Ghana.

Abstract

Soil moisture (SM) information that depicts its spatio-temporal variability, as influenced by physical soil properties, is one of the critical state variables necessary to improve food-crop production in water deficit landscapes. In essence, the survival of every crop is contingent on the available SM required to support it growing stages. However, such key soil parameter is missing in the Guinea savannah zone of Ghana, where the challenges of water deficit landscapes limit crop production potentials and impair rural livelihood options. This data article provides: (1) spatio-temporal *in situ* SM (n = 1,920), (2) gravimetric SM (n = 170) and (3) physical soil properties (i.e. clay, sand and silt contents, and bulk density; n = 170 each) from 36 validation locations across the Guinea savannah zone. Locations from where these datasets were generated were stratified via a coupled global weighted principal component analysis and a cost-constrained conditioned Latin hypercube algorithm. These locations also span prominent land use and key agricultural benchmark soils of the breadbasket of Ghana. Thus, such a comprehensive dataset can support future studies in improving food-crop productivity in the Guinea savannah zone of Ghana. It is worth noting that the settings of the Guinea savannah zone reflect the agro-ecology of major parts of West Africa.

Keywords: dry seasonal, soil moisture, spatio-temporal, physical soil properties, PR2/60 profile probe, Guinea savannah, Ghana

8.1 Specifications table

Subject	Agricultural and Biological Sciences.
Specific subject area	Soil Science for sustainable agriculture and food-crop production.
Type of data	Tables and excel file.
How data were acquired	Through stratified soil sampling from field survey using Delta-T PR2/60 SM probe and other soil sampling tools such as stainless-steel cylinder, soil auger and other soil digging tools. Others were obtained by standard laboratory analysis.
Data format	Raw.
Parameters for data collection	In situ soil conditions and soil properties.
Description of data collection	Data was collected by physically inserting SM probes into installed access tubes (up to 100 cm depth) and recording their respective SM for the various soil depths. In addition, soil samples were collected in close proximity to the installed access tubes for the various soil depths and analyzed in a laboratory.
Data source location	36 Geographical Position System (GPS) coordinates of validation locations in the Guinea savannah zone of Ghana are presented in appendix Table S8.1. Data to this article is also accessible on an open-access repository (DOI: 10.5281/zenodo.3662411).
Data accessibility	Data is with this article.
Related research article	1. Nketia KA, Asabere SB, Erasmi S, Sauer D. A new method for selecting sites for soil sampling, coupling global weighted principal component analysis and a cost-constrained conditioned Latin hypercube algorithm. MethodsX 2019; 6:284–99. DOI: 10.1016/j.mex.2019.02.005. 2. Nketia KA, Becker JN, Asabere SB, Erasmi S, Sauer D. Spatial and temporal dynamics of soil moisture in benchmark soils of the Guinea savannah zone of Ghana - is there an unused potential for food-crop production? Under review.

8.2 Value of the data

- The datasets provide detailed spatio-temporal SM (0–100 cm depth) of key agricultural benchmark soils of the Guinea savannah zone of Ghana.
- Thus far, no data exist on the spatio-temporal SM for an entire soil profile for the breadbasket of Ghana.

- It can be used for *in situ* water management, quantitative analysis to guide agricultural resource management in such water deficit landscapes, particularly in the field of soil science.
- The results of this article are useful and will directly benefits students and other agriculturist/researchers for future studies of SM variability throughout the dry-cropping season necessary for improved food-crop productivity.

8.3 Data description

The datasets described here: (1) spatio-temporal *in situ* SM measurements (n = 1,920), (2) gravimetric SM (n = 170) and (3) physical soil properties (i.e. clay, sand and silt contents, and bulk density; n = 170 each) are from the 36 SM validation locations (appendix Table S8.1) of the Guinea savannah zone (Figure 2.1). In addition, we present data on thresholds of SM demands by maize crops (*Zea mays*) for the dry season of the study area. The data covered a spatial extent of 150 x 80 km traversing various land uses such as grassland (*Pennisetum purpureum*), food-crops (i.e. maize, cowpea, sorghum, cassava and yam) and other tree species (i.e. Borassus palm – *Borassus aethiopum*, shea trees – *Vitellaria paradoxa* and Senegal mahogany – *Khaya senegalensis*). In addition, the data covers the variation of the peak of the dry season (in February) until the onset of heavy rains (in June). The morphological characteristics of the soils from which these datasets were collected are detailed in Adu (1999). We present the comprehensive findings as tables in the thesis' project online repository (DOI: 10.5281/zenodo.3662411). It should be stressed that all datasets reported in this paper are raw values. With regards to the *in situ* SM measurements, they are presented in a spatio-temporal domain, whereas, in one-time, the gravimetric SM and (3) physical soil properties are extracted from collected soil samples for each validation location in the laboratory.

8.4 Data source

Raw datasets are available on an open-access repository (DOI: 10.5281/zenodo.3662411).

Data on the spatio-temporal root zone soil water storage – the Guinea savannah zone of Ghana

Photo: Author

Chapter 9: Data on the spatio-temporal RZ-SWS in the Guinea savannah zone of Ghana

This manuscript is under review as:
Nketia KA, Sauer D. Data on the spatio-temporal root zone soil water storage – the Guinea savannah zone of Ghana.

Abstract

In between rainfall events or irrigations, soil water storage (SWS) of the root zone reflects the size of soil horizon's water reservoir and its storage adequacy. After field capacity is achieved, crops draw upon this SWS for its water requirements. SWS illustrates the capabilities of these soil horizons to buffer crops at specific growth stages against transient water deficits. Consequently, information on the spatio-temporal SWS is critical for the sustainability of food-crop production in rainfed agriculture, especially in semi-arid landscapes. In sub-Saharan Africa, where many national economies are driven by agriculture, such critical soil information is missing. In this data article, we provide data on the spatio-temporal SWS (n = 1,920) from 36 measurement locations across the Guinea savannah zone of Ghana. Data covers the 0–100 cm rooting depth and spans agricultural benchmark soils of the Guinea savannah zone. In addition, the data covers the various topographical units and prominent land use of the area, similar to most parts of West Africa. We further present algorithms used in deriving the SWS data presented in this article. The comprehensive dataset and algorithm can support crop production with regards to water management and optimal agricultural resource allocation in the Guinea savannah zone of Ghana and other parts of sub-Saharan African where water deficit conditions are common.

Keywords: root zone, semi-arid, soil water storage, spatio-temporal, Guinea savannah, Ghana, water deficit, West Africa

9.1 Specifications table

Subject	Agricultural and Biological Sciences.
Specific subject area	Soil Science for sustainable agriculture, food security and soil water management.
Type of data	Tables and 'R' data files.
How data were acquired	Through Delta-T PR2/60 *in situ* SM measurements.
Data format	Analyzed.
Description of data collection	Data was derived using a SWS estimation algorithm – it recursively profiled *in situ* SM as a function of bulk density and soil thickness, which accounted for the effect of measured soil depth (i.e. 0–100 cm) on the *in situ* SM measurements.
Data source location	Geographical Position System (GPS) coordinates of the 36 measurement locations are presented in appendix Table S8.1. Data to this article is also accessible on an open-access repository (DOI: 10.5281/zenodo.3662411).
Data accessibility	Data is with this article as supplementary material.
Related research article	1. Nketia KA, Asabere SB, Erasmi S, Sauer D. A new method for selecting sites for soil sampling, coupling global weighted principal component analysis and a cost-constrained conditioned Latin hypercube algorithm. MethodsX 2019; 6:284–99. DOI: 10.1016/j.mex.2019.02.005. 2. Nketia KA, Becker JN, Asabere SB, Erasmi S, Sauer D. Spatial and temporal dynamics of soil moisture in benchmark soils of the Guinea savannah zone of Ghana - is there an unused potential for food-crop production? Under review. 3. Nketia KA, Ramcharan A, Asabere SB, Herbold S, Erasmi S, Sauer D. Spatio-temporal mapping of SWS in a semi-arid landscape of Northern Ghana. Under review.

9.2 Value of the data

- Comprehensively, the data provides 0–100 cm soil depth information on spatio-temporal SWS for agricultural benchmark soils of the Guinea savannah zone of Ghana, similar other parts of the West African Guinea savannah agro-ecological zone.
- No data exist on the spatio-temporal SWS for 0–100 cm depth for soils of the Guinea savannah zone of Ghana.
- The data is critical for crop water resource management and quantitative analysis to guide food production under rainfed conditions during transient water deficit periods.

- It can be used to map the four-dimensional distribution of SWS, which can complement existing digital soil mapping initiatives of sub-Saharan Africa and Ghana.
- Further usefulness of the data, in the field of soil science, includes future food productivity research, modelling of water management for dry-season farming and guide policy making processes.

9.3 Data description

This data illustrates and describes the spatio-temporal SWS (n = 1,920) of 36 stratified locations of the Guinea savannah zone of Ghana (see appendix Table S8.1; Figure 2.1). The data covers an area of 150 x 80 km across key agricultural soil associations of the area. The Guinea savannah zone, from where the data was collected, reflects the settings of major part of West Africa (Benneh et al. 1996). Datasets covers land uses such as Borassus palm (*Borassus aethiopum)*, shea trees (*Vitellaria paradoxa),* Senegal mahogany (*Khaya senegalensis*) and natural grassland (*Pennisetum purpureum*). In addition, the data covers notable food-crops production systems of the Guinea savannah zone, i.e. maize, cowpea, sorghum, cassava, millet and yam.

This data further shows the variation of SWS from February (peak of the dry season) until June (onset of heavy rains) on key benchmark soils of the Guinea savannah zone. Adu (1999) reported details on the morphological properties of soils from which the SWS information was derived. Tables are used to show the comprehensive datasets on SWS in this article and can be found on the thesis' project online repository. We further illustrate the variability of SWS per each benchmark soil type along the various *in situ* measurement depths (see appendix Figure S9.1).

9.4 Data source

Raw datasets on the spatio-temporal root zone SWS and its associated set of algorithms used in its development are publicly assessible at DOI: 10.5281/zenodo.3662411.

Photo: Author

Section C: General conclusions, limitations and outlook

Chapter 10: Synthesis and implications

10.1 General conclusions

In this thesis, we investigated and modelled the spatio-temporal seasonal SM and SWS of the root zone of arable benchmark soils of the Guinea savannah zone of Ghana. In the context of the two specific objectives in pursuance of this main goal, our findings modernize and renew approaches on spatio-temporal scale analyses for understanding root zone SM and SWS in semi-arid landscapes. General implications of our findings are discussed.

In the newly designed sampling stratification (i.e. Chapter 4), we found that the ensemble-approach rather presented an alternative to represent both the feature and the geographic space in a soil sampling stratification. For this finding, a spatial extent of a study area can adequately be stratified in order to effectively and efficiently capture its localized spatio-temporal heterogeneities that influence the variability of the investigated soil property (in our case SM). Consequently, limitations documented in the literatures for soil sampling designs with regards to local structures and localized-spatial autocorrelation that explained the soil-landscape variability of the investigated soil property in a constrained environment have been addressed. Our proposed method effectively correlated the investigated soil property and its local environment, which depicts the geographic variations in the soil-environmental characteristics across space. Thus far, existing soil sampling designs were also challenged to address this limitation. In addition, our proposed sampling design further suppressed the occurrence of model-selected sampling locations in areas that were less representative of the soil property of interest, suggesting resource use efficiency. We therefore conclude that such an ensemble-approach for soil sampling presents a viable alternative to guide field data collection and installations of soil-monitoring stations.

By investigating the spatio-temporal variability of root zone SM and how it was influenced by internal- and external-soil factors (i.e. Chapter 5), we found that SM of the agricultural benchmark soils of the Guinea savannah was low and highly variable only in the top soil layers (≤ 10 cm), whereas in the bottom soil layers (≥ 20 cm), increased clay content with increasing soil depth kept high SM for longer periods. Internal factors such as bulk density, clay and silt content strongly influenced SM than external factors (i.e. slope, topographic wetness index, antecedent precipitation index and evapotranspiration). One key outcome of Chapter 5 is that almost all benchmark soils of the Guinea savannah zone (except for the Kumayili series) below

10 cm depth have SWS potentials that match the water requirements of at least some drought-tolerant crops grown in the area (i.e. cowpea – Bengpla cultivar; > 9 mm day^{-1}). Hence, the possibility of a dual farming system is not far-fetched. These findings are important to highlight because a dual farming system can enhance agricultural productivity, improve food security and increase the income of local farming communities. Our proposed legume-crop rotation brings an advantage over existing practices in that: (1) it can maintain soil fertility for the main cropping season, (2) the use of cowpea grain as food and nutritional supplements for the local communities and (3) residue from crop harvest can be used as mulch to conserve SM or be used as feed for livestock.

While we also sort to translate point-based measurements into spatial-temporal scales via remote sensing techniques (i.e. Chapter 6), we found that Sentinel-2 derived variables were better at discriminating low to extremely-low SM as compared to Sentinel-1 derived variables alone. Thus, they accounted for specific clusters of *in situ* SM whereas, Sentinel-1 derived variables focused on accounting for the entire range of *in situ* SM and microwave signal effects due to incidence angle. Also, the inclusion of DEM-derived variables better disentangled SM estimation ambiguities in order to improve SM estimation. Overall, our ensemble-based model, contrary to existing approaches, rather presented an enhanced alternative to estimate SM from synergetic use of remote sensing derived variables. The ensemble-based model, at certain SM regimes and soil conditions, is an effective tool to estimate SM from ESA's Sentinel-1 and -2 missions. Hence, new perspectives are opened to improve crop simulation models with detailed plot level spatio-temporal SM inputs.

Extending the applicability of findings from Chapter 5 and 6 (i.e. in Chapter 7), we found key implications that can improve food-crop production. (1) Modelled root zone SWS in the shallow soil layers were highly variable, unstable and consistently dry (high temporal instability) compared to the bottom soil layers, which showed time-stable wet cluster of locations due to increase SM and clay buildup that promoted stability. (2) Spatio-temporally, time-stable locations where crop water requirements can be met during crop growing periods can be explicitly identified for use. Our outcomes form a core support system necessary to guide the implementation of drought-adaptation measures and complement existing digital soil mapping initiatives around the world. Possible adoptable recommendations drawn from the four-dimensional SWS maps include, the possibility and practicability to either prolong the existing major single farming window or identify locations and durations where additional crop-specific farming is applicable in these semi-arid landscapes.

10.2 Limitations

This research is the first to investigate and demonstrate the applicability of spatio-temporal scale analyses for root zone SM and SWS in the Guinea savannah zone of Ghana and SSA at large. Some limitations were identified and future studies should be undertaken to address them.

- In Chapter 5, there is the need to undertake an on-farm validation and plausibility check of our proposed legume-crop rotation recommendation. This will ensure adaptability, workability and compatibility with existing farming practices that promote food security and livelihood empowerment.
- In Chapter 6, one novel aspect of the research task was the determination of the minimum SM thresholds at which the widely used machine-learning algorithms under-performed. However, we did not cover higher SM ranges in soils of the study area at the time of *in situ* measurements. Hence, further studies with extremely higher SM values is required to investigate the maximum SM thresholds at which these algorithms are also inefficient. This will ensure a full model performance recommendation.
- In Chapter 7, due to the sparse *in situ* measurement networks, neither a time-stable dry nor a time-stable wet location can adequately represent the temporal stability of SM for the various measurement soil depths. Consequently, a yearly expansive *in situ* measurement network at a shorter time lag and shorter soil thickness is important. This will reduce patchiness of temporal stability and inconsistencies reported with increasing soil depth with regards to SM and SWS. Furthermore, our modelling frameworks used in Chapter 4–7 can be very sensitive to slight changes in SM and SWS variability across the whole soil profile.

10.3 Outlook

Given the importance of root zone SM and SWS in water deficit landscapes, such benchmark studies, as demonstrated in this thesis, should not be overlooked in future crop production adaptation initiatives across SSA and the world as a whole. The spatio-temporally modelled SM and SWS present useful information for assessing the fluxes of SM and SWS necessary to guide and optimize rational use of agricultural resource in water-deficit landscapes. In addition, such valuable information can enhance the adaptive capacity of smallholder farmers to address site-specific concerns such as where, when, what quantity and stability of SM and SWS. That notwithstanding, plant-water management schemes can be designed with such SM and SWS flux information as benchmarks for crop suitability recommendation of an area.

The spatio-temporal scale analyses demonstrated in this thesis bring an advantage over point-based analysis and the practicability of understanding seasonal root zone SM and SWS in water deficit landscapes in the future. One critical outcome of this thesis is that our state-of-the-art modeling framework follows the *scorpan* concept for digital soil mapping. Hence, our approaches can easily be transferred to other geographical settings. However, the principles of the *scorpan* concept must be adhered in its usage. The automated modelling framework is open-sourced and can easily be updated with new *in situ* SM measurements. For these reasons and as part of open and reproducible science, all modelling framework and full datasets reported in this thesis are publicly available at DOI: 10.5281/zenodo.3662411.

References

Abass, M. (2015). Dry season farming in Northern and Upper East Dry season farming in Northern and Upper East Regions of Ghana: A potential to improve income and nutrition of smallholder farmers 5 (3), pp. 72–76.

Acheampong, E. N.; Ozor, N.; Owusu, E. S. (2014). Vulnerability assessment of Northern Ghana to climate variability. In Climatic Change 126 (1), pp. 31–44. DOI: 10.1007/s10584.014.1195.z.

Adeboye, O. B.; Schultz, B.; Adekalu, K. O.; Prasad, K. (2017). Soil water storage, yield, water productivity and transpiration efficiency of soybeans (Glyxine max L.Merr) as affected by soil surface management in Ile-Ife, Nigeria. In International Soil and Water Conservation Research 5 (2), pp. 141–150. DOI: 10.1016/j.iswcr.2017.04.006.

Adjei-Nsiah, S. (2008). Cropping system, land tenure and social diversity in Wenchi, Ghana: implications for soil fertility management. In: Bationo, A. Tubo, R., Wasiwa, B. Okoyo, J. and Fosil, M. Synthesis of soil water and nutrient management research in the Volta Basin.

Adjei-Nsiah, S.; Kuyper, T. W.; Leeuwis, C.; Abekoe, M. K.; Giller, K. E. (2007). Evaluating sustainable and profitable cropping sequences with cassava and four legume crops. Effects on soil fertility and maize yields in the forest/savannah transitional agro-ecological zone of Ghana. In Field crops research 103 (2), pp. 87–97. DOI: 10.1016/j.fcr.2007.05.001.

Adolwa, I. S.; Schwarze, S.; Bellwood-Howard, I.; Schareika, N.; Buerkert, A. (2017). A comparative analysis of agricultural knowledge and innovation systems in Kenya and Ghana. Sustainable agricultural intensification in the rural-urban interface. In Agriculture and Human Values 34 (2), pp. 453–472. DOI: 10.1007/s10460-016-9725-0.

Adu, S. V. (Ed.) (1995). Soils of the Nasia River Basin, Northern Region, Ghana. CSIR-Soil Research Institute. Memoir No. 11. Kwadaso, Kumasi.

Adu, S. V.; Asiamah, R. D. (Eds.) (2003). Soils of the Yapei – Sawla Road Area, Northern Region, Ghana. CSIR-Soil Research Institute. Memoir No. 16. Kwadaso, Kumasi.

AGRA (2017). Africa agriculture status report: The business of smallholder agriculture in sub-Saharan Africa. Alliance for a Green Revolution in Africa (AGRA). Nairobi, Kenya (5). Available online at https://agra.org/wp-content/uploads/2017/09/Final-AASR-2017-Aug-28.pdf, checked on 10/22/2018.

Agrawal, K. M.; Mehra, R.; Ryali, U. S. (2016). NISAR ISRO science data processing and products. In: Earth Observing Missions and Sensors: Development, Implementation, and Characterization IV, vol. 9881. International Society for Optics and Photonics, p. 988104.

Ahmad, S.; Kalra, A.; Stephen, H. (2010). Estimating soil moisture using remote sensing data. A machine learning approach. In Advances in Water Resources 33 (1), pp. 69–80. DOI: 10.1016/j.advwatres.2009.10.008.

Akaike, H. (2011). Akaike's information criterion. In: International encyclopedia of statistical science: Springer, p. 25.

Alexakis, D. D.; Mexis, F. K.; Vozinaki, A. K.; Daliakopoulos, I. N.; Tsanis, I. K. (2017). Soil moisture content estimation based on Sentinel-1 and auxiliary earth observation products. A Hydrological Approach. In Sensors 17 (6). DOI: 10.3390/s17061455.

Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. (2015). Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. In Remote Sensing 7 (12), pp. 16398–16421. DOI: 10.3390/rs71215841.

Alidu, A. F.; Tanko, M.; Iddrisu, A. (2016). Factors enhancing smallholders farmers` access to masara n'arziki project in the northern region of Ghana. 6 (10), pp. 9609–9615.

Allen, R. G.; Pereira, L. S.; Raes, D.; Smith, M.; et al. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. In Fao, Rome 300 (9), D05109.

Amazirh, A.; Merlin, O.; Er-Raki, S.; Gao, Q.; Rivalland, V.; Malbeteau, Y.; et al. (2018). Retrieving surface soil moisture at high spatio-temporal resolution from a synergy between Sentinel-1 radar and Landsat thermal data. A study case over bare soil. In Remote Sensing of Environment 211, pp. 321–337. DOI: 10.1016/j.rse.2018.04.013.

Amekudzi, L.; Yamba, E.; Preko, K.; Asare, O. E.; Aryee, J.; Baidu, M.; Codjoe, S. (2015). Variabilities in rainfall onset, cessation and length of rainy season for the various agro-ecological zones of Ghana 2015, pp. 416–434.

Antonino, A. C. D.; Sampaio, E. V. S. B.; Dall'Olio, A.; Salcedo, I. H. (2000). Balanço hidrico em solo com cultivos de subsistência no semi-árido do nordeste do Brasil. In Revista Brasileira de Engenharia Agr\'\icola e Ambiental 4 (1), pp. 29–34.

Antwi-Agyei, P.; Dougill, A. J.; Stringer, L. C.; Codjoe, S. N. A. (2018). Adaptation opportunities and maladaptive outcomes in climate vulnerability hotspots of northern Ghana. In Climate Risk Management 19, pp. 83–93. DOI: 10.1016/j.crm.2017.11.003.

Armah, F. A.; Odoi, J. O.; Yengoh, G. T.; Obiri, S.; Yawson, D. O.; Afrifa, E. K. A. (2011). Food security and climate change in drought-sensitive savannah zones of Ghana. In Mitigation and Adaptation Strategies for Global Change 16 (3), pp. 291–306. DOI: 10.1007/s11027-010-9263-9.

Arrouays, D.; McKenzie, N.; Hempel, J.; Richer de Forges, A.; McBratney, A. (Eds.) (2014). GlobalSoilMap. Basis of the global spatial soil information system. 1st Edition. London: CRC Press.

Asafo-Adjei, B.; Singh, B. B. (2005). Registration of 'Bengpla' cowpea. In Crop science 45 (6), pp. 2649–2651.

Asiamah, R. D.; Dedzoe, C. D. (1999). Plinthization-A threat to agricultural production. In Ghana Journal of Agricultural Science 32 (2), pp. 223–227.

Asuming-Brempong, S. (2013). Policy initiatives and agricultural performance in post-independent Ghana. In Journal of Social and Development Sciences 4 (9), pp. 425–434.

Attarzadeh, R.; Amini, J.; Notarnicola, C.; Greifeneder, F. (2018). Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at Plot Scale. In Remote Sensing 10 (8). DOI: 10.3390/rs10081285.

Bado, V. B.; Bationo, A. (2018). Chapter One - Integrated management of soil fertility and land resources in sub-Saharan Africa. Involving local communities. In Donald L. Sparks (Ed.), vol. 150: Academic Press (Advances in Agronomy), pp. 1–33.

Ballerine, C. (2017). Topographic wetness index urban flooding awareness act action support, Will & DuPage Counties, Illinois. Illinois State Water Survey.

Banson, K. E.; Nguyen, N. C.; Bosch, O. J. H. (2016). Using system archetypes to identify drivers and barriers for sustainable agriculture in Africa. A case study in Ghana. In Systems Research and Behavioral Science 33 (1), pp. 79–99.

Barrett, B. W.; Dwyer, E.; Whelan, P. (2009). Soil moisture retrieval from active spaceborne microwave observations. An evaluation of current techniques. In Remote Sensing 1 (3), pp. 210–242. DOI: 10.3390/rs1030210.

Barton, Kamil (2018). Package 'MuMIn'.

Baskan, O.; Kosker, Y.; Erpul, G. (2013). Spatial and temporal variation of moisture content in the soil profiles of two different agricultural fields of semi-arid region. In Environmental Monitoring and Assessment 185 (12), pp. 10441–10458.

Bates, D.; Mächler, M.; Bolker, B.; Walker, S. (2014). Fitting linear mixed-effects models using lme4. In arXiv preprint arXiv:1406.5823.

Belane, A. K.; Dakora, F. D. (2009). Measurement of N2 fixation in 30 cowpea (Vigna unguiculata L. Walp.) genotypes under field conditions in Ghana, using the15N natural abundance technique. In Symbiosis 48 (1), pp. 47–56. DOI: 10.1007/BF03179984.

Benneh, G.; Morgan, W. B.; Uitto, J. I. (1996). Sustaining the future: economic, social and environmental change in Sub-Saharan Africa: Tokyo, JP: United Nations University Press (338.967 S964s).

Bisgin, H.; Bera, T.; Ding, H.; Semey, H. G.; Wu, L.; Liu, Z.; et al. (2018). Comparing SVM and ANN based machine learning methods for species identification of food contaminating beetles. In Scientific Reports 8 (1), p. 6532. DOI: 10.1038/s41598-018-24926-7.

Biswas, A.; Si, B. C. (2011). Revealing the controls of soil water storage at different scales in a hummocky landscape. In Soil Science Society of America Journal 75 (4), pp. 1295–1306.

Biswas, A.; Zhang, Y. (2018). Sampling designs for validating digital soil maps. A review. In Pedosphere 28 (1), pp. 1–15. DOI: 10.1016/S1002-0160(18)60001-3.

Black, G. R.; Hartge, K. H. (1986). Bulk density. In methods of soil structure and migration of colloidal materials soils. In Soil Sci. Soc. Am. J 26, pp. 297–300.

Bohlool, B. B.; Ladha, J. K.; Garrity, D. P.; George, T. (1992). Biological nitrogen fixation for sustainable agriculture. A perspective. In Plant and Soil 141 (1), pp. 1–11. DOI: 10.1007/BF00011307.

Bourne Jr, Joel K. (2014). The next breadbasket. In National Geographic 226 (1), pp. 46–77.

Bouyoucos, G. J. (1962). Hydrometer method improved for making particle size analyses of soils. In Agronomy Journal 54, pp. 464–465. DOI: 10.2134/agronj1962.00021962005400050028x.

Breiman, L. (2001). Random Forests. In Machine Learning 45 (1), pp. 5–32. DOI: 10.1023/A:1010933404324.

Brenning, A.; Bangs, D.; Becker, M.; Schratz, P.; Polakowski, F. (2018). Package 'RSAGA'.

Brink, A. B.; Eva, H. D. (2009). Monitoring 25 years of land cover change dynamics in Africa. A sample based remote sensing approach. In Applied Geography 29 (4), pp. 501–512. DOI: 10.1016/j.apgeog.2008.10.004.

Brocca, L.; Ciabatta, L.; Massari, C.; Camici, S.; Tarpanelli, A. (2017). Soil moisture for hydrological applications. Open questions and new opportunities. In Water 9 (2), p. 140. DOI: 10.3390/w9020140.

Buakum, B.; Limpinuntana, V.; Vorasoot, Nimitr; Pannangpetch, Krirk; Bell, Richard (2018). Effects of Sowing Depths on Growth and Rooting Patterns of Cowpea and Mungbean in a Deep Sand with Shallow Water Table.

Burk, L.; Dalgliesh, N. P. (2013). Estimating plant available water capacity. In GRDC: Canberra Available at https://www. grdc. com. au/ /media/3AA0BA9C5D6E4EED8BF0C3313B0AD498. pdf [Verified 20 November 2015].

Cafarelli, B.; Castrignanò, A.; Benedetto, D.; Palumbo, A. D.; Buttafuoco, G. (2015). A linear mixed effect (LME) model for soil water content estimation based on geophysical sensing. A comparison of an LME model and kriging with external drift. In Environmental Earth Sciences 73 (5), pp. 1951–1960. DOI: 10.1007/s12665-014-3543-8.

Chamberlin, J. (2008). It's small world after all. Defining smallholder agriculture in Ghana: Intl Food Policy Res Inst (823).

Charlton, M.; Brunsdon, C.; Demsar, U.; Harris, P.; Fotheringham, S. (2010). Principal Components Analysis: from Global to Local. In: 13th AGILE International Conference on Geographic Information Science 2010, 1st Oct. 2010, Portugal.

Chen, T.; Guestrin, C. (2016). XGBoost. A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: ACM (KDD '16), pp. 785–794.

Chivenge, P.; Mabhaudhi, T.; Modi, A. T.; Mafongoya, P. (2015). The Potential Role of Neglected and Underutilised Crop Species as Future Crops under Water Scarce Conditions in sub-Saharan Africa. In International Journal of Environmental Research and Public Health 12 (6), pp. 5685–5711. DOI: 10.3390/ijerph120605685.

Choker, M.; Baghdadi, N.; Zribi, M.; El Hajj, M.; Paloscia, S.; Verhoest, N.; et al. (2017). Evaluation of the Oh, Dubois and IEM backscatter models using a large dataset of SAR data and experimental soil measurements. In Water 9 (1), p. 38. DOI: 10.3390/w9010038.

Colliander, A.; Jackson, T. J.; Bindlish, R.; Chan, S.; Das, N.; Kim, S. B.; et al. (2017). Validation of SMAP surface soil moisture products with core validation sites. In Remote Sensing of Environment 191, pp. 215–231. DOI: 10.1016/j.rse.2017.01.021.

Comber, A. J.; Harris, P.; Tsutsumida, N. (2016). Improving land cover classification using input variables derived from a geographically weighted principal components analysis. In ISPRS Journal of Photogrammetry and Remote Sensing 119, pp. 347–360. DOI: 10.1016/j.isprsjprs.2016.06.014.

Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; et al. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. In Geoscientific Model Development 8 (7), pp. 1991–2007. DOI: 10.5194/gmd-8-1991-2015.

Cooper, J. David (2016). Soil water measurement. A practical handbook: John Wiley & Sons.

Critchley, W.; Siegert, K.; Chapman, C.; Finkel, M. (1991). Water harvesting. A manual for the design and construction of water harvesting schemes for plant production. Scientific Publishers.

Dai, A.; Trenberth, K. E.; Qian, T. (2004). A global dataset of Palmer Drought Severity Index for 1870-2002. Relationship with soil moisture and effects of surface warming. In Journal of Hydrometeorology 5 (6), pp. 1117–1130.

Daly, E.; Porporato, A. (2005). A review of soil moisture dynamics. from rainfall infiltration to ecosystem response. In Environmental Engineering Science 22 (1), pp. 9–24. DOI: 10.1089/ees.2005.22.9.

Dewitte, O.; Jones, A.; Spaargaren, O.; Breuning-Madsen, H.; Brossard, M.; Dampha, A.; et al. (2013). Harmonisation of the soil map of Africa at the continental scale. In Geoderma 211-212, pp. 138–153. DOI: 10.1016/j.geoderma.2013.07.007.

Dobson, M. C.; Ulaby, F. T.; Hallikainen, M. T.; El-rayes, M. A. (1985). Microwave dielectric behavior of wet soil-part II. Dielectric mixing models. In IEEE Transactions on Geoscience and Remote Sensing GE-23 (1), pp. 35–46. DOI: 10.1109/TGRS.1985.289498.

Duan, L.; Huang, M.; Zhang, L. (2016). Use of a state-space approach to predict soil water storage at the hillslope scale on the Loess Plateau, China. In Catena 137, pp. 563–571. DOI: 10.1016/j.catena.2015.11.003.

El Hajj, M.; Baghdadi, N; Zribi, M.; Bazzi, H. (2017). Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. In Remote Sensing 9 (12). DOI: 10.3390/rs9121292.

El-Sharkawy, M. A. (1993). Drought-tolerant cassava for Africa, Asia, and Latin America. In BioScience 43 (7), pp. 441–451. DOI: 10.2307/1311903.

Entekhabi, D.; Njoku, E. G.; O'Neill, P. E.; Kellogg, K. H.; Crow, W. T.; Edelstein, W. N.; et al. (2010). The soil moisture active passive (SMAP) mission. In Proceedings of the IEEE 98 (5), pp. 704–716.

Ezui, K. S.; Franke, A. C.; Mando, A.; Ahiabor, B. D. K.; Tetteh, F. M.; Sogbedji, J.; et al. (2016). Fertiliser requirements for balanced nutrition of cassava across eight locations in West Africa. In Field crops research 185, pp. 69–78. DOI: 10.1016/j.fcr.2015.10.005.

Fang, X.; Zhao, W.; Wang, L.; Feng, Q.; Ding, J.; Liu, Y.; Zhang, X. (2016). Variations of deep soil moisture under different vegetation types and influencing factors in a watershed of the Loess Plateau, China. In Hydrology and Earth System Sciences 20 (8), pp. 3309–3323. DOI: 10.5194/hess-20-3309-2016.

Fening, J. O. (2016). The status of Ghana's agricultural soils and potential for sustainable food production. Kumasi, CSIR-Soil Research Institute, Academy Post Office, Kumasi-Ghana.

Fening, J. O.; Adjei Gyapong, T.; Ampontuah, E. O.; Yeboah, E.; Gaisie, E. (2006). Fertilizing for profit. The case of cassava cultivation in Ghana. In Tropical Science 45 (2), pp. 97–99. DOI: 10.1002/ts.57.

Fisher, M.; Abate, T.; Lunduka, R. W.; Asnake, W.; Alemayehu, Y.; Madulu, R. B. (2015). Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa. Determinants of adoption in eastern and southern Africa. In Climatic Change 133 (2), pp. 283–299. DOI: 10.1007/s10584-015-1459-2.

Foley, J. A.; Ramankutty, N.; Brauman, K. A.; Cassidy, E. S.; Gerber, J. S.; Johnston, M.; et al. (2011). Solutions for a cultivated planet. In Nature 478 (7369), pp. 337–342. DOI: 10.1038/nature10452.

Forkuor, G.; Hounkpatin, O. K. L.; Welp, G.; Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso. A comparison of machine learning and multiple linear regression models. In Plos One 12 (1), pp. 1–21. DOI: 10.1371/journal.pone.0170478.

Franceschetti, G.; Lanari, R. (2018). Synthetic aperture radar processing: CRC Press.

Franzluebbers, A. J. (2002). Water infiltration and soil structure related to organic matter and its stratification with depth. In Soil and Tillage Research 66 (2), pp. 197–205. DOI: 10.1016/S0167-1987(02)00027-2.

Friedman, J.; Hastie, T.; Tibshirani, R.; Simon, N.; Narasimhan, B.; Qian, J. (2017). GLMNET. Lasso and Elastic-Net Regularized Generalized Linear Models. R package version 2.0-13.

Frisvold, G.; Sanchez, C.; Gollehon, N.; Megdal, S. B.; Brown, P. (2018). Evaluating gravity-flow irrigation with lessons from Yuma, Arizona, USA. In Sustainability 10 (5). DOI: 10.3390/su10051548.

Fung, A. K.; Li, Z.; Chen, K. S. (1992). Backscattering from a randomly rough dielectric surface. In IEEE Transactions on Geoscience and Remote Sensing 30 (2), pp. 356–369. DOI: 10.1109/36.134085.

Gao, L.; Shao, M. (2012). Temporal stability of soil water storage in diverse soil layers. In Catena 95, pp. 24–32. DOI: 10.1016/j.catena.2012.02.020.

Gao, Q.; Zribi, M.; Escorihuela, M. J.; Baghdadi, N. (2017). Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at 100 m resolution. In Sensors 17 (9). DOI: 10.3390/s17091966.

Geng, S. M.; Yan, D. H.; Zhang, T. X.; Weng, B. S.; Zhang, Z. B.; Qin, T. L. (2015). Effects of drought stress on agriculture soil. In Natural Hazards 75 (2), pp. 1997–2011. DOI: 10.1007/s11069-014-1409-8.

Ghana Ministry of Finance (2017). 2017 Mid-year budget review. Accra.

Godfray, H. C. J.; Beddington, J. R.; Crute, I. R.; Haddad, L.; Lawrence, D.; Muir, J. F.; et al. (2010). Food Security: The challenge of feeding 9 billion people. In Science 327 (5967), pp. 812–818. DOI: 10.1126/science.1185383.

Grabs, T.; Seibert, J.; Bishop, K.; Laudon, H. (2009). Modeling spatial patterns of saturated areas. A comparison of the topographic wetness index and a dynamic distributed model. In Journal of Hydrology 373 (1-2), pp. 15–23.

Graham, A. J.; Harris, R. (2003). Extracting biophysical parameters from remotely sensed radar data. A review of the water cloud model. In Progress in Physical Geography: Earth and Environment 27 (2), pp. 217–229. DOI: 10.1191/0309133303pp378ra.

Grayson, R. B.; Blöschl, G.; Western, A. W.; McMahon, T. A. (2002). Advances in the use of observed spatial patterns of catchment hydrological response. In Advances in Water Resources 25 (8), pp. 1313–1334. DOI: 10.1016/S0309-1708(02)00060-X.

Greenland, S. J.; Dalrymple, J.; Levin, E.; O'Mahony, B. (2018). Improving agricultural water sustainability. Strategies for effective farm water management and encouraging the uptake of drip irrigation. In David Crowther, Shahla Seifi, Abdul Moyeen (Eds.). The goals of sustainable development: Responsibility and Governance. Singapore: Springer Singapore, pp. 111–123.

Griggs, D.; Stafford-Smith, M.; Gaffney, O.; Rockström, J.; Öhman, M. C.; Shyamsundar, P. et al. (2013). Sustainable development goals for people and planet. In Nature 495 (7441), pp. 305–307. DOI: 10.1038/495305a.

Grote, K.; Anger, C.; Kelly, B.; Hubbard, S.; Rubin, Y. (2010). Characterization of soil water content variability and soil texture using GPR groundwave techniques. In Journal of Environmental and Engineering Geophysics 15 (3), pp. 93–110.

Grunwald, S.; Thompson, J. A.; Boettinger, J. L. (2011). Digital soil mapping and modeling at continental scales. Finding solutions for global issues. In Soil Science Society of America Journal 75 (4), pp. 1201–1213.

Guilpart, N.; Grassini, P.; van Wart, J.; Yang, H.; van Ittersum, M. K.; van Bussel, L. G. J.; et al. (2017). Rooting for food security in sub-Saharan Africa. In Environmental Research Letters 12 (11), p. 114036. DOI: 10.1088/1748-9326/aa9003.

Gutierrez-Lopez, J.; Asbjornsen, H.; Helmers, M.; Isenhart, T. (2014). Regulation of soil moisture dynamics in agricultural fields using strips of native prairie vegetation. In Geoderma 226, pp. 238–249.

Gwak, Y.; Kim, S. (2017). Factors affecting soil moisture spatial variability for a humid forest hillslope. In Hydrological Processes 31 (2), pp. 431–445. DOI: 10.1002/hyp.11039.

Haboudane, D.; Miller, J. R.; Pattey, E.; Zarco-Tejada, P. J.; Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies. Modeling and validation in the context of precision agriculture. In Remote Sensing of Environment 90 (3), pp. 337–352. DOI: 10.1016/j.rse.2003.12.013.

Hall, A. J.; Vilella, F.; Trapani, N.; Chimenti, C. (1982). The effects of water stress and genotype on the dynamics of pollen-shedding and silking in maize. In Field crops research 5, pp. 349–363. DOI: 10.1016/0378-4290(82)90036-3.

Harris, P.; Brunsdon, C.; Charlton, M. (2011). Geographically weighted principal components analysis. In International Journal of Geographical Information Science 25 (10), pp. 1717–1736.

Harris, P.; Clarke, A.; Juggins, S.; Brunsdon, C.; Charlton, M. (2015). Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set. In Geographical Analysis 47 (2), pp. 146–172. DOI: 10.1111/gean.12048.

Hauck, C.; Mühll, D. V. (2003). Inversion and interpretation of two-dimensional geoelectrical measurements for detecting permafrost in mountainous regions. In Permafrost and Periglacial Processes 14 (4), pp. 305–318. DOI: 10.1002/ppp.462.

Haykin, S. S. (2011). Neural networks and learning machines: Pearson Upper Saddle River (3).

Hazelton, P.; Murphy, B. (2016). Interpreting soil test results. What do all the numbers mean?: CSIRO publishing.

He, Z.; Zhao, M.; Zhu, X.; Du, J.; Chen, L.; Lin, P.; Li, J. (2019). Temporal stability of soil water storage in multiple soil layers in high-elevation forests. In Journal of Hydrology 569, pp. 532–545. DOI: 10.1016/j.jhydrol.2018.12.024.

Hengl, T.; Heuvelink, G. B. M.; Kempen, B.; Leenaars, J. G. B.; Walsh, M. G.; Shepherd, K. D.; et al. (2015). Mapping Soil Properties of Africa at 250 m Resolution. Random Forests Significantly Improve Current Predictions. In Plos One 10 (6), pp. 1–26. DOI: 10.1371/journal.pone.0125814.

Hengl, T.; Kempen, B.; Heuvelink, G. B.M.; Malone, B. (2014). GSIF. Global Soil Information Facilities. R package version 0.4-1.

Hengl, T.; Jesus, J. M.; Heuvelink, G. B. M.; Gonzalez, M. R.; Kilibarda, M.; Blagotic, A.; et al. (2017a). SoilGrids250m. Global gridded soil information based on machine learning. In Plos One 12 (2). DOI: 10.1371/journal.pone.0169748.

Hengl, T.; Leenaars, J. G. B.; Shepherd, K. D.; Walsh, M. G.; Heuvelink, G. B. M.; Mamo, T.; et al. (2017b). Soil nutrient maps of sub-Saharan Africa. Assessment of soil nutrient content at 250 m spatial resolution using machine learning. In Nutrient Cycling in Agroecosystems 109 (1), pp. 77–102. DOI: 10.1007/s10705-017-9870-x.

Hengl, T.; Nussbaum, M.; Wright, M. N.; Heuvelink, G. B.M.; Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. In PeerJ Preprints 6, e26693v3. DOI: 10.7287/peerj.preprints.26693v3.

Hengl, T.; Rossiter, D. G.; Stein, A. (2003). Soil sampling strategies for spatial prediction by correlation with auxiliary maps. In Soil Research 41 (8), pp. 1403–1422.

Hijmans, R. J.; van Etten, J.; Cheng, J.; Mattiuzzi, M.; Sumner, M.; Greenberg, J. A.; et al. (2015). Package 'raster'. In R package.

Holtgrave, A.; Förster, M.; Greifeneder, F.; Notarnicola, C.; Kleinschmit, B. (2018). Estimation of soil moisture in vegetation-covered floodplains with Sentinel-1 SAR data using support vector regression. In PFG-Journal of Photogrammetry, Remote Sensing and Geoinformation Science, pp. 1–17.

Hounkpatin, K. O. L.; Schmidt, K.; Stumpf, F.; Forkuor, G.; Behrens, T.; Scholten, T.; et al. (2018). Predicting reference soil groups using legacy data. A data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). In Scientific Reports 8 (1), p. 9959. DOI: 10.1038/s41598-018-28244-w.

Howell, T. A. (2001). Enhancing Water Use Efficiency in Irrigated Agriculture Contrib. from the USDA-ARS, Southern Plains Area, Conserv. and Production Res. Lab., Bushland, TX 79012. Mention of trade or commercial names is made for information only and does not imply an endorsement, recommendation, or exclusion by USDA-ARS. In Agronomy Journal 93 (2), pp. 281–289. DOI: 10.2134/agronj2001.932281x.

Huang, X.; Shi, Z. H.; Zhu, H. D.; Zhang, H. Y.; Ai, L.; Yin, W. (2016). Soil moisture dynamics within soil profiles and associated environmental controls. In Catena 136, pp. 189–196. DOI: 10.1016/j.catena.2015.01.014.

Huang, Z.; Liu, Y.; Cui, Z.; Fang, Y.; He, H.; Liu, B.; Wu, G. (2018). Soil water storage deficit of alfalfa (Medicago sativa) grasslands along ages in arid area (China). In Field Crops Research 221, pp. 1–6. DOI: 10.1016/j.fcr.2018.02.013.

Hupet, F.; Vanclooster, M. (2002). Intraseasonal dynamics of soil moisture variability within a small agricultural maize cropped field. In Journal of Hydrology 261 (1), pp. 86–101. DOI: 10.1016/S0022-1694(02)00016-1.

IPCC (2007). Climate Change 2007. The physical science basis. In Summary for policymakers. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Paris: Summary for Policymakers formally approved at the 10th Session of Working Group I of the IPCC.

Jayne, T. S.; Mather, D.; Mghenyi, E. (2010). Principal challenges confronting smallholder agriculture in sub-Saharan Africa. In World Development 38 (10), pp. 1384–1398.

Jia, X.; Shao, M.; Wei, X.; Wang, Y. (2013). Hillslope scale temporal stability of soil water storage in diverse soil layers. In Journal of Hydrology 498, pp. 254–264. DOI: 10.1016/j.jhydrol.2013.05.042.

Jost, G.; Heuvelink, G.B.M.; Papritz, A. (2005). Analysing the space–time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. In Geoderma 128 (3), pp. 258–273. DOI: 10.1016/j.geoderma.2005.04.008.

Junner, N. R. (1940). Geology of the Gold Coat and Western Togoland. In Bull. Gold Coast Geol. Surv. 11, p. 40.

Kachanoski, R. G.; Jong, E. de (1988). Scale dependence and the temporal persistence of spatial patterns of soil water storage. In Water Resources Research 24 (1), pp. 85–91. DOI: 10.1029/WR024i001p00085.

Karam, M. A.; Fung, A. K.; Lang, R. H.; Chauhan, N. S. (1992). A microwave scattering model for layered vegetation. In IEEE Transactions on Geoscience and Remote Sensing 30 (4), pp. 767–784. DOI: 10.1109/36.158872.

Karatzoglou, A.; Smola, A.; Hornik, K.; Maniscalco, M. A.; Teo, C. H. (2018). Kernlab. Kernel-Based Machine Learning Lab. R package version 0.9-27.

Kassambara, A.; Mundt, F. (2016). Package ‘factoextra’. In Extract and Visualize the Results of Multivariate Data Analyses.

Keerthi, S. S.; Lin, C. (2003). Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel. In Neural Computation 15 (7), pp. 1667–1689. DOI: 10.1162/089976603321891855.

Kermah, M.; Franke, A. C.; Adjei-Nsiah, S.; Ahiabor, B. D. K.; Abaidoo, R. C.; Giller, K. E. (2017). Maize-grain legume intercropping for enhanced resource use efficiency and crop productivity in the Guinea savannah of northern Ghana. In Field Crops Research 213, pp. 38–50. DOI: 10.1016/j.fcr.2017.07.008.

Kim, S. (2012). Characterization of annual soil moisture response pattern on a hillslope in Bongsunsa Watershed, South Korea. In Journal of Hydrology 448-449, pp. 100–111. DOI: 10.1016/j.jhydrol.2012.04.030.

Klinke, R.; Kuechly, H.; Frick, A.; Förster, M.; Schmidt, T.; Holtgrave, A.; et al. (2018). Indicator-Based Soil Moisture Monitoring of Wetlands by Utilizing Sentinel and Landsat Remote Sensing Data. In PFG-Journal of Photogrammetry, Remote Sensing and Geoinformation Science, pp. 1–14.

Koning, N.B.J.; van M.K. Ittersum; G.A., Becx; van M.A.J.S. Boekel; W.A. Brandenburg; J.A. Van Den Broek et al. (2008). Long-term global availability of food: continued abundance or new scarcity? In NJAS - Wageningen Journal of Life Sciences 55 (3), pp. 229–292. DOI: 10.1016/S1573-5214(08)80001-2.

Korres, W.; Reichenau, T. G.; Fiener, P.; Koyama, C. N.; Bogena, H. R.; Cornelissen, T. et al. (2015). Spatio-temporal soil moisture patterns – A meta-analysis using plot to catchment scale data. In Journal of Hydrology 520, pp. 326–341. DOI: 10.1016/j.jhydrol.2014.11.042.

Kotu, V.; Deshpande, B. (2018). Data Science: Concepts and Practice: Elsevier Science. Available online at https://books.google.de/books?id=-nt8DwAAQBAJ.

Kuhn, M. (2018). caret. Classification and Regression Training. R package version 6.0-81. With assistance of J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, et al.

Kumar, P.; Prasad, R.; Choudhary, A.; Gupta, D. K.; Mishra, V. N.; Vishwakarma, A. K. et al. (2018). Comprehensive evaluation of soil moisture retrieval models under different crop cover types using C-band synthetic aperture radar data. In Geocarto International, pp. 1–20.

Kumar, P.; Singh, S.; Kumar, R.; Rawal, S.; Singh, B. (2015). Effect of tuber planting depth on yield, quality and profitability of potato (Solanum tuberosum) processing varieties 60, pp. 139–144.

Kumar, S.; Lal, R.; LLoyd, C. D. (2012). Assessing spatial variability in soil characteristics with geographically weighted principal components analysis. In Computational Geosciences 16 (3), pp. 827–835.

Lark, R. M.; Marchant, B. P. (2018). How should a spatial-coverage sample design for a geostatistical soil survey be supplemented to support estimation of spatial covariance parameters? In Geoderma 319, pp. 89–99. DOI: 10.1016/j.geoderma.2017.12.022.

Leenaars, J. G. B.; Claessens, L.; Heuvelink, G. B. M.; Hengl, T.; González, M.; van Bussel, L. G. J.; et al. (2018). Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. In Geoderma 324, pp. 18–36. DOI: 10.1016/j.geoderma.2018.02.046.

Lei, S.; Chen, H.; Bian, Z.; Liu, Z. (2016). Evaluation of integrating topographic wetness index with backscattering coefficient of TerraSAR-X image for soil moisture estimation in a mountainous region. In Ecological Indicators 61, pp. 624–633.

Lenth, R. V. (2016). Least-squares means. The R package lsmeans. In Journal of statistical software 69 (1), pp. 1–33.

Levi, M. R.; Rasmussen, C. (2014). Covariate selection with iterative principal component analysis for predicting physical soil properties. In Geoderma 219, pp. 46–57. DOI: 10.1016/j.geoderma.2013.12.013.

Li, Z.; Wu, P.; Feng, H.; Zhao, X.; Huang, J.; Zhuang, W. (2009). Simulated experiment on effect of soil bulk density on soil infiltration capacity. In Transactions of the Chinese Society of Agricultural Engineering 25 (6), pp. 40–45.

Liakos, K. G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. (2018). Machine learning in agriculture. A review. In Sensors 18 (8). DOI: 10.3390/s18082674.

Liangxia D.; Mingbin H.; Zhenwei L.; Zhongdian Z.; Luodan Z. (2017). Estimation of spatial mean soil water storage using temporal stability at the hillslope scale in black locust (Robinia pseudoacacia) stands. In Catena 156, pp. 51–61. DOI: 10.1016/j.catena.2017.03.023.

Lievens, H.; Verhoest, N. E. C. (2011). On the retrieval of soil moisture in wheat fields from L-band SAR based on water cloud modeling, the IEM, and effective roughness parameters. In IEEE Geoscience and Remote Sensing Letters 8 (4), pp. 740–744. DOI: 10.1109/LGRS.2011.2106109.

Liujun, Z.; Jeffrey, W.; Leung, T.; Huanting, H.; Nan, Y.; Christoph, R. (2019). Soil moisture retrieval from time series multi-angular radar data using a dry down constraint. In Remote Sensing of Environment 231, p. 111237. DOI: 10.1016/j.rse.2019.111237.

Loew, A.; Ludwig, R.; Mauser, W. (2006). Derivation of surface soil moisture from ENVISAT ASAR wide swath and image mode data in agricultural areas. In IEEE Transactions on Geoscience and Remote Sensing 44 (4), pp. 889–899. DOI: 10.1109/TGRS.2005.863858.

Lowder, S. K.; Skoet, J.; Raney, T. (2016). The number, size, and distribution of farms, smallholder farms, and family farms worldwide. In World Development 87, pp. 16–29. DOI: 10.1016/j.worlddev.2015.10.041.

Martínez-Fernández, J. (2003). Temporal stability of soil moisture in a large-field experiment in Spain. In Soil Science Society of America Journal v. 67 (no. 6), pp. 1647-1656-2003 v.67 no.6. DOI: 10.2136/sssaj2003.1647.

Mashnik, D.; Jacobus, H.; Barghouth, A.; Wang, E. J.; Blanchard, J.; Shelby, R. (2017). Increasing productivity through irrigation. Problems and solutions implemented in Africa and Asia. In Sustainable Energy Technologies and Assessments 22, pp. 220–227. DOI: 10.1016/j.seta.2017.02.005.

McBratney, A. B.; Santos, M. L. Mendonça; Minasny, B. (2003). On digital soil mapping. In Geoderma 117 (1), pp. 3–52. DOI: 10.1016/S0016-7061(03)00223-4.

Megersa, G.; Abdulahi, J. (2015). Irrigation system in Israel. A review. In International Journal of Water Resources and Environmental Engineering 7 (3), pp. 29–37.

Mei, X.; Ma, L.; Zhu, Q.; Li, B.; Zhang, D.; Liu, H.; et al. (2019). The variability in soil water storage on the loess hillslopes in China and its estimation. In Catena 172, pp. 807–818. DOI: 10.1016/j.catena.2018.09.045.

Mei, X.; Zhu, Q.; Ma, L.; Zhang, D.; Liu, H.; Xue, M. (2018). The spatial variability of soil water storage and its controlling factors during dry and wet periods on loess hillslopes. In Catena 162, pp. 333–344. DOI: 10.1016/j.catena.2017.10.029.

Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. In Environmental Modelling & Software 101, pp. 1–9. DOI: 10.1016/j.envsoft.2017.12.001.

Michailidis, M. (2017). Investigating machine learning methods in recommender systems: Doctoral dissertation, UCL (University College London).

Minasny, B.; McBratney, A. B. (2010). Methodologies for global soil mapping. In Janis, L. B., David, W. H., Amanda, C. M., Alfred, E. H., Suzann, K. (Eds.). Digital soil mapping: Bridging research, environmental application, and operation. Dordrecht: Springer Netherlands, pp. 429–436.

Minasny, B.; McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. In Computers & Geosciences 32 (9), pp. 1378–1388.

Mitchell, T. (2014). Geospatial Power Tools. GDAL Raster & Vector Commands: Locate Press.

Mohamad, I. B.; Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. In Research Journal of Applied Sciences, Engineering and Technology 6 (17), pp. 3299–3303.

Mohammed, T.; Salifu, I.; Saeed, A. S.; Goodness, A. (2019). Planting for Food and Jobs (PFJ). A panacea for productivity and welfare of rice farmers in Northern Ghana. In Cogent Economics & Finance 7 (1), p. 1693121. DOI: 10.1080/23322039.2019.1693121.

Mohanty, B. P.; Cosh, Michael H.; Lakshmi, V.; Montzka, C. (2017). Soil moisture remote sensing. State-of-the-science. In Vadose Zone Journal 16 (1). DOI: 10.2136/vzj2016.10.0105.

Molatudi, R. L.; Mariga, I. K. (2009). The effect of maize seed size and depth of planting on seedling emergence and seedling vigour. In Journal of Applied Sciences Research 5 (12), pp. 2234–2237.

Montanarella, L.; Vargas, R. (2012). Global governance of soil resources as a necessary condition for sustainable development. In Current Opinion in Environmental Sustainability 4 (5), pp. 559–564. DOI: 10.1016/j.cosust.2012.06.007.

Morris, M.; Binswanger-Mkhize, H. P.; Byerlee, D. (2009). Awakening Africa's sleeping giant. Prospects for commercial agriculture in the Guinea Savannah Zone and beyond: The World Bank.

Nedkov, R. (2017). Orthogonal transformation of segmented images from the satellite Sentinel-2. In Comptes rendus de l'Acad´emie bulgare des Sciences 70 (5), pp. 687–692.

Nketia, K. A.; Asabere, S. B.; Erasmi, S.; Sauer, D. (2019). A new method for selecting sites for soil sampling, coupling global weighted principal component analysis and a cost-constrained conditioned Latin hypercube algorithm. In MethodsX 6, pp. 284–299. DOI: 10.1016/j.mex.2019.02.005.

Notarnicola, C.; Angiulli, M.; Posa, F. (2008). Soil moisture retrieval from remotely sensed data. Neural network approach versus Bayesian method. In IEEE Transactions on Geoscience and Remote Sensing 46 (2), pp. 547–557.

OECD (2016). OECD-FAO, United Nations - Agricultural Outlook 2016-2025. Available online at https://www.oecd-ilibrary.org/content/publication/agr_outlook-2016-en.

Ogundare, S. K. (2017). Effect of depth of planting, methods of planting and animal residues application on the growth and yield performance of cassava in Ejiba, Kogi state, Nigeria. In Nigeria Agricultural Journal 48 (1), pp. 17–25.

Oki, T.; Kanae, S. (2006). Global Hydrological Cycles and World Water Resources. In Science 313 (5790), pp. 1068–1072. DOI: 10.1126/science.1128845.

Omar, H.; Misman, M. A.; Kassim, A. R. (2017). Synergetic of PALSAR-2 and Sentinel-1A SAR Polarimetry for Retrieving Aboveground Biomass in Dipterocarp Forest of Malaysia. In Applied Sciences 7 (7). DOI: 10.3390/app7070675.

Padarian, J.; Minasny, B.; McBratney, A. B. (2019a). Using deep learning for digital soil mapping. In Soil 5 (1), pp. 79–89. DOI: 10.5194/soil-5-79-2019.

Padarian, J.; Minasny, B.; McBratney, A. B. (2019b). Using deep learning to predict soil properties from regional spectral data. In Geoderma Regional 16, e00198. DOI: 10.1016/j.geodrs.2018.e00198.

Pal, M. (2005). Random forest classifier for remote sensing classification. In International Journal of Remote Sensing 26 (1), pp. 217–222. DOI: 10.1080/01431160412331269698.

Paloscia, S.; Pettinato, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Reppucci, A. (2013). Soil moisture mapping using Sentinel-1 images. Algorithm and preliminary validation. In Remote Sensing of Environment 134, pp. 234–248. DOI: 10.1016/j.rse.2013.02.027.

Pasolli, L.; Notarnicola, C.; Bertoldi, G.; Bruzzone, L.; Remelgado, R.; Greifeneder, F. et al. (2015). Estimation of soil moisture in mountain areas using SVR technique applied to multiscale active radar images at C-band. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8 (1), pp. 262–283.

Pasolli, L.; Notarnicola, C.; Bertoldi, G.; Della Chiesa, S.; Niedrist, G.; Bruzzone, L. et al. (2014). Soil moisture monitoring in mountain areas by using high-resolution SAR images. Results from a feasibility study. In European Journal of Soil Science 65 (6), pp. 852–864. DOI: 10.1111/ejss.12189.

Pasolli, L.; Notarnicola, C.; Bruzzone, L. (2011). Estimating soil moisture with the support vector regression technique. In IEEE Geoscience and Remote Sensing Letters 8 (6), pp. 1080–1084.

Pasolli, L.; Notarnicola, C.; Bruzzone, L. (2012). Multi-objective parameter optimization in support vector regression. General formulation and application to the retrieval of soil moisture from remote sensing data. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (5), pp. 1495–1508.

Passioura, J. B.; Angus, J. F. (2010). Chapter 2 - Improving productivity of crops in water-limited environments. In Donald L. Sparks (Ed.). Advances in Agronomy v106, vol. 106: Academic Press (Advances in Agronomy), pp. 37–75.

Peng, J.; Loew, A. (2017). Recent advances in soil moisture estimation from remote sensing. In Water 9 (7). DOI: 10.3390/w9070530.

Pérez, R. P.; Gianola, D. (2018). BRNN. Bayesian regularization for feed-forward neural networks. R package version 0.7.

Petropoulos, G. P. (2013). Remote Sensing of Energy Fluxes and Soil Moisture Content: CRC Press.

Piao, Y.; Park, H. W.; Jin, C. H.; Ryu, K. H. (2014). Ensemble method for classification of high-dimensional data. In: Big Data and Smart Computing (BIGCOMP), 2014 International Conference on. IEEE, pp. 245–249.

Piles, M.; Entekhabi, D.; Camps, A. (2009). A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations. In IEEE Transactions on Geoscience and Remote Sensing 47 (12), pp. 4125–4131. DOI: 10.1109/TGRS.2009.2022088.

Pradhan, P.; Fischer, G.; van Velthuizen, H.; Reusser, D. E.; Kropp, J. P. (2015). Closing yield gaps: How sustainable can we be? In Plos One 10 (6), pp. 1–18. DOI: 10.1371/journal.pone.0129487.

Quegan, S.; Yu, J. J. (2001). Filtering of multichannel SAR images. In IEEE Transactions on Geoscience and Remote Sensing 39 (11), pp. 2373–2379.

R Core Team (2017). R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Ramcharan, A.; Hengl, T.; Nauman, T.; Brungard, C.; Waltman, S.; Wills, S.; Thompson, J. (2018). Soil property and class maps of the conterminous United States at 100-meter spatial resolution. In Soil Science Society of America Journal 82 (1), pp. 186–201.

Rawls, W. J.; Pachepsky, Y. A.; Ritchie, J. C.; Sobecki, T. M.; Bloodworth, H. (2003). Effect of soil organic carbon on soil water retention. In Geoderma 116 (1-2), pp. 61–76.

Reid, S.; Grudic, G. (2009). Regularized Linear Models in Stacked Generalization. In Benediktsson, J. A.; Kittler, J.; Roli, F. (Eds.). Multiple Classifier Systems. Berlin, Heidelberg, 2009: Springer Berlin Heidelberg, pp. 112–121.

Ren, A.; Sun, M.; Xue, L.; Deng, Y.; Wang, P.; Lei, M.; et al. (2019). Spatio-temporal dynamics in soil water storage reveals effects of nitrogen inputs on soil water consumption at different growth stages of winter wheat. In Agricultural Water Management 216, pp. 379–389. DOI: 10.1016/j.agwat.2019.01.023.

Revelle, W. (2018). Psych package [program].

Reynolds, S. G. (1970). The gravimetric method of soil moisture determination Part III An examination of factors influencing soil moisture variability. In Journal of Hydrology 11 (3), pp. 288–300.

Rosich, B.; Zink, M.; Torres, R.; Closa, J.; Buck, C. (2003). ASAR instrument performance and product quality status. In : Geoscience and Remote Sensing Symposium, 2003. IGARSS’03. Proceedings. 2003 IEEE International, vol. 2. IEEE, pp. 1109–1111.

Rossato, L.; Alvalá, R. C. d. S.; Marengo, J. A.; Zeri, M.; Cunha, A. P. M. d. A.; Pires, L. B. M.; Barbosa, H. A. (2017). Impact of soil moisture on crop yields over Brazilian semiarid. In Frontiers in Environmental Science 5, p. 73. DOI: 10.3389/fenvs.2017.00073.

Roudier, P.; Beaudette, D. E.; Hewitt, A. E. (2012). A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In Digital soil assessments and beyond, pp. 227–231.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. In arXiv preprint arXiv:1706.05098.

Sanchez, P. A.; Ahamed, S.; Carré, F.; Hartemink, A. E.; Hempel, J.; Huising, J.; et al. (2009). Digital Soil Map of the World. In Science 325 (5941), pp. 680–681. DOI: 10.1126/science.1175084.

Sánchez, N.; Alonso-Arroyo, A.; Martínez-Fernández, J.; Piles, M.; et al. (2015). On the Synergy of Airborne GNSS-R and Landsat 8 for Soil Moisture Estimation. In Remote Sensing 7 (8), pp. 9954–9974. DOI: 10.3390/rs70809954.

Sánchez-Ruiz, S.; Piles, M.; Sánchez, N.; Martínez-Fernández, J.; et al. (2014). Combining SMOS with visible and near/shortwave/thermal infrared satellite data for high resolution soil moisture estimates. In Journal of Hydrology 516, pp. 273–283. DOI: 10.1016/j.jhydrol.2013.12.047.

Schilt-van, E. C.; Giller, K. E.; Thuijsman, E. (2017). N2Africa Putting nitrogen fixation to work for smallholder farmers in Africa, Podcaster no. 47, August and September 2017. N2Africa project.

Schlund, M.; Scipal, K.; Quegan, S. (2018). Assessment of a power law relationship between P-band SAR backscatter and aboveground biomass and Its Implications for biomass mission performance. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (99), pp. 1–10.

Sebastian, K. (2009). Agro-ecological zones of Africa. With assistance of Harvest Choice. DRAFT VERSION: Harvard Dataverse.

Seneviratne, S. I.; Corti, T.; Davin, E. L.; Hirschi, M.; Jaeger, E. B.; Lehner, I. et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. In Earth-Science Reviews 99 (3), pp. 125–161. DOI: 10.1016/j.earscirev.2010.02.004.

Sheaffer, C. C.; Moncada, K. M. (2012). Introduction to agronomy. Food, crops, and environment: Cengage Learning.

Shepherd, K. D.; Shepherd, G.; Walsh, M. G. (2015). Land health surveillance and response. A framework for evidence-informed land management. In Agricultural Systems 132, pp. 93–106. DOI: 10.1016/j.agsy.2014.09.002.

Smilovic, M.; Gleeson, T.; Siebert, S. (2015). The limits of increasing food production with irrigation in India. In Food Security 7 (4), pp. 835–856. DOI: 10.1007/s12571-015-0477-2.

Smola, A. J.; Schölkopf, B. (2004). A tutorial on support vector regression. In Statistics and Computing 14 (3), pp. 199–222. DOI: 10.1023/B: STCO.0000035301.49549.88.

SRID (2016). Agriculture in Ghana; Facts and figures. Statistics, Research and Information Directorate, Ministry of Food and Agriculture, Ghana. Accra.

Srivastava, P. K.; Petropoulos, G. P.; Kerr, Y. H. (2016). Satellite soil moisture retrieval. Techniques and applications / edited by Prashant K. Srivastava, George P. Petropoulos, Yann H. Kerr. Amsterdam, Netherlands: Elsevier.

Steichen, T. J.; Cox, N. J. (2002). A note on the concordance correlation coefficient. In the Stata Journal 2 (2), pp. 183–189. DOI: 10.1177/1536867X0200200206.

Strobl, C.; Boulesteix, A.; Zeileis, A.; Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. In BMC Bioinformatics 8 (1), p. 25. DOI: 10.1186/1471-2105-8-25.

Sure, A.; Dikshit, O. (2019). Estimation of root zone soil moisture using passive microwave remote sensing. A case study for rice and wheat crops for three states in the Indo-Gangetic basin. In Journal of Environmental Management 234, pp. 75–89. DOI: 10.1016/j.jenvman.2018.12.109.

Tallon, L. K.; Si, B. C. (2015). Representative soil water benchmarking for environmental monitoring. In Journal of Environmental Informatics 4 (1), pp. 31–39.

Tilman, D.; Balzer, C.; Hill, J.; Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. In Proceedings of the National Academy of Sciences 108 (50), pp. 20260–20264. DOI: 10.1073/pnas.1116437108.

Tomar, D.; Arya, R.; Agarwal, S. (2011). Prediction of profitability of industries using weighted SVR. In International Journal on Computer Science and Engineering 3 (5), pp. 1938–1945.

Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; et al. (2012). GMES Sentinel-1 mission. In Remote Sensing of Environment 120, pp. 9–24.

Ulaby, F. T.; Moore, R. K.; Fung, A. K. (1982). Microwave remote sensing. Active and passive. Vol. 2, Radar remote sensing and surface scattering and emission theory: Addison-Wesley Reading, MA.

Ulaby, F. T.; Long, D. G.; Blackwell, W. J.; University of Michigan Press. (2014). Microwave radar and radiometric remote sensing. Ann Arbor: The University of Michigan Press.

Urban, M.; Berger, C.; Mudau, T. E.; Heckel, K.; Truckenbrodt, J.; Onyango O., V.; et al. (2018). Surface moisture and vegetation cover analysis for drought monitoring in the southern Kruger National Park using Sentinel-1, Sentinel-2, and Landsat-8. In Remote Sensing 10 (9). DOI: 10.3390/rs10091482.

Vachaud, G.; Passerat D. S. A.; Balabanis, P.; Vauclin, M. (1985). Temporal stability of spatially measured soil water probability density function. In Soil Science Society of America Journal 49 (4), pp. 822–828. DOI: 10.2136/sssaj1985.03615995004900040006x.

Vågen, T. G.; Shepherd, Keath D.; Walsh, Markus G.; Winowiecki, Leigh; Desta, Lulseged Tamene; Tondoh, Jerome E. (2010). AfSIS technical specifications. Soil health surveillance. In Africa Soil Information Service (AfSIS). Nairobi, Kenya: World Agroforestry Centre.

van Ittersum, M. K.; van Bussel, L. G. J.; Wolf, J.; Grassini, P.; van Wart, J.; Guilpart, N.; et al. (2016). Can sub-Saharan Africa feed itself? In Proceedings of the National Academy of Sciences 113 (52), pp. 14964–14969. DOI: 10.1073/pnas.1610359113.

Vanlauwe, B.; Coyne, D.; Gockowski, J.; Hauser, S.; Huising, J.; Masso, C.; et al. (2014). Sustainable intensification and the African smallholder farmer. In Current Opinion in Environmental Sustainability 8 (0), pp. 15–22.

Vapnik, V. (2013). The nature of statistical learning theory: Springer Science & Business Media.

Veste, M.; Staudinger, M.; Küppers, M. (2008). Spatial and temporal variability of soil water in drylands. Plant water potential as a diagnostic tool. In Forestry Studies in China 10 (2), p. 74. DOI: 10.1007/s11632-008-0022-x.

Vorontsov, K. V. (2009). Splitting and similarity phenomena in the sets of classifiers and their effect on the probability of overfitting. In Pattern Recognition and Image Analysis 19 (3), pp. 412–420. DOI: 10.1134/S1054661809030055.

Wagner, W.; Dorigo, W.; Jeu, Ri.; Fernandez, D.; Benveniste, J.; Haas, E.; et al. (2012). Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals) 7, pp. 315–321.

Wagner, W.; Pathe, C.; Doubkova, M.; Sabel, D.; Bartsch, A.; Hasenauer, S.; et al. (2008). Temporal stability of soil moisture and radar backscatter observed by the Advanced Synthetic Aperture Radar (ASAR). In Sensors 8 (2), pp. 1174–1197. DOI: 10.3390/s80201174.

Wang, A.; Liu, B.; Wang, Z.; Liu, G. (2016). Monitoring and predicting the soil water content in the deeper soil profile of Loess Plateau, China. In International Soil and Water Conservation Research 4 (1), pp. 6–11. DOI: 10.1016/j.iswcr.2016.02.001.

WCSP (2014). World checklist of selected plant families. In Facilitated by the Royal Botanic Gardens, Kew.

Weil, R.; Brady, N. (2016). The Nature and Properties of Soils. 15th edition.

Western, A. W.; Grayson, R. B.; Blöschl, G.; Willgoose, G. R.; McMahon, T. A. (1999). Observed spatial organization of soil moisture and its relation to terrain indices. In Water Resources Research 35 (3), pp. 797–810.

Wickham, H. (2018). Package 'plyr'. In Obtenido de https://cran. rproject. org/web/packages/dplyr/dplyr. pdf.

Wigneron, J. P; Calvet, J. C; Pellarin, T.; van de Griend, A. A.; Berger, M.; Ferrazzoli, P. (2003). Retrieving near-surface soil moisture from microwave radiometric observations. Current status and future plans. In Remote Sensing of Environment 85 (4), pp. 489–506. DOI: 10.1016/S0034-4257(03)00051-8.

Williams, P. A.; Crespo, O.; Abu, M. (2018). Assessing vulnerability of horticultural smallholders' to climate variability in Ghana: applying the livelihood vulnerability approach. In Environment, Development and Sustainability. DOI: 10.1007/s10668-018-0292-y.

World Bank (2017a). Agricultural irrigated land (% of total agricultural land) 2001-2014. World Bank Group. Washington, D.C.

World Bank (2017b). Ghana: Agriculture sector policy note. transforming agriculture for economic growth, job creation and food security. World Bank Group. Washington, DC.

Wright, M. N.; Ziegler, A. (2015). Ranger. A fast implementation of random forests for high dimensional data in C++ and R. In arXiv preprint arXiv:1508.04409.

Xie, H.; Perez, N.; Anderson, W.; Ringler, C.; You, L. (2018). Can Sub-Saharan Africa feed itself? The role of irrigation development in the region’s drylands for food security. In Water International 43 (6), pp. 796–814. DOI: 10.1080/02508060.2018.1516080.

Yang, L.; Chen, L.; Wei, W. (2015). Effects of vegetation restoration on the spatial distribution of soil moisture at the hillslope scale in semi-arid regions. In Catena 124, pp. 138–146. DOI: 10.1016/j.catena.2014.09.014.

Yusuf, A. A.; Iwuafor, E. N. O.; Abaidoo, R. C.; Olufajo, O. O.; Sanginga, N. (2009). Grain legume rotation benefits to maize in the northern Guinea savannah of Nigeria. Fixed-nitrogen versus other rotation effects. In Nutrient Cycling in Agroecosystems 84 (2), pp. 129–139. DOI: 10.1007/s10705-008-9232-9.

Zhang, Y.; Shangguan, Z. (2016). The change of soil water storage in three land use types after 10years on the Loess Plateau. In Catena 147, pp. 87–95. DOI: 10.1016/j.catena.2016.06.036.

Zhao, Y.; Peth, S.; Wang, X. Y.; Lin, H.; Horn, R. (2010). Controls of surface soil moisture spatial patterns and their temporal stability in a semi-arid steppe. In Hydrological Processes 24 (18), pp. 2507–2519. DOI: 10.1002/hyp.7665.

Zhao, Y.; Wang, Y.; Wang, L.; Fu, Z.; Zhang, X.; Cui, B. (2017). Soil-water storage to a depth of 5m along a 500-km transect on the Chinese Loess Plateau. In Catena 150, pp. 71–78. DOI: 10.1016/j.catena.2016.11.008.

Zribi, M.; Taconet, O.; Le Hégarat-Mascle, S.; Vidal-Madjar, D.; Emblanch, C.; Loumagne, C.; Normand, M. (1997). Backscattering behavior and simulation comparison over bare soils using SIR-C/X-SAR and ERASME 1994 data over Orgeval. In Remote Sensing of Environment 59 (2), pp. 256–266. DOI: 10.1016/S0034-4257(96)00158-7.

Appendix

Equations

Equation S6.1. Sentinel-1 image calibration.

$$C_i = \frac{[DN_i]^2}{A_i^2} \qquad \text{(S6.1)}$$

where, depending on the selected look up table, C_i is the Sigma0 (σ_i^0) at a pixel location *i*, DN_i is the digital number and A_i is the Sentinel-1 final scaling from internal single look complex to ground range detected.

Equation S6.2. Converting calibrated Sigma0 (σ) to decibels.

$$\sigma_{db}^0 = 10.\, log_{10}\sigma^0 \qquad \text{(S6.2)}$$

where σ_{db}^0 is the Sigma0 in decibels (dB).

Equation S6.3. Estimating the coefficient of determination of variations (R^2) explained by the base MLAs.

$$R^2 = \left[\frac{\sum(SM_p - \overline{SM_p}\,)\; x\; (SM_e - \overline{SM_e}\,)}{\sqrt{\sum(SM_p - \overline{SM_p}\,)^2}\; x\; \sqrt{\sum(SM_e - \overline{SM_e}\,)^2}}\right]^2 \qquad \text{(S6.3)}$$

where SM_p and SM_e are the *in situ* and estimated SM respectively, whereas $\overline{SM_p}$ and $\overline{SM_e}$ are also the mean *in situ* and estimated SM respectively.

Equation S6.4. Estimating the root mean squared error (RMSE) of the MLAs.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[SM_{e_i} - SM_{p_i}\right]^2} \qquad \text{(S6.4)}$$

where at a location, SM_{e_i} and SM_{p_i} are the estimated and *in situ* SM respectively.

Equation S6.5. Estimating the percent bias (PBias) of the MLAs for over- or underestimated SM.

$$PBias = 100\; x \left[\frac{\sum_{i=1}^{n} SM_{e_i} - SM_{p_i}}{\sum_{i=1}^{n} SM_{p_i}}\right] \qquad \text{(S6.5)}$$

where at a location, SM_{e_i} and SM_{p_i} are the estimated and *in situ* SM respectively.

Equation S7.1. Coefficient of determination (R^2).

$$R^2 = 1 - \frac{SS_{error}}{SS_{total}} \qquad \text{(S7.1)}$$

where SS_{error} is the residual sum of squares at the LLTO-CV locations and the SS_{total} represents the total sum of squares. A model with R^2 = 1 (i.e. 100%) indicates a perfect *goodness-of-fit* and vice versa.

Equation S7.2. Root mean square error (RMSE).

$$RMSE = \left[\frac{1}{N}\sum_{k=1}^{N}(pred_i - obs_i)^2\right]^{1/2} \tag{S7.2}$$

where N is the number of LLTO-CV points, $pred_i$ is the predicted SWS value at the estimated SWS (obs_i) location (i).

Equation S7.3. Lin's concordance correlation coefficient (CCC).

$$CCC = \frac{2 * \rho * \sigma_{pred} * \sigma_{obs}}{\sigma_{pred}^2 + \sigma_{obs}^2 + (\mu_{pred} - \mu_{obs})^2} \tag{S7.3}$$

where μ_{pred} and μ_{obs} are the predicted and estimated SWS means respectively, σ_{pred} and σ_{obs} are the variances at the LLTO-CV locations, and ρ is the correlation coefficient between the predicted and estimated SWS values.

Tables

Table S4.1. Principal component matrix of the GWPCA

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
EV	5.249	3.916	2.079	1.598	1.325	1.117
CPoV	0.747		0.812	0.847	0.881	0.915
Loadings						
AWC	-0.150	**0.382**	-0.177	-0.118	-0.079	**0.259**
BD	0.145	**-0.417**	0.141	-0.006	-0.026	-0.197
Bedrck	0.175	0.040	0.289	0.295	-0.031	-0.261
Clay	0.179	0.365	0.162	-0.002	0.270	-0.102
DEM	**-0.368**	-0.081	0.214	-0.017	0.035	-0.144
Drainage	-0.257	-0.150	-0.229	-0.146	0.083	-0.129
Geology	-0.025	-0.051	-0.326	0.046	0.027	-0.077
Landforms	-0.168	-0.004	-0.140	-0.107	**0.552**	-0.147
Lithology	-0.168	0.056	**0.297**	0.008	0.176	-0.408
Precip	-0.325	0.181	0.266	0.077	-0.092	0.224
Riverdist	-0.129	0.024	0.059	**-0.445**	**-0.187**	**-0.453**
Sand	-0.220	-0.388	-0.118	0.138	-0.152	0.034
Sent1A_VH	-0.150	0.261	**-0.355**	0.317	-0.141	-0.264
Sent1A_VV	-0.197	0.213	-0.339	0.335	-0.160	-0.254
Silt	0.216	0.349	0.053	-0.248	0.016	0.044
Soil_types	-0.076	-0.173	0.167	**0.446**	0.131	0.258
Temp	**0.374**	-0.083	-0.270	0.000	-0.002	0.081
WatBal	-0.345	0.172	0.249	0.067	-0.090	0.199
WATCov	0.155	0.115	0.014	0.373	0.393	-0.168
Slope	-0.260	-0.099	-0.162	-0.132	0.411	0.145
Maximum	0.374	0.382	0.297	0.446	0.552	0.259
Minimum	-0.368	-0.417	-0.355	-0.445	-0.187	-0.453

PC = principal component, EV = eigenvalues, CPoV = cumulative proportion of variation and bold values represent minimum and maximum values.

Table S5.1. Mean values of soil moisture and physical soil properties of the different soil types in the study area

Statistic	Moisture content		Particle size distribution		BD
	SM_g	SM_p	Silt	Clay	
	%				g cm^{-3}
R^2	0.68	0.66	0.16	0.64	0.18
Upland slopes					
Kpelesawgu series (n = 15)					
LSM	10.67	11.80	34.74	8.91	1.54
SE (±)	1.13	1.63	3.33	1.15	0.03
Min	1.79	0.15	26.00	4.00	1.29
Max	24.28	28.58	50.00	20.00	1.82
Middle to lower slopes					
Changnalili series (n = 10)					
LSM	11.92	12.89	28.68	11.42	1.51
SE (±)	1.40	2.03	4.14	1.44	0.03
Min	1.28	1.08	14.00	4.00	1.24
Max	34.21	37.70	52.00	44.00	1.83
Kumayili series (n = 15)					
LSM	11.12	12.43	25.76	7.35	1.41
SE (±)	1.16	1.67	3.36	1.17	0.03
Min	2.72	0.08	4.00	4.00	1.12
Max	19.63	26.50	48.00	16.00	1.62
Lima series (n = 105)					
LSM	9.23	15.07	36.09	9.61	1.50
SE (±)	0.43	0.63	1.27	0.44	0.01
Min	0.97	0.05	2.00	4.00	1.17
Max	34.97	53.58	60.00	34.00	1.77

LSM = least square mean, r^2 = coefficient of determination of the LSM at 0.95 confidence level, SE (±) = standard error, SM_g = volumetric soil moisture by gravimetric method, SM_p = volumetric soil moisture by PR2/60 moisture probe and BD = bulk density.

Table S5.1. Continued

Statistic	Moisture content		Particle size distribution		BD
	SM_g	SM_p	Silt	Clay	
	%				g cm^{-3}
R^2	0.68	0.66	0.16	0.64	0.18
Valley bottom slopes					
Dagare series (n = 5)					
LSM	10.61	12.58	14.30	7.02	1.38
SE (±)	2.03	2.94	5.80	2.07	0.05
Min	4.76	5.20	10.00	4.00	1.40
Max	6.99	12.35	18.00	8.00	1.51
Siare series (n = 5)					
LSM	19.67	15.25	32.87	16.19	1.40
SE (±)	2.35	3.58	7.32	2.49	0.06
Min	26.91	13.73	34.00	10.00	1.14
Max	42.74	60.75	50.00	40.00	1.55
Volta series (n = 15)					
LSM	8.11	14.75	33.93	13.01	1.50
SE (±)	1.14	1.66	3.39	1.15	0.03
Min	1.59	5.23	8.00	4.00	1.38
Max	24.38	33.73	56.00	40.00	1.73

LSM = least square mean, r^2 = coefficient of determination of the LSM at 0.95 confidence level, SE (±) = standard error, SM_g = volumetric soil moisture by gravimetric method, SM_p = volumetric soil moisture by PR2/60 moisture probe and BD = bulk density.

Table S5.2. Mean values of soil moisture and physical soil properties at standard soil depths of the study area

Statistic	**Moisture content**		**Particle size distribution**		**BD**
	SM_g	**SM_p**	**Silt**	**Clay**	
	%				**g cm^{-3}**
R^2	0.68	0.66	0.16	0.64	0.18
At 10 cm					
LSM	8.61	7.79	29.80	7.10	1.42
SE (±)	0.89	1.27	2.58	0.88	0.02
Min	0.97	0.05	4.00	4.00	1.14
Max	26.91	22.30	60.00	14.00	1.63
At 20 cm					
LSM	10.26	10.39	29.40	7.86	1.46
SE (±)	0.85	1.28	2.59	0.90	0.02
Min	1.79	0.15	6.00	4.00	1.17
Max	34.97	28.48	58.00	32.00	1.77
At 30 cm					
LSM	7.97	15.19	30.30	10.38	1.46
SE (±)	0.86	1.24	2.51	0.88	0.02
Min	2.07	3.40	10.00	4.00	1.31
Max	38.55	43.68	58.00	28.00	1.73
At 40 cm					
LSM	10.15	14.98	30.22	12.38	1.48
SE (±)	0.90	1.30	2.64	0.90	0.02
Min	2.76	3.95	2.00	4.00	1.12
Max	42.74	50.53	60.00	36.00	1.83
At 60 cm					
LSM	12.64	19.34	27.68	14.79	1.50
SE (±)	1.05	1.47	2.96	0.96	0.02
Min	3.98	5.78	8.00	6.00	1.32
Max	35.32	60.75	56.00	44.00	1.82

n = 34, LSM = least square mean, r^2 = coefficient of determination of the LSM at 0.95 confidence level, SE (±) = standard error, SM_g = volumetric soil moisture by gravimetric method, SM_p = volumetric soil moisture by PR2/60 moisture probe and BD = bulk density.

Table S5.3. Mean values of soil moisture, local slope and topographic wetness index of different soil types of the study area

Soil types	**SM_g**		**LS**		**TWI**	
	Mean	**SE (±)**	**Mean**	**SE (±)**	**Mean**	**SE (±)**
Upland slopes						
Kpelesawgu series *(n = 3)*	10.67	1.10	1.45	0.20	11.50	0.30
Middle to lower slopes						
Changnalili series *(n = 2)*	11.92	1.40	0.80	0.13	13.80	0.27
Kumayili series *(n = 3)*	11.12	1.20	1.40	0.10	11.30	0.10
Lima *(n = 21)*	9.23	0.40	1.00	0.05	12.30	0.08
Valley bottom slopes						
Siare series *(n = 1)*	19.67	2.30	0.98	0.00	15.20	0.00
Volta series *(n = 3)*	8.11	1.10	0.92	0.17	13.20	0.32
Dagare series *(n = 1)*	10.61	2.00	0.10	0.00	14.10	0.00

SE (±) = standard error, SM_g = volumetric soil moisture by gravimetric method, LS = local slope and TWI = topographic wetness index.

Table 5.4. Thresholds of SM demands (PAWC) for maize (*Zea mays)* in the dry season of the study

Station ID	Upper soil depth	Lower soil depth	Layer thickness	BD	Sand	Silt	Clay	SAT	DUL	CLL$_a$	CLL$_m$	PAWC$_m$	PAWC$_a$	PAWC$_m$*	PAWC$_a$*
	(cm)			g cm^{-3}	(%)							(mm)			
P1	0	10	10	1.39	80.16	14.38	5.46	40.64	23.66	12.70	6.29	10.95	17.36	57.16	162.96
P2	10	20	10	1.45	78.30	15.07	6.63	38.37	19.38	11.69	4.37	7.69	15.01		
P3	20	30	10	1.34	68.57	18.94	12.49	42.42	17.60	12.22	3.90	5.38	13.70		
P4	30	40	10	1.48	56.35	20.20	23.45	37.27	23.43	14.90	5.02	8.53	18.40		
P5	40	60	20	1.64	61.90	20.05	18.05	31.27	26.31	20.27	8.95	12.08	34.72		
P6	60	100	40	1.66	53.77	25.83	20.40	30.53	30.22	27.09	14.28	12.52	63.77		

BD = bulk density, SAT = saturation, DUL = drained upper limit, CLL = crop lower limit, ‘*’ presents PAWC for the entire soil profile (i.e. 0–100 cm soil depth). Subscripts ‘a’ and ‘m’ indicates anthesis and maturity stages of maize.

Table S5.5. Mean values of temporal SM$_p$ of the different soil types at standard depth intervals

Depth	Measurements									
intervals	1	2	3	4	5	6	7	8	9	10
cm	12 – day interval									
Upland slopes										
Kpelesawgu series										
0 — 10	0.10	0.05	0.43	0.64	2.74	0.49	0.24	0.23	0.55	1.55
10 — 20	6.64	6.96	7.93	15.16	13.60	10.49	9.65	10.45	7.83	17.13
20 — 30	15.86	15.56	15.13	19.11	18.63	20.54	16.66	19.69	16.26	25.04
30 — 40	23.29	22.26	21.61	23.30	23.46	25.53	23.33	23.89	22.08	29.16
40 — 60	28.61	27.54	25.95	27.64	28.15	30.33	29.11	31.67	27.85	29.65
60 — 100	16.36	17.81	17.75	17.75	17.83	17.95	17.85	17.75	17.55	17.54
Middle to lower slopes										
Changnalili series										
0 — 10	0.15	5.28	1.08	0.08	5.65	3.95	5.22	8.35	9.78	17.55
10 — 20	6.13	7.22	6.85	12.63	12.10	11.38	14.58	18.43	13.25	15.24
20 — 30	10.30	10.65	10.83	11.70	11.60	12.69	15.10	19.34	17.09	23.12
30 — 40	22.50	23.96	23.46	25.66	25.71	26.08	27.59	31.16	27.26	25.90
40 — 60	55.73	40.95	36.26	33.58	32.04	31.80	44.35	46.25	45.61	45.94
60 — 100	56.00	51.18	52.03	50.62	51.16	51.23	50.13	52.24	52.33	53.08
Kumayili series										
0 — 10	0.08	0.15	1.18	2.31	3.68	1.77	2.20	1.59	8.69	8.81
10 — 20	0.38	3.10	2.90	3.90	12.23	9.18	5.91	7.40	6.11	14.86
20 — 30	17.35	4.26	4.76	5.30	5.11	8.38	19.39	15.66	18.03	23.93
30 — 40	20.63	15.58	15.50	20.50	20.63	20.56	19.30	17.85	20.44	28.70
40 — 60	29.00	25.28	24.13	25.96	25.96	26.51	26.14	25.33	25.11	26.58
60 — 100	33.73	16.00	16.31	16.36	16.31	16.50	16.46	16.69	16.61	19.61
Lima series										
0 — 10	1.53	1.95	3.45	3.74	6.26	4.34	5.61	3.83	5.57	7.22
10 — 20	6.62	6.01	12.44	10.93	12.42	15.38	14.74	11.63	14.89	19.00
20 — 30	13.28	14.64	17.82	18.44	20.08	19.93	20.29	16.51	19.10	23.57
30 — 40	17.79	17.49	17.69	19.75	20.32	22.32	21.62	19.74	20.99	24.85
40 — 60	21.89	24.02	24.47	25.72	25.90	26.83	27.92	27.00	27.01	28.82
60 — 100	32.15	32.29	33.08	33.65	33.17	34.50	35.46	35.06	34.29	37.19
Valley bottom slopes										
Dagare series *										
0 — 10	3.10	4.25	5.20	10.38	7.03	-	-	-	-	-
10 — 20	6.05	6.48	6.35	15.80	9.10	-	-	-	-	-
20 — 30	7.75	7.55	7.63	16.15	9.90	-	-	-	-	-
30 — 40	8.05	7.50	7.50	14.23	9.48	-	-	-	-	-
40 — 60	13.33	12.48	12.35	12.35	12.40	-	-	-	-	-
60 — 100	24.05	23.23	23.15	23.10	23.20	-	-	-	-	-

* On Dagare series, SM$_p$ measurements were not completed due to destruction of installed access tube by local residents.

Table S5.5. Continued

Depth intervals	**Measurements**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
cm	**12 – day interval**									
Valley bottom slopes										
Siare series										
0 — 10	1.05	0.80	10.95	13.73	5.93	25.13	21.67	17.50	11.75	2.10
10 — 20	0.35	0.13	14.88	21.03	6.60	30.23	39.48	31.50	23.80	9.43
20 — 30	15.43	16.98	41.83	43.68	18.50	47.10	46.70	43.93	39.53	12.43
30 — 40	27.93	26.83	46.75	50.53	26.83	51.43	51.23	49.80	50.63	20.95
40 — 60	46.80	45.53	59.48	60.75	59.30	62.60	57.53	67.55	62.45	51.43
60 — 100	58.28	58.90	59.00	60.53	59.30	63.10	61.20	60.73	61.70	61.40
Volta series										
0 — 10	1.25	2.26	11.61	7.80	8.52	5.43	6.76	4.48	12.11	10.88
10 — 20	7.10	6.54	15.28	13.84	13.20	11.04	11.14	8.53	14.59	13.36
20 — 30	11.35	10.50	13.68	13.73	14.78	13.70	11.43	11.41	11.63	16.08
30 — 40	14.43	12.98	12.86	13.88	14.00	16.28	15.99	16.30	16.83	18.40
40 — 60	25.68	29.36	29.43	29.68	29.63	29.85	29.81	30.73	29.55	29.69
60 — 100	64.90	54.21	54.24	53.65	54.45	54.08	54.37	59.69	54.45	54.78

Table S5.6. Mean values of temporal soil water storage of the different soil types at standard depth intervals

Depth intervals	Measurements									
	1	2	3	4	5	6	7	8	9	10
cm	12-day interval									
Upland slopes										
Kpelesawgu series										
0 — 10	0.15	0.08	0.65	0.98	4.21	0.75	0.37	0.36	0.85	2.39
10 — 20	10.22	10.72	12.20	23.34	20.93	16.14	14.85	16.08	12.04	26.37
20 — 30	24.41	23.95	23.28	29.41	28.66	31.61	25.64	30.30	25.03	38.53
30 — 40	35.84	34.26	33.26	35.86	36.11	39.28	35.90	36.76	33.97	44.88
40 — 60	88.07	84.76	79.87	85.07	86.66	93.34	89.60	97.48	85.72	91.26
60 — 100	100.73	109.65	109.27	109.27	109.73	110.50	109.88	109.27	108.04	107.96
Middle to lower slopes										
Changnalili series										
0 — 10	0.23	7.97	1.62	0.11	8.53	5.96	7.88	12.61	14.76	26.50
10 — 20	9.25	10.90	10.34	19.08	18.27	17.18	22.01	27.82	20.01	23.01
20 — 30	15.55	16.08	16.35	17.67	17.52	19.16	22.80	29.21	25.81	34.91
30 — 40	33.98	36.18	35.42	38.74	38.82	39.39	41.66	47.06	41.16	39.11
40 — 60	168.29	123.67	109.50	101.40	96.77	96.04	133.95	139.68	137.74	138.74
60 — 100	338.24	309.10	314.23	305.74	309.01	309.43	302.77	315.52	316.09	320.62
Kumayili series										
0 — 10	0.11	0.21	1.66	3.26	5.20	2.49	3.11	2.25	12.27	12.44
10 — 20	0.53	4.38	4.09	5.51	17.26	12.96	8.35	10.45	8.63	20.98
20 — 30	24.50	6.02	6.72	7.48	7.22	11.83	27.38	22.12	25.45	33.79
30 — 40	29.12	21.99	21.89	28.95	29.12	29.03	27.25	25.20	28.86	40.52
40 — 60	81.90	71.38	68.13	73.32	73.32	74.87	73.81	71.52	70.92	75.05
60 — 100	190.48	90.37	92.13	92.42	92.13	93.19	92.98	94.25	93.83	110.77
Lima series										
0 — 10	2.30	2.92	5.18	5.61	9.39	6.50	8.41	5.74	8.35	10.83
10 — 20	9.92	9.02	18.64	16.38	18.62	23.05	22.09	17.43	22.32	28.49
20 — 30	19.91	21.94	26.71	27.65	30.10	29.88	30.41	24.75	28.63	35.34
30 — 40	26.67	26.22	26.52	29.60	30.46	33.46	32.41	29.59	31.47	37.24
40 — 60	65.63	72.02	73.37	77.10	77.64	80.44	83.70	80.96	80.97	86.40
60 — 100	192.77	193.64	198.36	201.74	198.91	206.85	212.63	210.24	205.62	222.96
Valley bottom slopes										
Dagare series *										
0 — 10	4.28	5.87	7.19	14.34	9.71	-	-	-	-	-
10 — 20	8.36	8.95	8.78	21.84	12.58	-	-	-	-	-
20 — 30	10.71	10.43	10.54	22.32	13.68	-	-	-	-	-
30 — 40	11.13	10.37	10.37	19.66	13.09	-	-	-	-	-
40 — 60	36.83	34.48	34.14	34.14	34.27	-	-	-	-	-
60 — 100	132.95	128.39	127.97	127.70	128.25	-	-	-	-	-

* On Dagare series, SM_p measurements were not completed due to destruction of installed access tube by local residents.

Table S5.6. Continued

Depth intervals	Measurements									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
cm	**12-day interval**									
Valley bottom slopes										
Siare series										
0 — 10	1.47	1.12	15.37	19.28	8.32	35.29	30.42	24.57	16.50	2.95
10 — 20	0.49	0.19	20.88	29.53	9.27	42.44	55.42	44.23	33.42	13.24
20 — 30	21.66	23.83	58.72	61.32	25.97	66.13	65.57	61.68	55.49	17.44
30 — 40	39.21	37.66	65.64	70.94	37.66	72.21	71.92	69.92	71.08	29.41
40 — 60	131.41	127.83	167.01	170.59	166.51	175.78	161.55	189.68	175.36	144.42
60 — 100	327.27	330.78	331.34	339.96	333.03	354.37	343.70	341.08	346.51	344.82
Volta series										
0 — 10	1.87	3.39	17.40	11.68	12.76	8.13	10.12	6.70	18.14	16.29
10 — 20	10.64	9.79	22.88	20.73	19.77	16.54	16.68	12.77	21.85	20.01
20 — 30	17.00	15.73	20.50	20.56	22.15	20.52	17.11	17.10	17.41	24.08
30 — 40	21.61	19.44	19.27	20.78	20.97	24.38	23.95	24.42	25.22	27.56
40 — 60	76.92	87.97	88.16	88.91	88.76	89.43	89.32	92.05	88.53	88.94
60 — 100	388.88	324.84	324.99	321.50	326.26	324.07	325.77	357.65	326.26	328.21

Table S6.1. Spatial variability of auxiliary datasets along topographical units

Statistic	TWI	ASP	SLP	SurfR	Elev	LIA
	-	d		-	m	d
Upper slopes (n = 3)						
Mean	11.469	226.491	1.340	0.575	197.555	42.680
SE (±)	0.769	51.369	0.500	0.084	25.935	1.303
Min	10.662	144.786	0.450	0.472	160.812	40.074
Max	13.005	321.285	2.170	0.741	247.634	44.025
Middle to lower slopes (n = 26)						
Mean	12.267	181.302	1.040	0.515	169.175	42.628
SE (±)	0.193	23.603	0.110	0.018	4.410	0.453
Min	11.055	4.498	0.220	0.333	128.890	36.833
Max	14.652	355.989	2.150	0.667	233.931	45.548
Toe slopes (n = 5)						
Mean	13.777	140.421	0.650	0.520	126.596	40.551
SE (±)	0.605	39.385	0.360	0.061	5.749	0.767
Min	12.301	18.586	0.000	0.378	109.484	38.433
Max	15.181	203.991	1.920	0.711	138.290	42.335

LIA = local incidence angle, TWI = topographic wetness index, ASP = aspect, SLP = slope, SurfR = surface roughness, Elev = elevation, d = degree, SE (±) = standard error, n = number of validation locations, Min = minimum and Max = maximum.

Table S7.1. List of a *priori* covariates used in the SWS modelling framework

Covariate ID used in regression matrix	Description of parameter	Spatial Resolution (m)	Resampling Method	Units	Source
Terrain variables					
DevM	Deviation from mean elevation value (surface roughness)	100	-	m	Derived[a]
FloAC	Flow accumulation – Upslope contributing area within the catchment.	100	-	-	Derived[a]
FloCRV	Flow Line Curvature	100	-	degree m^{-1}	Derived[a]
FloDir	Flow direction	100	-	-	Derived[a]
FloPath	Flow path length	100	-	-	Derived[a]
MrRTF	Multi-resolution of ridge top flatness index (MrRTF)	100	-	-	Derived[a]
MrVBF	Multi-resolution Index of Valley Bottom Flatness - Estimating depositional areas	100	-	-	Derived[a]
NegTO	Negative Topographic Openness - the enclosure of a landscape location	100	-	-	Derived[a]
PlaCRV	Plan Curvature - Horizontal (contour) curvature	100	-	degree m^{-1}	Derived[a]
PosTO	Positive Topographic Openness - the dominance of the coordinate of a pixel location	100	-	-	Derived[a]
ProCRV	Profile Curvature - Vertical rate of change of slope	100	-	degree m^{-1}	Derived[a]
SLP	Terrain slope - Inclination of the land surface from the horizontal	100	-	radians	Derived[a]
SRTM	Shuttle Radar Topography Mission (SRTM) 1 arc sec - Digital Elevation Model	30	bilinear	meter	https://earthexplorer.usgs.gov/

Derived[a] are process-based features derived from Shuttle Radar Terrain Mapper-Digital Elevation Model whilst Derived[b] are those from Sentinel-2 images. Covariates that were sourced at > 100 m spatial resolution were resampled to 100 m via a cubic spline algorithm (cubic) and those < 100 m spatial resolution resampled also to 100 m via a bilinear algorithm (bilinear).

Table S7.1. Continued

Covariate ID used in regression matrix	**Description of parameter**	**Spatial Resolution (m)**	**Resampling Method**	**Units**	**Source**
Terrain variables					
TWI	Topographical Wetness Index	100	-	-	Derived[a]
Vdepth	Valley depth - Vertical distance to a channel network base level elevation	100	-	m	Derived[a]
Landcover variables					
BSI	Bare soil index	100	-	-	Derived[b]
CI	Colour index to differentiate soil types of an area	100	-	-	Derived[b]
LULC	10 class landuse and landcover	20	bilinear	-	http://2016africalandcover20m.esrin.esa.int/
MSAVI	Modified Soil adjusted vegetation index	100	bilinear	-	Derived[b]
NDVI	Normalized Difference Vegetation Index (NDVI)	100	-	-	Derived[b]
NDWI2	The second normalized difference water index used to detect surface water and extent	100	-	-	Derived[b]
TCWI	Tasselled cap wetness developed from Tasselled cap coefficients for Sentinel-2 bands (1 – 12 and 8A)	100	-	-	Derived[b]

Derived[a] are process-based features derived from Shuttle Radar Terrain Mapper-Digital Elevation Model whilst Derived[b] are those from Sentinel-2 images. Covariates that were sourced at > 100 m spatial resolution were resampled to 100 m via a cubic spline algorithm (cubic) and those < 100 m spatial resolution resampled also to 100 m via a bilinear algorithm (bilinear).

Table S7.1. Continued

Covariate ID used in regression matrix	Description of parameter	Spatial Resolution (m)	Resampling Method	Units	Source
Climatic variables					
DLST	Mean monthly MODIS Land Surface Temperature - daytime	1000	cubic	Kelvin	https://earthdata.nasa.gov/
ETo	Subsurface soil moisture	1000	cubic	Kg m^{-2}	Google earth engine
NLST	Mean monthly MODIS Land Surface Temperature - nightime	1000	cubic	Kelvin	https://earthdata.nasa.gov/
Prep	Mean monthly precipitation at 1 km February	1000	cubic	mm	https://www.esrl.noaa.gov/psd/data/
Landscape variables					
ASSDT	Average soil and sedimentary deposit thickness	1000	cubic	m	https://daac.ornl.gov
AWC	Available soil water capacity (volumetric fraction) with FC = pF 2.0, 2.3, 2.5	250	cubic	Vol%	https://www.isric.org
BD2	Fine earth bulk density	250	cubic	kg m^{-3}	https://www.isric.org
Bdst	Geographical distances from and to each measurement location	100	-	m	Derived
CLAY	Soil texture fraction – clay content (0-2µm)	250	cubic	%	https://www.isric.org
Cos_Fg	Volumetric coarse fragments	250	cubic	%	https://www.isric.org
LandF	Landform class - Breaks/Foothills-Flat Plains-Hills-Low Hills-Low Mountains-Smooth Plains	250	-	-	CSIR-Soil Research Institute

Derived[a] are process-based features derived from Shuttle Radar Terrain Mapper-Digital Elevation Model whilst Derived[b] are those from Sentinel-2 images. Covariates that were sourced at > 100 m spatial resolution were resampled to 100 m via a cubic spline algorithm (cubic) and those < 100 m spatial resolution resampled also to 100 m via a bilinear algorithm (bilinear).

Table S7.1. Continued

Covariate ID used in regression matrix	Description of parameter	Spatial Resolution (m)	Resampling Method	Units	Source
Landscape variables					
LSM	Soil mapping units at series level	250	cubic	-	CSIR-Soil Research Institute
ORG	Soil organic carbon content (fine earth)	250	cubic	g kg^{-1}	https://www.isric.org
RivDist	River distance map	250	cubic	km	CSIR-Soil Research Institute
SAND2	Sand content (> 5µm)	250	cubic	%	https://www.isric.org
SILT2	Silt content (2-5µm)	250	cubic	%	https://www.isric.org
UVBLSDT	Upland valley bottom and lowland sedimentary deposit thickness	1000	cubic	m	https://daac.ornl.gov
WatCov	Surface water occurrence	250	cubic	-	CSIR-Soil Research Institute
In situ measured variables					
BD	In situ bulk density	point-based	-	g cm^{-3}	Authors
Clay	In situ clay content	point-based	-	%	Authors
Cum	Cumulative Days (from the start of SM measurements)	point-based	-	-	Authors
Depth	Measured standard soil depths	point-based	-	-	Authors
DOY	Day of the year for each measurement	point-based	-	-	Authors
Sand	In situ sand content	point-based	-	%	Authors
Silt	In situ silt content	point-based	-	%	Authors
SM_p	In situ volumetric SM by calibrated PR2/60 capacitance moisture probe.	point-based	-	Vol%	Authors

Derived[a] are process-based features derived from Shuttle Radar Terrain Mapper-Digital Elevation Model whilst Derived[b] are those from Sentinel-2 images. Covariates that were sourced at > 100 m spatial resolution were resampled to 100 m via a cubic spline algorithm (cubic) and those < 100 m spatial resolution resampled also to 100 m via a bilinear algorithm (bilinear).

Table S7.2. Correlation coefficient matrix of spatio-temporal SWS at standard depth intervals

Time-steps	20-Feb	4-Mar	16-Mar	28-Mar	9-Apr	21-Apr	3-May	15-May	27-May
At 0–5 cm									
4-Mar	0.9								
16-Mar	0.83	0.93							
28-Mar	0.96	0.92	0.85						
9-Apr	0.6	0.7	0.72	0.65					
21-Apr	0.81	0.91	0.97	0.84	0.75				
3-May	0.51	0.62	0.6	0.55	0.88	0.63			
15-May	0.49	0.59	0.6	0.54	0.88	0.63	0.89		
27-May	0.56	0.65	0.68	0.6	0.97	0.7	0.92	0.9	
8-Jun	0.32	0.32	0.37	0.31	0.28	0.36	0.31	0.4	0.32
At 5–15 cm									
4-Mar	0.9								
16-Mar	0.83	0.93							
28-Mar	0.96	0.92	0.85						
9-Apr	0.6	0.7	0.72	0.65					
21-Apr	0.81	0.91	0.97	0.84	0.75				
3-May	0.51	0.62	0.6	0.55	0.88	0.63			
15-May	0.49	0.59	0.6	0.54	0.88	0.63	0.89		
27-May	0.56	0.65	0.68	0.6	0.97	0.7	0.92	0.9	
8-Jun	0.32	0.32	0.37	0.31	0.28	0.36	0.31	0.4	0.32
At 15–30 cm									
4-Mar	0.94								
16-Mar	0.92	0.96							
28-Mar	0.99	0.95	0.93						
9-Apr	0.87	0.9	0.94	0.88					
21-Apr	0.92	0.94	0.99	0.93	0.95				
3-May	0.82	0.9	0.89	0.84	0.95	0.89			
15-May	0.81	0.86	0.86	0.82	0.94	0.87	0.95		
27-May	0.85	0.89	0.92	0.86	0.99	0.93	0.97	0.95	
8-Jun	0.47	0.5	0.54	0.51	0.52	0.57	0.53	0.61	0.52
At 30–40 cm									
4-Mar	0.98								
16-Mar	0.98	0.99							
28-Mar	1	0.98	0.98						
9-Apr	0.96	0.97	0.98	0.96					
21-Apr	0.98	0.99	1	0.98	0.98				
3-May	0.95	0.97	0.97	0.95	0.99	0.97			
15-May	0.95	0.96	0.96	0.95	0.98	0.96	0.98		
27-May	0.95	0.97	0.98	0.96	1	0.98	0.99	0.98	
8-Jun	0.79	0.81	0.81	0.8	0.81	0.81	0.81	0.85	0.81

All matrix comparisons were significant at $\rho < 0.01$. n = 2e+07.

Table S7.2. Continued

Time-steps	20-Feb	4-Mar	16-Mar	28-Mar	9-Apr	21-Apr	3-May	15-May	27-May
At 40–60 cm									
4-Mar	0.97								
16-Mar	0.95	0.99							
28-Mar	1	0.97	0.96						
9-Apr	0.93	0.95	0.95	0.93					
21-Apr	0.96	0.98	0.99	0.96	0.96				
3-May	0.92	0.93	0.93	0.92	0.97	0.94			
15-May	0.93	0.93	0.93	0.93	0.97	0.93	0.97		
27-May	0.93	0.95	0.95	0.93	0.99	0.95	0.98	0.97	
8-Jun	0.81	0.82	0.81	0.81	0.82	0.81	0.8	0.85	0.82
At 60–100 cm									
4-Mar	1								
16-Mar	0.99	0.99							
28-Mar	1	1	0.99						
9-Apr	0.98	0.99	0.99	0.98					
21-Apr	0.99	0.99	1	0.99	0.99				
3-May	0.95	0.94	0.94	0.95	0.96	0.93			
15-May	0.95	0.94	0.93	0.95	0.95	0.93	1		
27-May	0.95	0.95	0.94	0.95	0.97	0.94	1	0.99	
8-Jun	0.93	0.92	0.91	0.93	0.93	0.9	0.98	0.98	0.98

All matrix comparisons were significant at $\rho < 0.01$. n = 2e+07.

Table S7.3. Overview of descriptive statistics of spatio-temporal SWS at standard depth intervals

Time-steps	Mean	SE (±)	St. Dev	Var	CV	Min	Max	Median	CI (95%)
At 0–5 cm									
Feb 20	1.586	0.000	1.544	2.385	0.974	0.006	10.904	1.479	0.001
Mar 04	3.267	0.000	2.173	4.720	0.665	0.006	15.397	2.986	0.001
Mar 16	3.433	0.001	3.250	10.560	0.947	0.006	17.383	2.493	0.001
Mar 28	1.743	0.000	1.606	2.580	0.922	0.006	12.369	1.493	0.001
Apr 09	4.766	0.000	1.994	3.976	0.418	0.006	19.328	4.479	0.001
Apr 21	6.159	0.001	3.158	9.971	0.513	0.006	19.356	4.986	0.001
May 03	3.567	0.000	2.192	4.807	0.615	0.006	18.808	3.000	0.001
May 15	5.313	0.000	2.114	4.468	0.398	0.006	18.835	5.000	0.001
May 27	3.587	0.000	2.031	4.126	0.566	0.006	16.863	3.014	0.001
Jun 08	7.160	0.001	2.931	8.593	0.409	0.006	23.808	6.521	0.001
At 5–15 cm									
Feb 20	1.586	0.000	1.544	2.385	0.974	0.006	10.904	1.479	0.001
Mar 04	3.267	0.000	2.173	4.720	0.665	0.006	15.397	2.986	0.001
Mar 16	3.433	0.001	3.250	10.560	0.947	0.006	17.383	2.493	0.001
Mar 28	1.743	0.000	1.606	2.580	0.922	0.006	12.369	1.493	0.001
Apr 09	4.766	0.000	1.994	3.976	0.418	0.006	19.328	4.479	0.001
Apr 21	6.159	0.001	3.158	9.971	0.513	0.006	19.356	4.986	0.001
May 03	3.567	0.000	2.192	4.807	0.615	0.006	18.808	3.000	0.001
May 15	5.313	0.000	2.114	4.468	0.398	0.006	18.835	5.000	0.001
May 27	3.587	0.000	2.031	4.126	0.566	0.006	16.863	3.014	0.001
Jun 08	7.160	0.001	2.931	8.593	0.409	0.006	23.808	6.521	0.001
At 15–30 cm									
Feb 20	13.245	0.001	3.530	12.458	0.266	0.006	29.835	12.479	0.002
Mar 04	19.709	0.001	3.893	15.158	0.198	6.576	37.863	18.507	0.002
Mar 16	22.992	0.001	5.662	32.057	0.246	9.562	43.918	20.986	0.002
Mar 28	14.138	0.001	3.513	12.341	0.248	0.110	31.328	13.466	0.002
Apr 09	25.052	0.001	4.325	18.708	0.173	12.041	47.356	23.945	0.002
Apr 21	27.264	0.001	5.416	29.335	0.199	14.041	47.397	25.479	0.002
May 03	22.158	0.001	4.050	16.402	0.183	9.055	43.342	20.959	0.002
May 15	24.885	0.001	3.937	15.504	0.158	12.562	45.780	23.973	0.002
May 27	22.313	0.001	4.127	17.030	0.185	10.548	43.369	20.986	0.002
Jun 08	33.702	0.001	3.630	13.174	0.108	17.165	53.753	33.000	0.002

SE (±) = standard error, St. Dev = standard deviation, Var = variance, CV = coefficient of variation of spatial SWS, Min = minimum, Max = maximum and CI = confidence interval at 95% around the mean. n = 2e+07.

Table S7.3. Continued

Time-steps	Mean	SE (±)	St. Dev	Var	CV	Min	Max	Median	CI (95%)
At 30–40 cm									
Feb 20	41.012	0.001	6.499	42.237	0.158	15.220	76.876	38.973	0.003
Mar 04	44.689	0.002	7.341	53.892	0.164	18.617	93.397	42.438	0.003
Mar 16	50.627	0.002	9.407	88.492	0.186	24.589	105.383	47.411	0.004
Mar 28	41.987	0.001	6.688	44.730	0.159	17.192	79.876	39.945	0.003
Apr 09	55.876	0.002	8.013	64.208	0.143	30.562	105.918	53.397	0.003
Apr 21	51.786	0.002	9.241	85.398	0.178	25.096	103.835	48.424	0.004
May 03	52.470	0.002	7.552	57.032	0.144	27.069	100.397	49.945	0.003
May 15	55.813	0.002	7.430	55.204	0.133	29.082	101.411	53.904	0.003
May 27	52.670	0.002	7.974	63.582	0.151	28.576	103.918	49.986	0.003
Jun 08	67.498	0.001	6.081	36.984	0.090	37.165	97.739	66.424	0.003
At 40–60 cm									
Feb 20	25.693	0.001	5.995	35.938	0.233	4.644	61.684	23.973	0.003
Mar 04	21.864	0.001	5.889	34.684	0.269	2.631	52.794	20.055	0.003
Mar 16	20.692	0.001	6.756	45.638	0.326	1.151	49.315	18.521	0.003
Mar 28	24.791	0.001	5.941	35.293	0.240	4.631	60.191	23.000	0.003
Apr 09	25.079	0.001	5.323	28.338	0.212	6.576	51.808	23.493	0.002
Apr 21	20.724	0.001	6.777	45.924	0.327	1.617	47.342	18.493	0.003
May 03	25.676	0.001	5.589	31.241	0.218	6.096	51.808	24.452	0.002
May 15	26.485	0.001	5.963	35.558	0.225	6.096	57.780	24.986	0.003
May 27	24.845	0.001	5.672	32.169	0.228	6.603	52.794	23.466	0.002
Jun 08	29.237	0.001	5.751	33.071	0.197	6.110	61.739	27.973	0.002
At 60–100 cm									
Feb 20	121.135	0.004	17.295	299.103	0.143	56.151	223.971	114.890	0.007
Mar 04	117.737	0.004	16.813	282.675	0.143	53.672	217.957	111.631	0.007
Mar 16	117.532	0.004	17.442	304.233	0.148	51.206	217.930	111.424	0.008
Mar 28	120.225	0.004	17.293	299.048	0.144	55.165	222.971	113.904	0.007
Apr 09	122.569	0.004	16.669	277.866	0.136	61.069	221.464	116.890	0.007
Apr 21	118.112	0.004	17.443	304.257	0.148	50.699	216.916	111.931	0.008
May 03	129.063	0.004	19.118	365.515	0.148	61.576	250.875	123.959	0.008
May 15	127.719	0.004	19.397	376.245	0.152	60.589	251.368	122.411	0.008
May 27	126.423	0.004	18.595	345.777	0.147	62.069	245.861	121.014	0.008
Jun 08	131.855	0.004	18.680	348.937	0.142	63.082	257.313	126.521	0.008

SE (±) = standard error, St. Dev = standard deviation, Var = variance, CV = coefficient of variation of spatial SWS, Min = minimum, Max = maximum and CI = confidence interval at 95% around the mean. n = 2e+07.

Table S8.1. Site characteristics and GPS coordinates for all sampling locations from the Guinea savannah zone of Ghana

Station ID	Latitude	Longitude	Soil type	Soil association	Geology	District block
SM and physical soil properties sites						
AT01	9.38209	-0.68264	Lima	Sambu-Pasga	Shale, Mudstone, Sandstone	Mion
AT02	9.38980	-1.02133	Kpelesawgu	Sambu-Pasga	Shale, Mudstone, Sandstone	Tolon
AT03	9.24358	-0.62165	Changnalili	Lima-Volta	Alluvium sediments	Karaga
AT04	9.30885	-0.71828	Kpelesawgu	Techiman-Tampu	Voltain sandstone	Tamale Metro
AT05	9.40523	-1.23727	Changnalili	Kpelesawgu-Changnalili	Voltain shale	Tolon
AT06	9.55798	-0.96041	Lima	Lima-Volta	Alluvium sediments	Kumbungu
AT07	9.55211	-1.17127	Lima	Sambu-Pasga	Shale, Mudstone, Sandstone	Tolon
AT08	9.34742	-0.75396	Lima	Techiman-Tampu	Voltain sandstone	Tamale Metro
AT09	9.25980	-0.72064	Lima	Lima-Volta	Alluvium sediments	East Gonja
AT10	9.45722	-1.29907	Lima	Kpelesawgu-Changnalili	Voltain shale	North Gonja
AT11	9.63135	-1.18874	Lima	Kpelesawgu-Changnalili	Voltain shale	Kumbungu
AT12	9.39602	-0.48972	Kumayili	Techiman-Tampu	Voltain sandstone	Tamale Metro
AT13	9.41191	-0.23344	Changnalili	Gushiagu-kasele	Voltain shale	Karaga
AT15	9.32232	-0.81182	Lima	Lima- Volta	Alluvium sediments	East Gonja
AT16	9.08735	-1.13139	Siare	Siare-dagare	Alluvium sediments	Central Gonja
AT17	9.80794	-0.42220	Lima	Kpelesawgu-Changnalili	Voltain shale	Karaga
AT18	9.08612	-1.03072	Volta	Kpelesawgu-Changnalili	Voltain shale	Central Gonja
AT19	9.57707	-0.77180	Lima	Techiman-Tampu	Voltain sandstone	Savelugu Nanton
AT20	9.31643	-0.25106	Lima	Sambu-Pasga	Shale, Mudstone, Sandstone	Mion
AT22	9.75427	-0.82686	Volta	Lima-Volta	Alluvium sediments	Kumbungu
AT23	9.51278	-0.75376	Lima	Lima-Volta	Alluvium sediments	Savelugu Nanton
AT24	9.47611	-1.13107	Kumayili	Sambu-Pasga	Shale, Mudstone, Sandstone	Savelugu Nanton
AT25	9.38885	-0.27060	Kpelesawgu	Techiman-Tampu	Voltain sandstone	Tamale Metro
AT26	9.53645	-1.37973	Dagare	Sambu-Pasga	Shale, Mudstone, Sandstone	North Gonja
AT27	9.52581	-0.92455	Lima	Lima-Volta	Alluvium sediments	Kumbungu

Table S8.1. Continued

Station ID	Latitude	Longitude	Soil type	Soil association	Geology	District block
AT28	9.73252	-0.47761	Lima	Sambu-Pasga	Shale, Mudstone, Sandstone	Karaga
AT29	9.15954	-1.42628	Lima	Techiman-Tampu	Voltain sandstone	Central Gonja
AT30	9.40799	-0.45041	Lima	Lima-Volta	Alluvium sediments	Mion
AT31	9.07524	-0.54386	Lima	Sambu-Pasga	Shale, Mudstone, Sandstone	East Gonja
AT32	9.32838	-0.93585	Lima	Lima-Volta	Alluvium sediments	Central Gonja
AT33	9.60072	-0.84122	Kumayili	Mimi-Techiman	Voltain sandstone	North Gonja
AT34	9.59978	-0.63474	Lima	Kpelesawgu-Changnalili	Voltain shale	Savelugu Nanton
AT35	9.38973	-0.33578	Lima	Lima-Volta	Alluvium sediments	Mion
AT36	9.49393	-0.69697	Lima	Lima-Volta	Alluvium sediments	Karaga
AT37	9.47358	-0.84833	Lima	Techiman-Tampu	Voltain sandstone	Sagnerigu
AT38	9.20374	-0.97482	Volta	Kpelesawgu-Changnalili	Voltain shale	Central Gonja
Plant available water content sites						
Wet 1	9.65884	-0.57731	Lima	Lima-Volta	Alluvium sediments	Karaga
Wet 2	9.40714	-0.98608	Kpelesawgu	Sambu-Pasga	Shale, Mudstone, Sandstone	Tolon

Figures

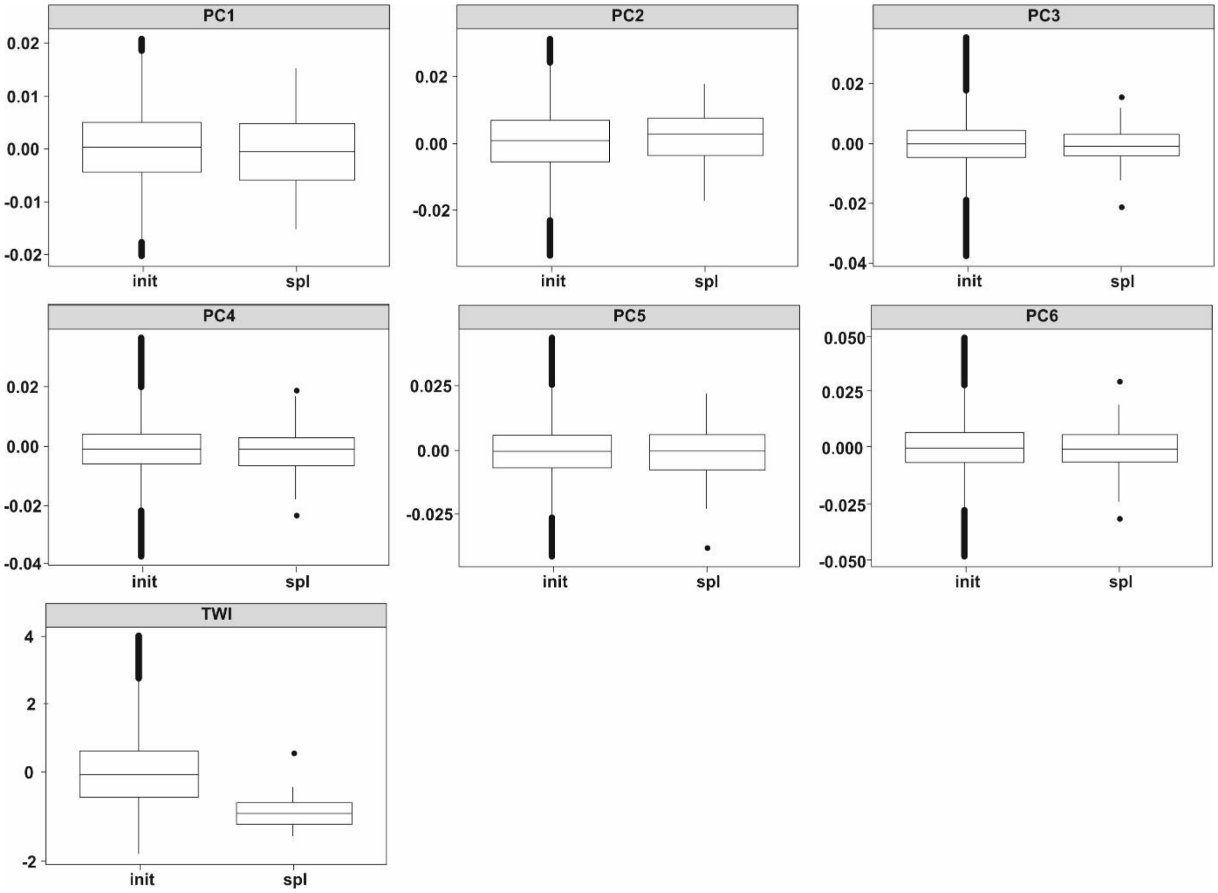


Figure S4.1. Distribution of subsamples (spl) drawn from the selected PCs of the GWPCA (init) used in the cLHC simulating annealing process. TWI = SAGA topographic wetness index.

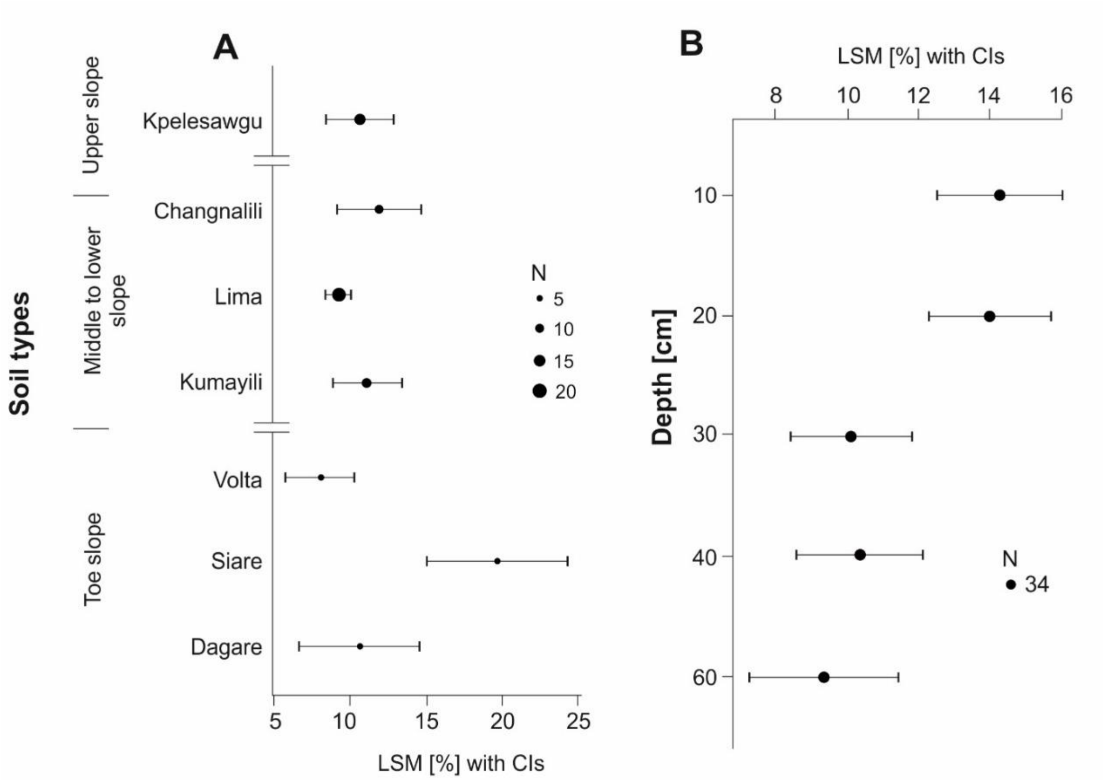


Figure S5.1. Modelled least square means (LSM) and 95% confidence intervals (CI) for gravimetric soil moisture (SM_g) and physical soil properties per soil type (A) and at different standard soil depths (B). N = number of samples used per groups of measurements.

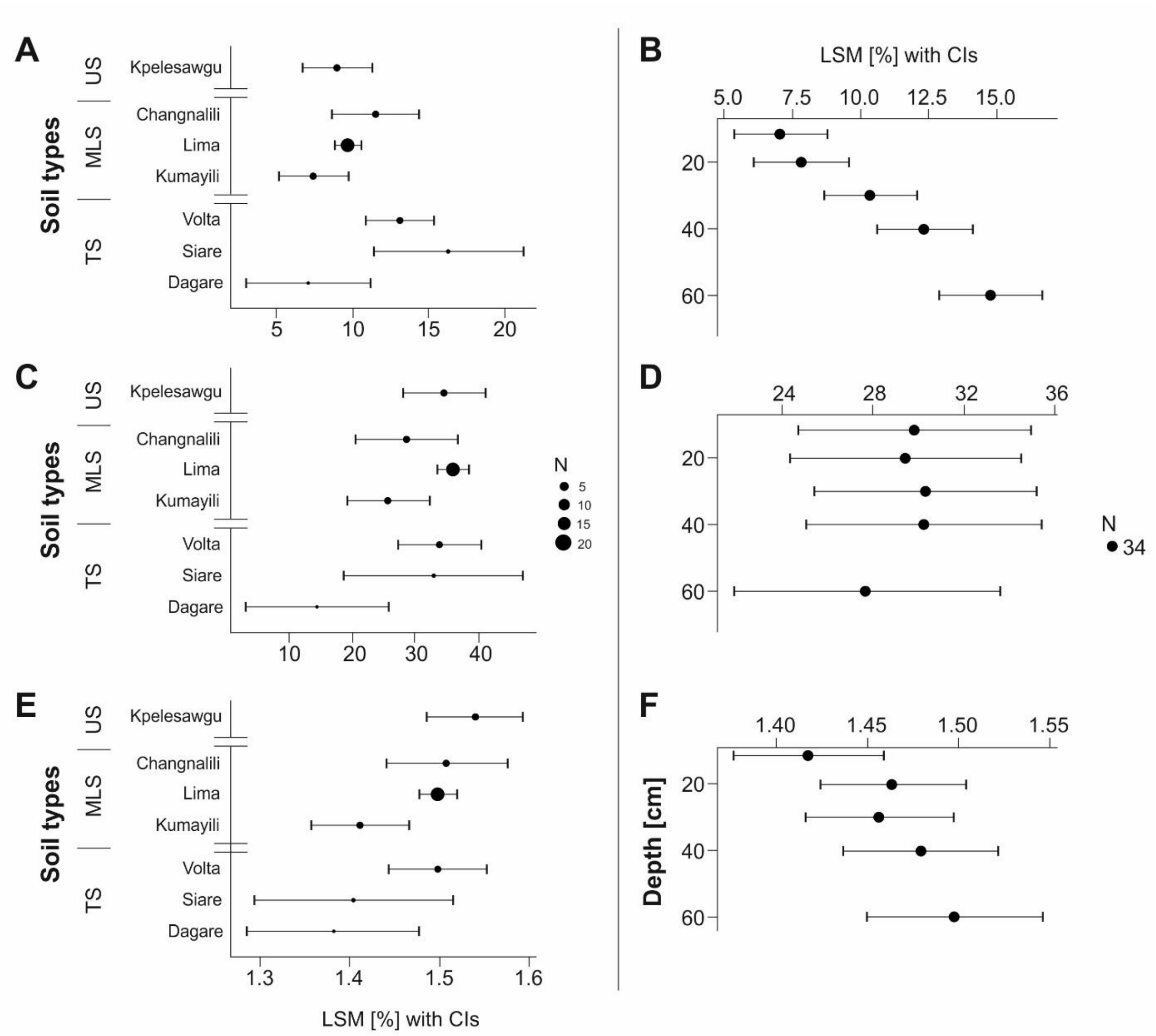


Figure S5.2. Modelled least square means (LSM) and 95% confidence intervals (CI) for physical soil properties and gravimetric soil moisture per soil type (A – clay content, C – silt content and E – bulk density) and at different standard soil depths (B – clay content, D – silt content and F – bulk density), N = number of samples used per groups of measurements. TS = toe slope, MLS = middle to lower slope and UP = upper slope.

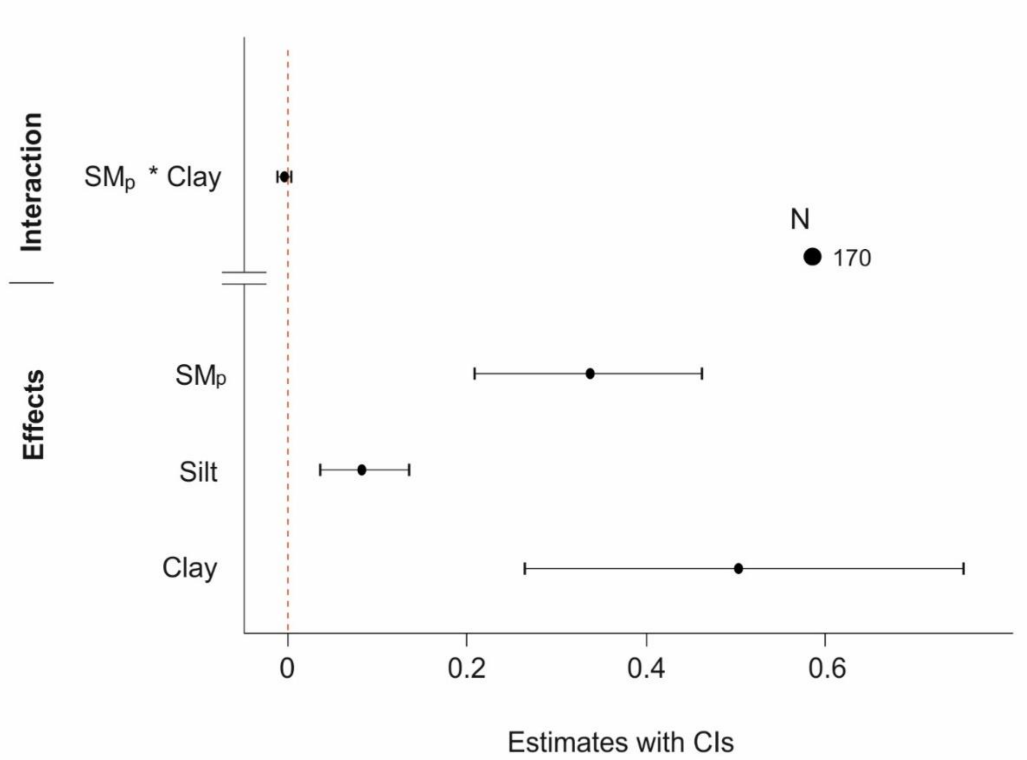


Figure S5.3. Estimates and 95% confidence intervals (CI) for fixed and random effects of the mixed effect models used to explain the relationship between gravimetric (SM_g) and PR2 profile probe (SM_p) soil moisture measurements.

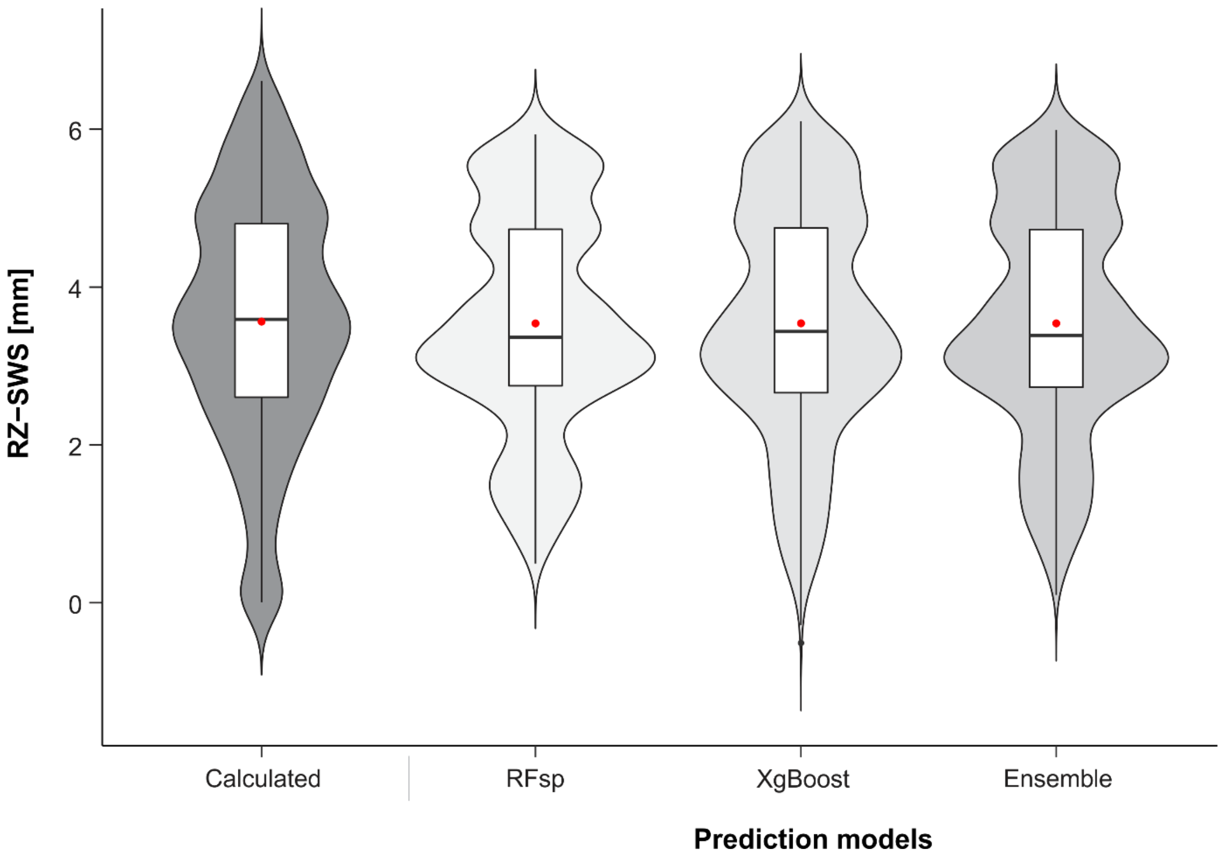


Figure S7.1. Differences in model performances for predicted SWS from the *target-oriented* LLTO-CV. Arithmetic means (red dots), ranges (length of whiskers), medians (horizontal bars) and their corresponding distribution of predicted SWS are presented.

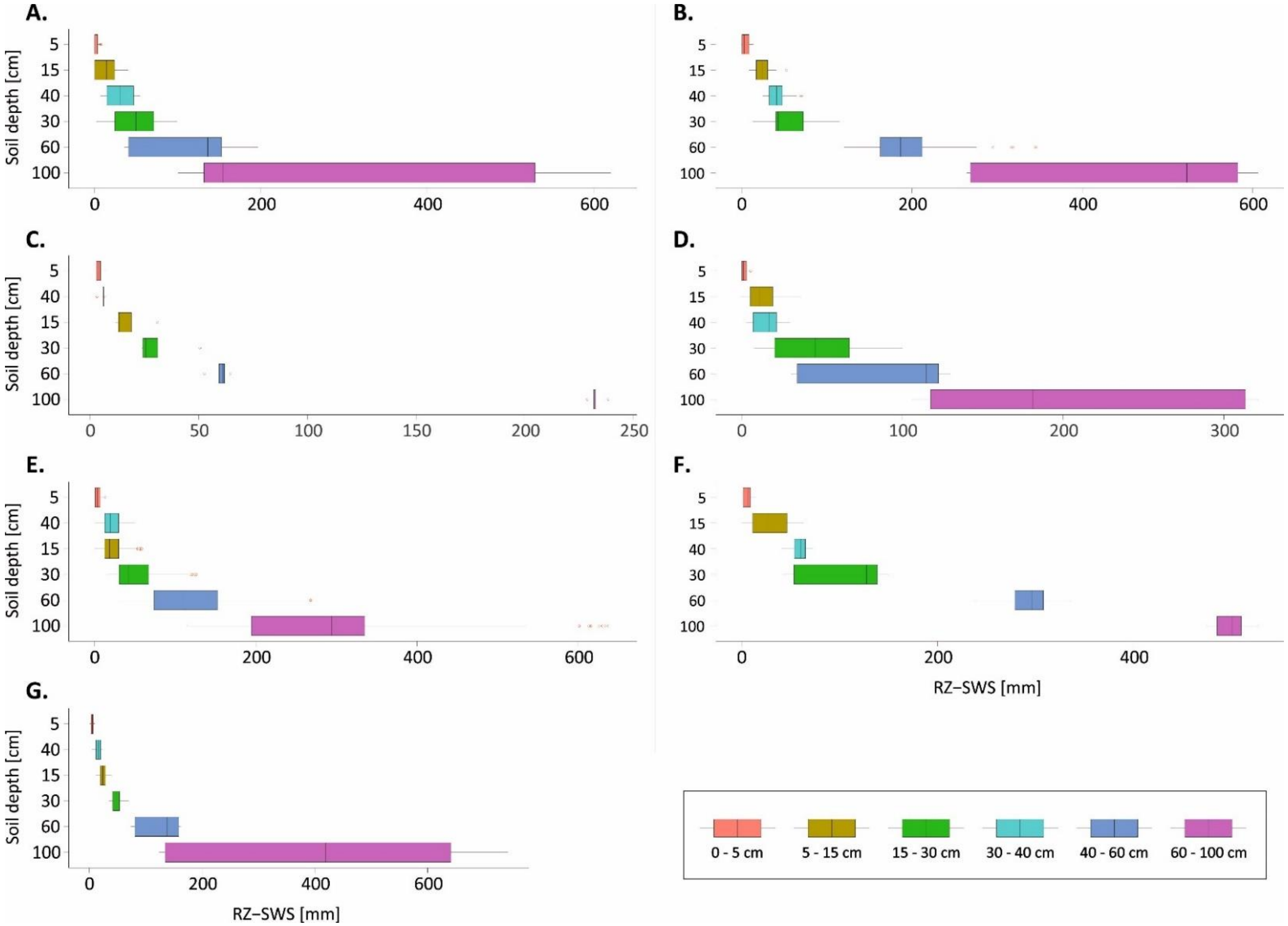


Figure S9.1. Spatio-temporal variability of SWS for (A) Kpelesawgu series, (B) Changnalili series, (C) Dagare series, (D) Kumayili series, (E) Lima series, (F) Siare series and (G) Volta series along the various in situ measurement depths. Soil names are in Ghanaian soil classification system. Statistical analysis focussed on the spatio-temporal variability of SWS. Ranges (length of whiskers) and medians (vertical bars) are shown.