A near-linear transect sampling method for validation of SMAP surface soil moisture in non-complex terrains

Cheruiyot, Elijah; Mito, Collins; Menenti, Massimo

DOI
10.1080/01431161.2024.2388866

Publication date
2024

Document Version
Final published version

Published in
International Journal of Remote Sensing

INTERNATIONAL JOURNAL OF REMOTE SENSING
2024, VOL. 45, NO. 18, 6285–6302
https://doi.org/10.1080/01431161.2024.2388866

A near-linear transect sampling method for validation of SMAP surface soil moisture in non-complex terrains

Elijah Cheruiyot [a], Collins Mito [a] and Massimo Menenti [b]

[a]Department of Physics, University of Nairobi, Nairobi, Kenya; [b]Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands

ABSTRACT

The global validation for SMAP (Soil Moisture Active Passive) Level-4 surface soil moisture using well-established core validation sites does not comprehensively account for all landscapes on earth's surface due to the diverse nature of their intrinsic characteristics. Due to the inhibitive cost of implementing a standard validation site, this study presents an alternative sampling approach suitable for localized validation of SMAP data in non-complex terrains. It involves clustering a large heterogeneous study area to smaller units of non-complex terrains where landscape-defining characteristics are largely homogeneous, thus permitting the computation of areal soil moisture as a simple arithmetic mean of near-linear point measurements. This allows optimization of limited resources (a hand-held moisture sensor with site-specific calibrations) to balance the need for spatial representativeness of samples in the sampling unit and the need for temporal proximity of sampled measurements to the aggregation time of the satellite product being validated. For ease of movement, transect sampling is implemented along access roads that run across the sampling units to allow sufficient measurement replications within a reasonably short time. Validations with four different landscapes in Kenya show a good agreement between *in situ* measurements and SMAP with R^2 of 0.76, 0.72, 0.80, and 0.82, and biases of −0.0246, +0.0113, 0.0004, and + 0.0035 m^3 m^{-3}, respectively for Mawego, Kuresoi, Sotik and Kapsuser sites. These results only marginally differ from those obtained with a spatially distributed sampling method, indicating the potential of the proposed sampling design for time and cost effectiveness in validations at non-complex terrains. An analysis of the temporal variability of SMAP soil moisture in the watershed is also presented, with an assessment of its significance in the selection of sampling sites for validation. Particularly, the concept of temporal stability of soil moisture as a basis for clustering validation sites is evaluated.

ARTICLE HISTORY
Received 19 February 2024
Accepted 24 July 2024

1. Introduction

Soil moisture is a state variable that is highly dynamic both in space and time (Narasimhan et al. 2005), so that the most practical way to measure it on a large scale is by satellite

CONTACT Elijah Cheruiyot ekcheru@gmail.com Department of Physics, University of Nairobi, P.O. Box 30197, Nairobi 00100, Kenya

remote sensing. To this end, significant technological progress has been made to provide such satellite-based data, albeit at coarse spatial resolutions, and have found relevance in a wide range of applications. Remotely sensed soil moisture has found particular interest in hydrological modelling since soil water controls many water balance variables such as surface runoff, deep drainage and the rate of evapotranspiration (Melesse et al. 2007; Wang and Qu 2009). Many open access satellite-based soil moisture products are currently available, and they present a vital resource that allows operationalization of systems such as hydrological, agricultural, drought or flood monitoring, etc., at a large scale.

But because of heterogeneous land surfaces, remotely sensed soil moisture estimates suffer inevitable biases. Land surface-related factors affecting satellite remote sensing retrieval of soil moisture include soil texture, vegetation density and topography (Choudhury et al. 1979; Jackson, Schmugge, and Wang 1982; Konings et al. 2017; Mohanty 2013; Ulaby, Bradley, and Dobson 1979; Wang et al. 1983). Thus, satellite-based soil moisture data must be validated at regional and local level to account for surface heterogeneity. The estimation error induced by surface heterogeneity on retrieved soil moisture is a bulk behaviour resulting from a complex combination of these factors, so that the most practical way to correct for them is to make a synoptic evaluation of the biases at a landscape scale rather than a small-scale parametrization of each factor.

Colliander et al. (2017) report a comprehensive campaign to validate SMAP surface soil moisture products, in which several landscapes featuring different land cover types, climatic regimes, etc., were considered. The report describes an effort undertaken to ensure geographic distribution and diversity of site conditions. However, it is not realistically feasible to account for all possible landscape scenarios on earth's land surface. Owing to the diverse nature of the salient features that define landscapes, each landscape is unique so that the generic landscapes sampled for large-scale (global) validation of satellite products are only representative of the world's major landscapes but are neither exhaustive nor can precisely match any other landscape outside of the specific ones that were sampled. Kolassa et al. (2017) note that global validation and bias correction approaches may be more robust compared to those at local scales, but they are more vulnerable to retrieval errors due to uncertainties in the retrieval inputs. The quality of soil moisture satellite products validated in such a generic manner suffices for most applications, but in some applications such as surface runoff modelling where soil moisture is a key initialization variable (Shahrban et al. 2018), a more precise landscape-adapted validation may be required. Antecedent soil moisture has a direct influence on quantification of surface runoff, as it determines whether or not precipitation is sufficient to trigger a surface runoff (Meißl, Zieher, and Geitner 2020; Song and Wang 2019), and remotely sensed soil moisture has proved the feasibility and rationality to provide this information (Crow and Ryu 2009; Jiang and Wang 2019; Kofidou and Gemitzi 2023; Minet et al. 2011; Yu et al. 2018).

While landscape-specific validation of satellite soil moisture products is desirable, the elaborate infrastructure for collecting ground measurements required to implement it can be a major impediment. The SMAP validation campaign employed a broad network of core validation sites (CVS), which meet certain requirements in terms of sensor calibrations and representativeness of sample locations within the pixel (Colliander et al. 2017).

Few regions in the world, however, have sampling sites that meet the requirements for a CVS; it is costly to establish and maintain them (Lakshmi 2013). Myeni, Moeletsi, and Clulow (2019) highlighted a mismatch between the highly diverse climatological, biogeographical, pedological and lithological characteristics of landscapes over the African continent, and the disproportionately poor soil moisture ground sampling infrastructure that precludes any comprehensive validation of satellite-based soil moisture data in the region. This limits the operational application of SMAP, as many regions are forced to rely on validations carried out in unmatching environments. Despite the efforts made to instal soil moisture monitoring networks in Africa in the past two decades (Myeni, Moeletsi, and Clulow 2019), there remains a data gap that can only be filled by less conventional sampling approaches.

This study proposes an alternative sampling approach that optimizes limited resources (a single sensor) to collect the required *in situ* soil moisture data in order to fill the data gap left by the standard validation method for SMAP, without significant compromise on the need for spatial representativeness of samples in the sampling site and the need for temporal proximity of sampled measurements to the satellite product being validated.

2. Materials and method

2.1. Study area

The validation of SMAP surface soil moisture product presented in this study was carried out in four sites within Sondu-Miriu watershed in Western Kenya, shown in Figure 1. The watershed was defined using the Hydrology tool of ArcGIS 10.3 to generate flow channels and define watershed boundaries, with SRTM digital elevation model as input parameter.

Sondu-Miriu watershed extends over an area of about 3500 square kilometres, and together with Lake Victoria, the watershed forms a complex ecosystem. It is one of the major watersheds feeding the Kenyan side of the trinational fresh water lake. Growth in

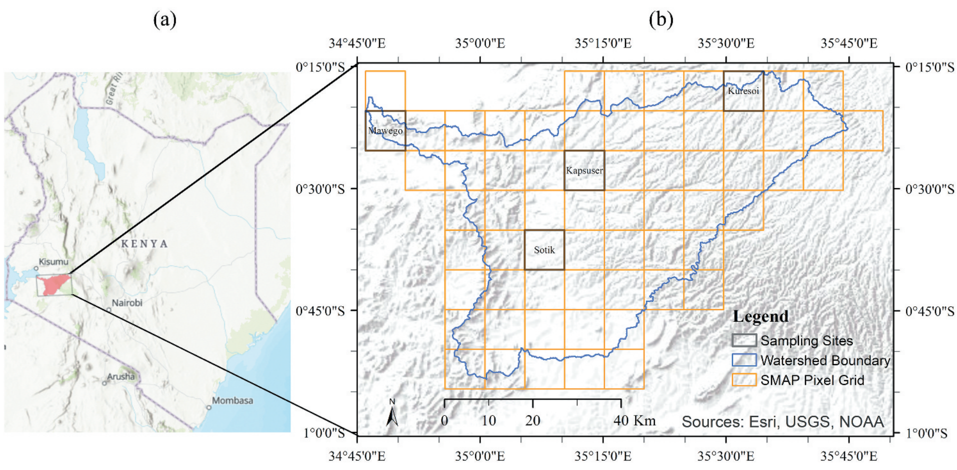

Figure 1. Map showing the geographic location of Sondu-Miriu watershed in Western Kenya (a), and the four soil moisture sampling sites within the watershed (b). The grid of SMAP soil moisture pixels that overlap the watershed is shown.

the riparian population has led to an increase in anthropogenic activities such as agriculture (Olokotum et al. 2020), and surface runoff has the potential to sweep nutrients into the lake, exacerbating the periodic occurrences of explosive growth of invasive aquatic plants (Cheruiyot et al. 2014). Hypothetically, a hydrological simulation of surface runoff in the watershed could help link the extraordinary phenomena of periodic explosions of invasive aquatic weed in the lake to its apparent triggers, and remotely sensed soil moisture measurements would be a valuable resource for this purpose.

2.2. SMAP soil moisture

SMAP is a set of earth observation products derived from NASA's (National Aeronautics and Space Administration) environmental monitoring satellite (SMAP), that was launched in January 2015. SMAP's instruments operate in the L-band (1.26 GHz and 1.41 GHz respectively for radar and radiometer), which permits observations of soil moisture through clouds and moderate vegetation cover, as these layers are nearly transparent at the microwave frequencies (Monerris et al. 2009). SMAP soil moisture product was originally planned to be a combined product of SMAP's radar (3 km) and radiometric (36 km) measurements to produce an intermediate resolution (9 km) product (Colliander et al. 2017), but the radar component failed in July 2015, barely three months into operation. SMAP mission objective is to retrieve surface soil moisture for the top 5 cm of the soil profile with estimation errors no larger than 0.04 cm^3 cm^{-3} (Colliander et al. 2017).

SMAP Level-4 Surface Soil Moisture (Reichle et al. 2022) is an open access soil moisture simulated product available on a 9 km grid and simulation intervals of 3 hours, making it an important resource for near-real time study of hydrological processes at a regional scale. It is a value-added product generated by a land data assimilation system that utilizes L-band brightness temperature observations from SMAP satellite, precipitation observations, and land surface modelling (Reichle et al. 2017); production of SMAP soil moisture is therefore not limited by the radar data availability (Colliander et al. 2017). The land model is based on conservation principles of water and energy, which together with realistic forcing data ensures reliability of the simulated product that is further refined by satellite observations to provide fairly reliable moisture estimates compared to other space-borne moisture products (Duygu and Akyürek 2019; Reichle et al. 2017).

The current study targets validation of SMAP Level-4 'SPL4SMGP sm_surface' product at two simulation time intervals: 09:00–12:00 hours (average-centred at 10:30 hours) and 15:00–18:00 hours (average-centred at 16:30 hours).

2.3. Sampling design and measurement

According to Colliander et al. (2017), the basic requirements for ground soil moisture point measurements averaged to obtain an areal estimate value of a SMAP grid pixel are (1) the ground sensor must be location-specific calibrated, and (2) there is an adequate number of ground measurement replications with a representative distribution within the pixel grid. Meeting the latter requirement would conventionally require simultaneous point measurements from an extensive network of sensors representatively distributed within the pixel grid as depicted in Figure 2(a). These sensors are normally installed on the ground for extended periods of time and continuously

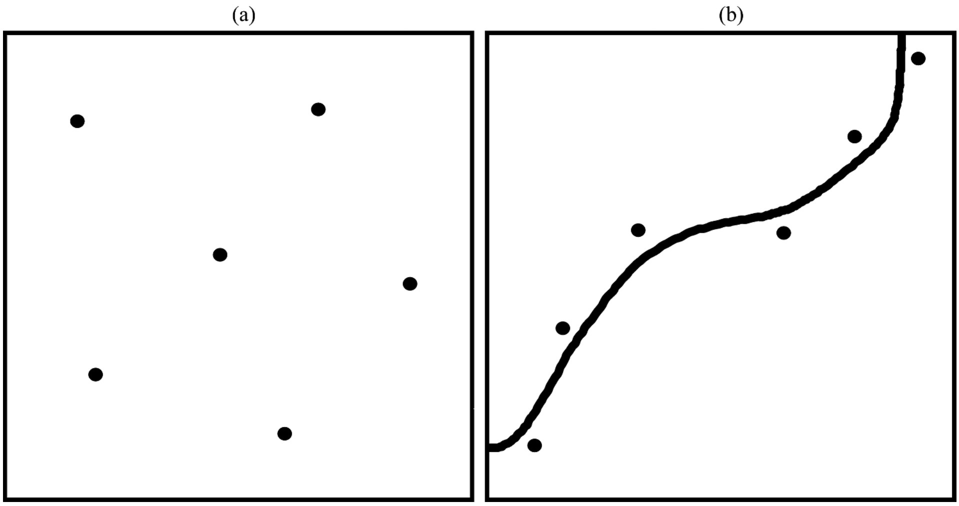

Figure 2. A comparison of two sampling designs: a conventional sampling design (a) with installed sensors that simultaneously take measurements at spatially distributed sample points to enhance spatial representativeness, and the near-linear transect sampling design (b) that allows low-cost sampling of soil moisture through the use of a single sensor along access roads to achieve multiple replications in a short time.

collect data at regular intervals, with or without automated data transmission capabilities. The distribution of sensors is designed to capture the spatial heterogeneity of the sampling unit, and therefore enhance their representativeness. The point measurements can be averaged to obtain an area representative soil moisture value of the sampling unit, or interpolated to obtain a spatially continuous distribution of the state variable. But this infrastructure does not presently exist in the study area and it is costly to instal one.

In the absence of the conventional sampling design for validation of SMAP, the idea is to implement an alternative sampling strategy, shown in Figure 2(b), that allows gathering of quality ground truth soil moisture data at low cost. To attain sufficient measurement replications using a single sensor, measurements are collected along access roads that allow ease of movements. But the resultant sample points may follow a near-linear pattern, that is known to be less representative of the sampling unit. This is mitigated by identifying close to homogeneous pixels, within which near-linear sample points may be representative enough. To achieve this, the following procedure was followed: 1) Evaluate the spatial variability of soil moisture in the study area with respect to the spatial scale of the remotely sensed soil moisture data; 2) Identify the key driving factors of the soil moisture spatial heterogeneity, and use that information to cluster pixels with similar characteristics to form regions within which validation is to be carried out; 3) From each cluster, select a representative pixel with an access road that cuts across a substantial pixel area, from which sampling data is to be collected; 4) Perform site-specific soil moisture meter calibration for each cluster; and 5) Implement a near-linear transect sampling approach to collect soil moisture data along the access roads within the sampling pixels.

2.3.1. Spatial variability of SMAP surface soil moisture

Spatial variability was evaluated by defining a linear transect featuring 13 SMAP pixels that cut across the study area from downstream where the Sondu-Miriu River drains to Lake Victoria, to upstream which is the main source of the river's water, as shown in Figure 3(a). For a specific SMAP dataset, the change of soil moisture across space was determined as the difference between the moisture values of adjacent pixels. This resulted in 12 pixel-to-pixel transitions per dataset, and was repeated for 23,376 SMAP datasets representing a period covering the first eight complete years of the product from 31 March 2015 01:30:00 to 30 March 2023 22:30:00, at three hours intervals, as presented in Figure 3(b).

There is a pixel-to-pixel variation of SMAP soil moisture that range from −0.05 to +0.15 $m^3\ m^{-3}$. This highlights the sharp change in soil moisture values between adjacent SMAP pixels, which is an indication that the pixel-to-pixel landscape-defining characteristics vary so significantly across the study area that they cannot be validated as a block. The study area must therefore be broken down to sub-regions with similar characteristics.

2.3.2. Clustering of SMAP pixels

Soil texture, vegetation density and topography, which are the land surface factors with greater influence on satellite remote sensing retrieval of soil moisture (Mohanty 2013), were considered in characterizing pixels. Soil texture information presented in Figure 4(a), obtained from World Soil Information (ISRIC)'s SoilGrids (Hengl et al. 2017), shows four texture types: coarse, medium, fine, and very fine, and was considered the primary factor in characterizing pixels. Four clusters were created, and a pixel was assigned to a cluster if one of the four soil types covered at least half its area: Cluster 1 (coarse), Cluster 2 (medium), Cluster 3 (fine) and Cluster 4 (very fine). Vegetation density implied from an Earth Resources Observation and Science (EROS) Center Visible Infrared Imaging Radiometer Suite (eVIIRS) Global Normalized Difference Vegetation Index (NDVI) image shown in Figure 4(b), was used to assign the pixels that remained unclassified by soil texture criterion. Pixels with low vegetation density were assigned to Cluster 1, and those with high vegetation density to Cluster 4. Finally, topography implied from ASTER digital elevation model shown in Figure 4(c) was used to assign the pixels that remained

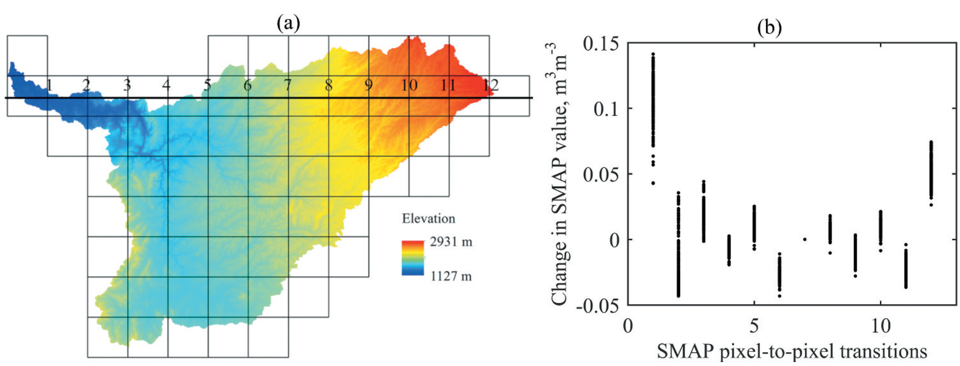

Figure 3. A linear transect across Sondu-Miriu watershed from downstream to upstream showing 12 pixel-to-pixel transitions (a), and the spatial variability of SMAP surface soil moisture depicted as the difference in moisture values of adjacent SMAP pixels (b).

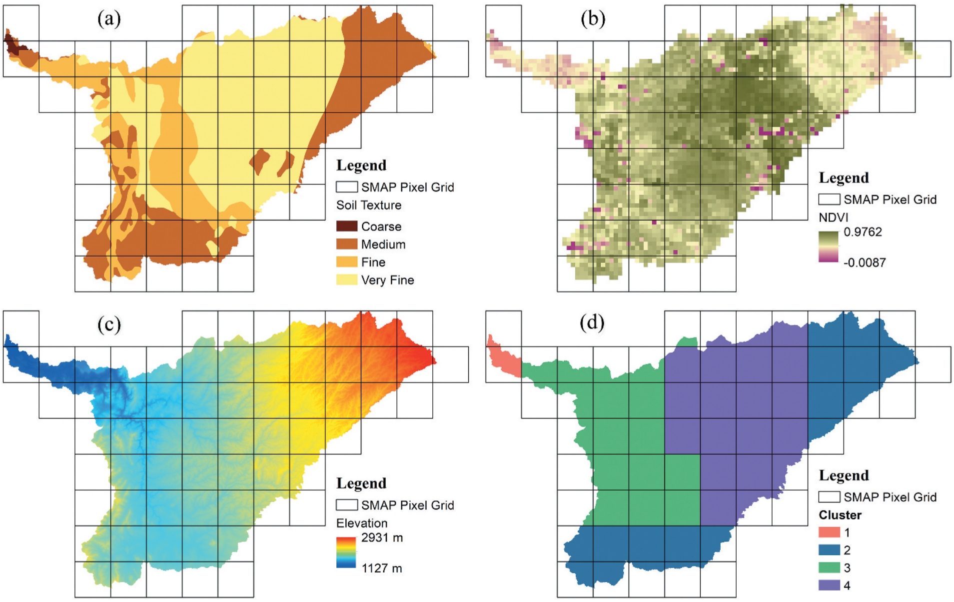

Figure 4. Characterization of SMAP pixels in Sondu-Miriu watershed based on soil texture (a), vegetation density implied from NDVI (b), and topography implied from digital elevation model (c), resulted in clustering of the pixels into four regions of similar landscape-defining characteristics (d).

unclassified by soil texture and vegetation density criteria. Downstream pixels were assigned to Cluster 1, upstream pixels to Cluster 2, and midstream pixels to Cluster 3 (lower midstream) or Cluster 4 (upper midstream). This led to grouping of pixels to four clusters as shown in Figure 4(d).

2.3.3. Selection of validation sites

Constrained further by the need to make rapid successive measurements so that the ground measurements generally fall within a reasonable temporal proximity to SMAP simulation time, the final step in selection of sampling sites was dictated by accessibility. From each cluster, a pixel with an access road cutting across it was selected as a representative sampling site. The selected validation sites, each approximately 9 km × 9 km corresponding to SMAP grid pixel, are shown in Figure 1(b) and summarized in Table 1. Each of the four sites represents one of the major soil texture types in the watershed: coarse (Mawego), medium (Kuresoi), fine (Sotik), and very fine (Kapsuser). The sites represent four different land use types: predominantly sparsely vegetated shrublands with low vegetation density (Mawego), predominantly small-scale annual croplands with medium vegetation density (Sotik), predominantly large-scale perennial

Table 1. Characteristics of SMAP validation sites in Sondu-Miriu watershed.

Site Name	Soil Texture Type	Vegetation Density	Stream Flow Stage
Mawego	Coarse	Low	Downstream
Kuresoi	Medium	Medium	Upstream
Sotik	Fine	Medium	Midstream
Kapsuser	Very fine	High	Midstream

croplands with high vegetation density (Kapsuser), and predominatly annual croplands with patches of forest cover and overall medium vegetation density (Kuresoi). The sites represent different stages of stream flow, upstream (Kuresoi), midstream (Kapsuser and Sotik), and downstream (Mawego).

2.3.4. Moisture meter calibration

Soil moisture ground measurements were collected using a portable soil moisture meter, Lutron PMS-714 (Master Instruments Pty Ltd, Australia). The moisture meter was calibrated to obtain soil-specific calibration equations suitable for the soil type in each of the four sampling sites. This was achieved through a laboratory process in which soil samples from the validation sites were saturated and allowed to naturally dry out, allowing multiple simultaneous measurements of soil moisture by both the moisture meter and the reference gravimetric method. The two sets of data were fitted to a linear model to produce a calibration equation for each soil type.

In addition to the calibration equations, the calibration procedure produced for each soil type two moisture parameters, one on each end of the calibration curve: on the upper end the soil saturation value, which is the maximum possible moisture content of the soil, and on the lower end the minimum moisture value, which is the lowest water content the soil retains when it reaches its driest level under natural conditions. This yielded saturation values of 0.34, 0.55, 0.45, 0.58 $m^3\ m^{-3}$ and minimum moisture values of 0.08, 0.24, 0.22, 0.21 $m^3\ m^{-3}$, respectively for Mawego, Kuresoi, Sotik and Kapsuser sampling sites. A detailed description of the calibration is available at Cheruiyot, Mito, and Menenti (2024).

2.3.5. Data collection

The requirement for spatial representativeness was met by collecting ground data along near-linear transects of about 8–12 km across the SMAP pixel grid, as shown in Figure 5(a)–(d). Measurements were taken at about 400–600 m lateral intervals along the transect, resulting in about 10–20 sampling points across the sampling site, providing sufficient replications over the spatial domain that exceed the SMAP recommended number of at least five for a 9×9 km square grid size (Colliander et al. 2017).

Measurements were taken from naturally occurring soils about 50–100 m off the access roads. For each sampling site, soil moisture measurements were taken twice

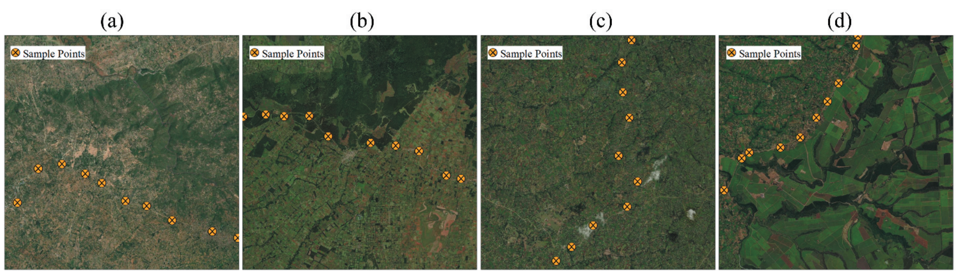

Figure 5. Picture showing a near-linear transect sampling of soil moisture, indicating sampling points along access roads in each of the four sampling sites: Mawego (a), Kuresoi (b), Sotik (c), and Kapsuser (d). The background image is a satellite imagery of the sites obtained from ArcMap basemaps.

a day for seven consecutive days; the first run between 10:00–11:00 hours and the second run between 16:00–17:00 hours, respectively designed to coincide with SMAP soil moisture simulated products for the time ranges 09:00–12:00 hours (average-centred at 10:30 hours) and 15:00–18:00 hours (average-centred at 16:30 hours). The second objective of completing each run of measurements within an hour is to minimize the effect of diurnal soil temperature variations between measurements on the sensor consistency.

While the sampling design adopted in this study results in almost linearly distributed sampling points that are known to be less representative of the sampling area compared to scattered sampling points (Oliver and Webster 1986), it is a low-cost alternative to a well distributed sensor network infrastructure. Transects as a spatial sampling technique for soil characteristics is not a new concept (Famiglietti, Rudnicki, and Rodell 1998; Wang et al. 2012), though most of the applications are at a much smaller spatial scale. In essence, the objective of a sampling design is to capture the spatial heterogeneity of soil moisture in the pixel grid, which is influenced by precipitation distribution, soil texture, land cover and topography (Colliander et al. 2017). The sampling sites in this study have almost homogenous soil texture within the pixel grid (less so in Mawego site), with generally the same land cover type (less so in Kuresoi site), and topography, and the 9 km × 9 km grid size is less than about 10–100 km spatial-temporal variability scale for precipitation, where a uniform rainfall distribution in one hour can be assumed (Cristiano, ten Veldhuis, and van de Giesen 2017). Therefore, considering the non-complex terrains of the sampling sites, the sampling design adopted provides to a reasonable extent a good representation of the sampled area.

Microwave remote sensing is generally limited to a soil sensing depth of approximately five centimetres (Mohanty 2013); variations depend on the soil moisture content and its distribution (Escorihuela et al. 2010). Based on a relationship between soil effective temperature and the sensing depth of microwave soil moisture radiometry, Lv et al. (2018) recommend 5 cm depth of soil moisture measurement as a ground reference to calibrate and validate satellite-based soil moisture products, as this depth captures the main signal source on average. Therefore, soil moisture ground measurements were collected from the top 5 cm of soil surface using a portable soil moisture meter with suitable site-specific calibrations.

2.4. Validation analysis

For each transect run of ground measurements, the measured meter values were averaged to obtain a single moisture value representing the soil moisture state of the grid at the field time. This value was compared with the corresponding SMAP value for the same grid over the same period. Linear correlation of the two sets of soil moisture values was performed to generate correlation statistics (R, RMSE). Statistical biases were also calculated to determine over- or under-estimation of SMAP simulations as compared to the reference ground measurements.

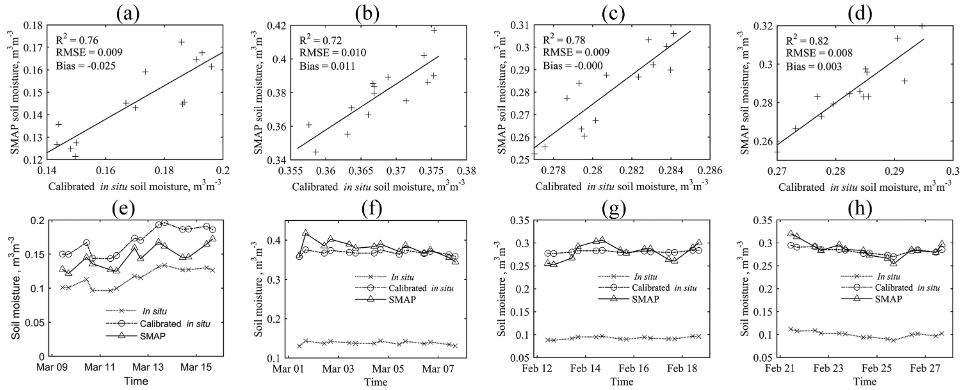

Figure 6. Graphical presentation of validation results for SMAP soil moisture at four validation sites. (a)–(d) show a comparison of SMAP and calibrated in situ soil moisture measurements, while (e)–(h) show a time series of the same moisture data over the field period, respectively for Mawego, Kuresoi, Sotik and Kapsuser sampling sites.

3. Results and discussion

3.1. SMAP validation results

Figure 6 is a graphical presentation of validation results for SMAP surface soil moisture at the four validation sites in Sondu-Miriu watershed. Figure 6(a)–(d) shows a comparison between calibrated soil moisture ground measurements and SMAP soil moisture, while Figure 6(e)–(h) shows a time series comparison of the same moisture data over the field period, respectively for Mawego, Kuresoi, Sotik and Kapsuser sampling sites. The two datasets compare well with R^2 of 0.76, 0.72, 0.78, 0.82 and RMSE of 0.009, 0.010, 0.009, 0.008 $m^3\ m^{-3}$, respectively.

The validation results presented in this work only differ marginally with those of Cheruiyot, Mito, and Menenti (2024), which produced RMSE in the range 0.007–0.010 $m^3\ m^{-3}$ and bias in the range −0.002–0.030 $m^3\ m^{-3}$ using spatially distributed sample points over the same study area (see Table 2). By contrast, Zhang, He, and Zhang (2017) obtained comparatively larger validation errors (RMSE of 0.074–0.077 $m^3\ m^{-3}$ and bias of 0.024–0.028 $m^3\ m^{-3}$) from a complex mountainous terrain. This shows that for a non-complex terrain, the sampling design presented in this work where areal values are obtained by a simple arithmetic mean of near-linear point measurements may suffice. This is consistent with the findings of Wu et al. (2020), which showed lower representativeness of *in situ* measurements in regions with higher topographic complexity.

Table 2. Comparison of validation errors obtained with near-linear transect sampling design proposed in this study, and those obtained from a spatially distributed sampling approach in Cheruiyot, Mito, and Menenti (2024).

	Near-linear transect sampling (present study)			Spatially distributed sampling (Cheruiyot, Mito, and Menenti (2024))		
Site	R^2	RMSE ($m^3\ m^{-3}$)	Bias ($m^3\ m^{-3}$)	R^2	RMSE ($m^3\ m^{-3}$)	Bias ($m^3\ m^{-3}$)
Mawego	0.76	0.009	−0.0246	0.79*	0.009*	−0.007*
Kuresoi	0.72	0.010	+0.0113	0.81	0.010	+0.030
Sotik	0.78	0.009	−0.0004	0.83	0.007	−0.014
Kapsuser	0.82	0.008	+0.0035	0.88	0.008	−0.002

*Not included in Cheruiyot, Mito, and Menenti (2024), but produced in a similar manner.

3.2. Impact of surface heterogeneity on accuracy of soil moisture sampling methods

With a bias of $-0.0004\,m^3\ m^{-3}$ and $+0.0035\,m^3\ m^{-3}$ respectively for Sotik and Kapsuser, SMAP soil moisture is insignificantly biased at these sites. These midstream sites have moderate vegetation cover and fine to very fine soil texture. This is consistent with findings of Mousa and Shu (2020), that SMAP shows best performance over moderate vegetation cover. SMAP soil moisture is however underestimated at Mawego site by $0.0246\,m^3\ m^{-3}$ and is overestimated at Kuresoi site by $0.0113\,m^3\ m^{-3}$. These results are in agreement with the findings of Zhang, He, and Zhang (2017), that SMAP posts largest errors in shrublands compared to other vegetation landscapes, where it underestimates soil moisture with the largest negative bias. Mawego is a downstream site characterized by the presence of uniformly sparse vegetation cover and a soil texture that is predominantly coarse but features patches of medium and fine textured soils. Kuresoi is an upstream site characterized by a predominantly medium textured soil and a medium density vegetation cover that is predominantly croplands with a small section of forest cover. These factors may have contributed to the reduced validation accuracy for the Mawego and Kuresoi sites, compared to Sotik and Kapsuser sites where all the three landscape-defining characteristics considered in this study are largely homogeneous. While the validation errors in the two less homogeneous sites are only marginally elevated, it is evident that a linear sampling method does not capture the spatial heterogeneity of the sampling area. It can be surmised from these results that the errors will increase with increase in the spatial heterogeneity, so that a spatially balanced sampling method is ideal for such complex terrains.

3.3. Temporal variability of SMAP surface soil moisture

Temporal variability of SMAP was evaluated for a period covering the first eight complete years of the product from 31 March 2015 01:30:00 to 30 March 2023 22:30:00, featuring 23,376 datasets at three hours intervals. A time series plot depicting the surface soil moisture variation of each sampling site over a selected representative period, the 2017 calendar year, is presented in Figure 7(a)–(d), respectively for Mawego, Kuresoi, Sotik and Kapsuser sampling sites, featuring SMAP soil moisture (black) and the bias-corrected SMAP moisture values (blue). Cyan circles indicate instances when the land surface was saturated, and the magenta circles indicate instances when the land surface attained minimum moisture content under natural conditions. The soil saturation value and the minimum moisture value of each site were determined through a soil meter calibration process described in the 'Moisture Meter Calibration' section. These results show that the fine and very fine texture soils of Sotik and Kapsuser sites have a notably higher water holding capacity compared to coarse texture soils of Mawego site. The medium textured soils showed a water holding capacity that is higher than would relatively compare with other texture types. Kuresoi site features a forested or recently deforested landscape whose soils have a high organic matter content, which together with clay have a combined effect of enhancing its water holding capacity (Kirkham 2005) and significantly influencing its wilting point (Bouyoucos 1939; Garg et al. 2017).

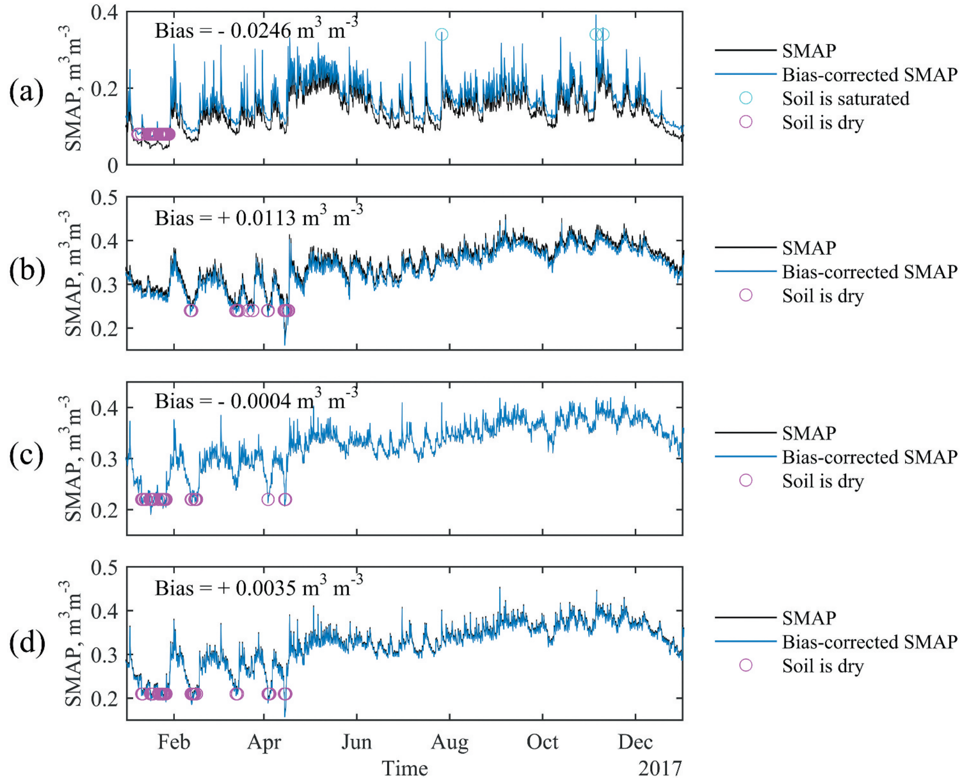

Figure 7. Time series plots showing temporal variation of SMAP soil moisture and the bias-corrected SMAP moisture values for Mawego (a), Kuresoi (b), Sotik (c), and Kapsuser (d) sites for the calendar year 2017. Instances when the bias-corrected SMAP soil moisture reached the minimum and maximum values for the soil types of the respective sites are indicated, showing periods when the soil was dry or saturated.

Of the 23,376 SMAP datasets analysed, the downstream site of Mawego and that of Sotik respectively experienced 60 and 159 occurrences of soil saturation during the eight-year period, in comparison to 13 and 153 occurrences during the same period before the datasets were corrected for biases. The upstream site of Kuresoi and that of Kapsuser experienced no saturation events during the period, with and without bias correction of the datasets. All sites experienced instances of extremely dry soil over the same period, with 365, 410, 113, and 216 occurrences respectively for Mawego, Kuresoi, Sotik and Kapsuser sites, in comparison to 1738, 268, 114, and 184 before the datasets were bias-corrected. These results demonstrate the possibility to utilize SMAP soil moisture data to detect episodes of flooding and drought at a landscape scale, but that the accuracy of these detections can change significantly with marginal adjustments to the estimated values of soil moisture.

Crow and Ryu (2009) developed a surface runoff prediction model that is based on three ideal observation characteristics of the input remotely sensed soil moisture data: a retrieval RMSE accuracy of 0.03 m^3 m^{-3}, a soil sensing depth of 10 cm, and an observation frequency of a day. SMAP has an excellent product simulation frequency of 3 hours, and can meet the RMSE accuracy of 0.03 m^3 m^{-3} with a proper validation and bias correction, as presented in this work. SMAP surface soil moisture product with a sensing depth of 5 cm is inadequate for this application, though a root zone product with a depth of 1 m is also available. Therefore, the retrieval accuracy and the temporal resolution of SMAP data seem suitable for hydrological applications. Assimilation of remotely sensed soil moisture in hydrological models has been shown to improve simulations of streamflow (López López et al. 2016; Wakigari and Leconte 2023).

Temporal stability is the concept that the spatial patterns of soil moisture are stable over time (Vanderlinden et al. 2012), so that while the areal soil moisture varies over time, there is a consistency in the geographic locations of relatively wetter and drier areas (Brocca et al. 2017). Temporal stability of SMAP data was evaluated by plotting the monthly moving average values of SMAP soil moisture for all pixels within the watershed over the eight-year period, as presented in Figure 8(a). This produced identifiable patterns of wetter and drier pixels, which were further verified by obtaining the mean soil moisture value for each pixel over the eight-year period. This led to the clustering of the watershed into four representative classes A, B, C and D, in terms of wetness as presented in Figure 8(b), with the soil moisture ranges: less than 0.20 $m^3\ m^{-3}$, 0.20–0.30 $m^3\ m^{-3}$, 0.30–0.35 $m^3\ m^{-3}$, and more than 0.35 $m^3\ m^{-3}$, respectively. This allows for a time- and cost-efficient monitoring of soil moisture in the watershed in a long-term basis by focusing only on four representative pixels, as proposed by Brocca et al. (2017).

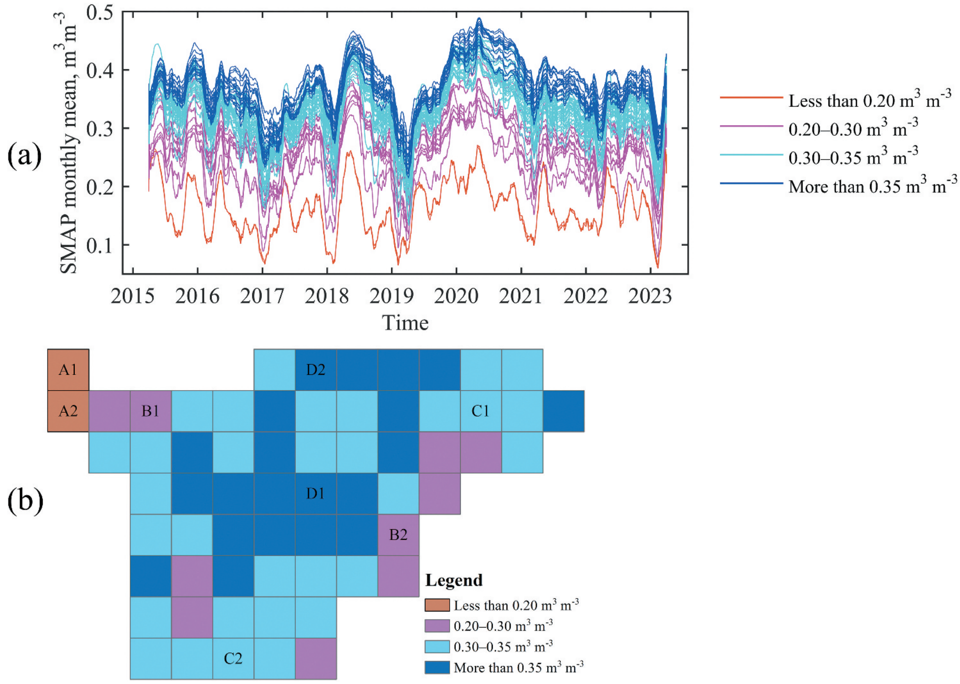

Figure 8. (a) Shows the temporal stability of SMAP surface soil moisture at Sondu-Miriu watershed over the period 31 March 2015 to 30 March 2023, obtained using monthly moving average soil moisture values. Each line represents a SMAP pixel, and is represented in a colour that indicates the range of its mean soil moisture value over the period. (b) shows the wetness clustering of SMAP pixels in the watershed based on their eight-year mean SMAP moisture values. A1, B1, C1, and D1 are randomly selected representative pixels for each cluster, while A2, B2, C2, and D2 are pixels with median moisture values for each cluster.

In order to test the concept of using temporal stability as a basis for selecting sampling areas, four SMAP pixels marked A1, B1, C1, and D1 in Figure 8(b) were randomly selected as the representative pixels for their respective classes. If the soil moisture values of these pixels were applied to unsampled areas, the magnitude of error in the estimation of soil moisture of each pixel in the watershed was computed as the difference between the actual SMAP soil moisture value of that pixel and that of the representative pixel in the respective class. The resultant errors presented in Figure 9(a) show values ranging from −0.2354 to +0.1818 $m^3\ m^{-3}$, which average to a range of −0.0736 to +0.0325 $m^3\ m^{-3}$ over an eight-year period. The relative errors shown in Figure 9(b) range from −0.8267 to 1.9000, and average to a range of −0.2525 to 0.0978, while standard deviation of the relative errors are 0.0124, 0.0893, 0.0318, and 0.0332, respectively for classes A, B, C, and D. When a SMAP pixel with the median soil moisture value of each class was selected as the representative pixel of the respective class, marked with A2, B2, C2, and D2 in Figure 8(b), the resulting errors presented in Figure 9(c) reduced to values ranging from −0.2264 to +0.1706 $m^3\ m^{-3}$,

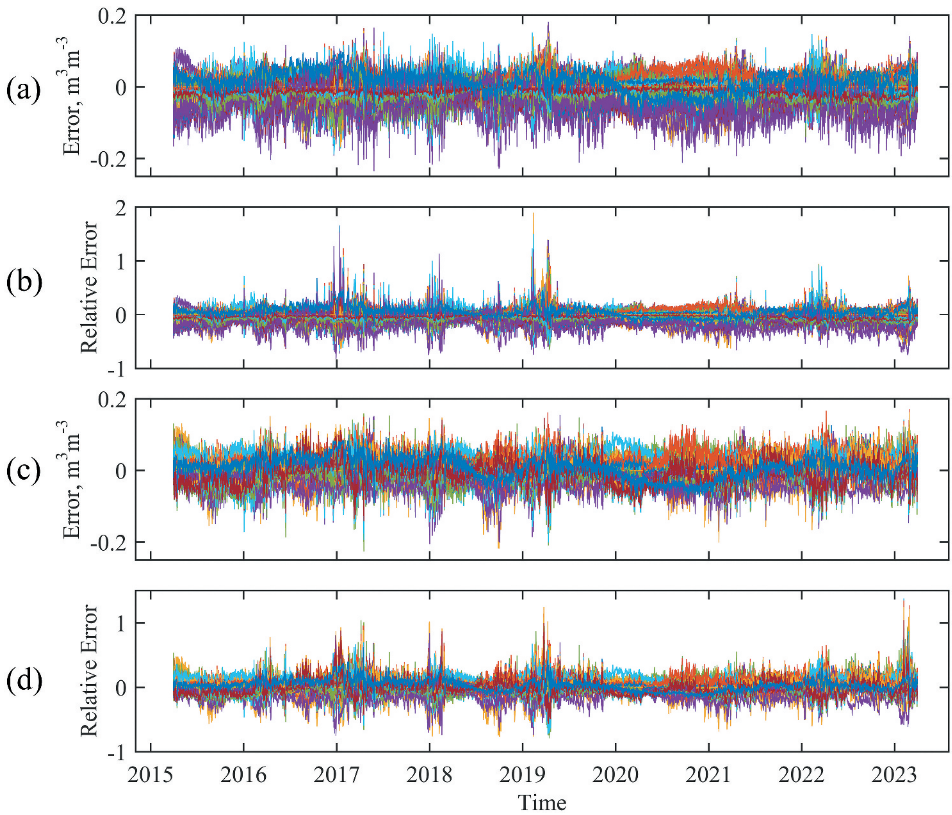

Figure 9. Magnitude of errors in the estimation of soil moisture of unsampled pixels when temporal stability wetness clustering is applied to the selection of sampling areas using randomly selected representative pixels within a class (a), and using pixels of median soil moisture values as representative pixels for the class (c). Each line represents a SMAP pixel in the watershed, and is represented in a unique random colour for visualization purposes only. (b) and (d) are the relative errors for the former and latter methods, respectively.

and averaging to a range of −0.0450 to +0.0318 m^3 m^{-3} over the eight-year period. The relative errors shown in Figure 9(d) range from −0.7891 to 1.3772, and average to a range of −0.1812 to 0.1215, while the standard deviation of the relative errors are 0.0150, 0.0987, 0.0335, and 0.0326, respectively for classes A, B, C, and D. The average errors from this latter method are closer to the SMAP target value of 0.04 m^3 m^{-3} (Colliander et al. 2017). These results show that this method results in large approximation errors in the short-term, but produce reasonable errors in the long-term.

4. Conclusions

While SMAP (Soil Moisture Active Passive) Level-4 surface soil moisture has been validated for select representative landscapes using well established core validation sites, these landscapes are not nearly as exhaustive since each landscape exhibits unique salient features determined largely by terrain, vegetation cover, soils, climate, etc. Localized validation of SMAP could enhance its reliability, but it is impractical to implement for every landscape the elaborate ground sampling infrastructure that meets the requirements of a standard SMAP core validation site in terms of spatial representativeness of sampling points within the sampling unit and the frequency of measurements. This study proposes a sampling approach that utilizes limited resources (a single sensor) to balance the need for spatial distribution of samples across the sampling site and the need for temporal proximity of the sampled measurements to the aggregation time of the satellite product being validated. It involves transect sampling along access roads that run across the sampling sites, with measurements taken at about 400–600 m lateral intervals and within ±30 minutes of the SMAP product simulation time. The proposed sampling approach is suitable for validation of SMAP in non-complex terrains where the landscape-defining characteristics do not substantially change within the SMAP grid, permitting the computation of areal soil moisture as simple arithmetic means of near-linear point measurements. Therefore, large and inhomogeneous landscapes must first be clustered to smaller near homogeneous units before application of the proposed method.

This sampling approach was applied to the validation of SMAP data at Sondu-Miriu watershed in Western Kenya. A spatial variability of SMAP soil moisture data revealed sharp changes between soil moisture values of adjacent SMAP pixels, indicating that the watershed could not be validated as a single unit, therefore it was clustered to four different landscapes. Sampled soil moisture measurements were corrected with soil-specific calibration equations, and averaged to obtain a single moisture value for the site that corresponds to SMAP product for that specific time. Validation analysis show a good agreement between *in situ* measurements and SMAP with R^2 of 0.76, 0.72, 0.80, and 0.82, respectively for Mawego, Kuresoi, Sotik and Kapsuser sites. With biases of −0.0246, +0.0113, 0.0004, and +0.0035 m^3 m^{-3}, it shows that SMAP is significantly underestimated at Mawego site, marginally overestimated at Kuresoi site, and insignificantly biased at Sotik and Kapsuser sites. These results are not significantly different from those obtained for the same sites using the standard sampling approach of spatially distributed sample points across the sampling site, indicating that the proposed simple sampling approach suffices in non-complex terrains.

An analysis of the temporal variability of SMAP soil moisture revealed the potential of the data as a resource to map flooding and drought events particularly because of its high repeat frequency of three hours, but a proper validation is required to remove marginal biases that can cause significant interpretation errors. An analysis of the temporal stability of SMAP data in the watershed showed the possibility to cluster the watershed to a few wetness classes, permitting time- and cost-efficient monitoring of soil moisture in a long-term basis by focusing on a few representative areas.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Elijah Cheruiyot http://orcid.org/0000-0003-0949-5161

Data availability statement

The data that support the findings of this study are available from the corresponding author, EC, upon reasonable request.

References

Bouyoucos, G. J. 1939. "Effect of Organic Matter on the Water-Holding Capacity and the Wilting Point of Mineral Soils." *Soil Science* 47 (5): 377–384. https://doi.org/10.1097/00010694-193905000-00005 .

Brocca, L., L. Ciabatta, C. Massari, S. Camici, and A. Tarpanelli. 2017. "Soil Moisture for Hydrological Applications: Open Questions and New Opportunities." *Water* 9 (2): 140. https://doi.org/10.3390/w9020140.

Cheruiyot, E. K., C. Mito, M. Menenti, B. Gorte, R. Koenders, and N. Akdim. 2014. "Evaluating MERIS-Based Aquatic Vegetation Mapping in Lake Victoria." *Remote Sensing* 6 (8): 7762–7782. https://doi.org/10.3390/rs6087762.

Cheruiyot, E., C. Mito, and M. Menenti. 2024. "An Improved Method of Soil Moisture Meter Calibration for Satellite Data Validation at Watershed Scale." *Earth Science Informatics* 17 (1): 117–129. https://doi.org/10.1007/s12145-023-01170-w.

Choudhury, B. J., T. J. Schmugge, A. Chang, and R. W. Newton. 1979. "Effect of Surface Roughness on the Microwave Emission from Soils." *Journal of Geophysical Research Oceans* 84 (C9): 5699–5706. https://doi.org/10.1029/JC084iC09p05699.

Colliander, A., T. J. Jackson, R. Bindlish, S. Chan, N. Das, S. B. Kim, M. H. Cosh, et al. 2017. "Validation of SMAP Surface Soil Moisture Products with Core Validation Sites." *Remote Sensing of Environment* 191 (March): 215–231. https://doi.org/10.1016/j.rse.2017.01.021.

Cristiano, E., M.-C. ten Veldhuis, and N. van de Giesen. 2017. "Spatial and Temporal Variability of Rainfall and Their Effects on Hydrological Response in Urban Areas – a Review." *Hydrology and Earth System Sciences* 21 (7): 3859–3878. https://doi.org/10.5194/hess-21-3859-2017 .

Crow, W. T., and D. Ryu. 2009. "A New Data Assimilation Approach for Improving Runoff Prediction Using Remotely-Sensed Soil Moisture Retrievals." *Hydrology and Earth System Sciences* 13 (1): 1–16. https://doi.org/10.5194/hess-13-1-2009.

Duygu, M. B., and Z. Akyürek. 2019. "Using Cosmic-Ray Neutron Probes in Validating Satellite Soil Moisture Products and Land Surface Models." *Water* 11 (7): 1362. https://doi.org/10.3390/w11071362.

Escorihuela, M. J., A. Chanzy, J. P. Wigneron, and Y. H. Kerr. 2010. "Effective Soil Moisture Sampling Depth of L-Band Radiometry: A Case Study." *Remote Sensing of Environment* 114 (5): 995–1001. https://doi.org/10.1016/j.rse.2009.12.011.

Famiglietti, J. S., J. W. Rudnicki, and M. Rodell. 1998. "Variability in Surface Moisture Content Along a Hillslope Transect: Rattlesnake Hill, Texas." *Journal of Hydrology* 210 (1–4): 259–281. https://doi.org/10.1016/S0022-1694(98)00187-5.

Garg, A., J. Li, J. Hou, C. Berretta, and A. Garg. 2017. "A New Computational Approach for Estimation of Wilting Point for Green Infrastructure." *Measurement* 111 (December): 351–358. https://doi.org/10.1016/j.measurement.2017.07.026.

Hengl, T., J. M. de Jesus, G. B. M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, et al. 2017. "SoilGrids250m: Global Gridded Soil Information Based on Machine Learning." *PLoS One* 12 (2): e0169748. https://doi.org/10.1371/journal.pone.0169748.

Jackson, T. J., T. J. Schmugge, and J. R. Wang. 1982. "Passive Microwave Sensing of Soil Moisture Under Vegetation Canopies." *Water Resources Research* 18 (4): 1137–1142. https://doi.org/10.1029/WR018i004p01137.

Jiang, D., and K. Wang. 2019. "The Role of Satellite-Based Remote Sensing in Improving Simulated Streamflow: A Review." *Water* 11 (8): 1615. https://doi.org/10.3390/w11081615.

Kirkham, M. B. 2005. "8 - Field Capacity, Wilting Point, Available Water, and the Non-Limiting Water Range." In *Principles of Soil and Plant Water Relations*, edited by M. B. Kirkham, 101–115. Burlington: Academic Press. https://doi.org/10.1016/B978-012409751-3/50008-6.

Kofidou, M., and A. Gemitzi. 2023. "Assimilating Soil Moisture Information to Improve the Performance of SWAT Hydrological Model." *Hydrology* 10 (8): 176. https://doi.org/10.3390/hydrology10080176.

Kolassa, J., R. H. Reichle, Q. Liu, M. Cosh, D. D. Bosch, T. G. Caldwell, A. Colliander, et al. 2017. "Data Assimilation to Extract Soil Moisture Information from SMAP Observations." *Remote Sensing* 9 (11): 1179. https://doi.org/10.3390/rs9111179.

Konings, A. G., M. Piles, N. Das, and D. Entekhabi. 2017. "L-Band Vegetation Optical Depth and Effective Scattering Albedo Estimation from SMAP." *Remote Sensing of Environment* 198 (September): 460–470. https://doi.org/10.1016/j.rse.2017.06.037.

Lakshmi, V. 2013. "Remote Sensing of Soil Moisture." *International Scholarly Research Notices* 2013 (March): e424178. https://doi.org/10.1155/2013/424178.

López López, P., N. Wanders, J. Schellekens, L. J. Renzullo, E. H. Sutanudjaja, and M. F. P. Bierkens. 2016. "Improved Large-Scale Hydrological Modelling Through the Assimilation of Streamflow and Downscaled Satellite Soil Moisture Observations." *Hydrology and Earth System Sciences* 20 (7): 3059–3076. https://doi.org/10.5194/hess-20-3059-2016.

Lv, S., Y. Zeng, J. Wen, H. Zhao, and Z. Su. 2018. "Estimation of Penetration Depth from Soil Effective Temperature in Microwave Radiometry." *Remote Sensing* 10 (4): 519. https://doi.org/10.3390/rs10040519.

Meißl, G., T. Zieher, and C. Geitner. 2020. "Runoff Response to Rainfall Events Considering Initial Soil Moisture – Analysis of 9-Year Records in a Small Alpine Catchment (Brixenbach Valley, Tyrol, Austria)." *Journal of Hydrology: Regional Studies* 30 (August): 100711. https://doi.org/10.1016/j.ejrh.2020.100711.

Melesse, A. M., Q. Weng, P. S. Thenkabail, and G. B. Senay. 2007. "Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling." *Sensors (Basel, Switzerland)* 7 (12): 3209–3241. https://doi.org/10.3390/s7123209.

Minet, J., E. Laloy, S. Lambot, and M. Vanclooster. 2011. "Effect of High-Resolution Spatial Soil Moisture Variability on Simulated Runoff Response Using a Distributed Hydrologic Model." *Hydrology and Earth System Sciences* 15 (4): 1323–1338. https://doi.org/10.5194/hess-15-1323-2011.

Mohanty, B. P. 2013. "Soil Hydraulic Property Estimation Using Remote Sensing: A Review." *Vadose Zone Journal* 12 (4): 1–9. https://doi.org/10.2136/vzj2013.06.0100.

Monerris, A., T. Schmugge, A. Monerris, and T. Schmugge. 2009. "Soil Moisture Estimation Using L-Band Radiometry." In *Advances in Geoscience and Remote Sensing*, IntechOpen. https://doi.org/10.5772/8334.

Mousa, B. G., and H. Shu. 2020. "Spatial Evaluation and Assimilation of SMAP, SMOS, and ASCAT Satellite Soil Moisture Products Over Africa Using Statistical Techniques." *Earth & Space Science* 7 (1): e2019EA000841. https://doi.org/10.1029/2019EA000841.

Myeni, L., M. E. Moeletsi, and A. D. Clulow. 2019. "Present Status of Soil Moisture Estimation Over the African Continent." *Journal of Hydrology: Regional Studies* 21 (February): 14–24. https://doi.org/10.1016/j.ejrh.2018.11.004.

Narasimhan, B., R. Srinivasan, J. G. Arnold, and M. Di Luzio. 2005. "Estimation of Long-Term Soil Moisture Using a Distributed Parameter Hydrologic Model and Verification Using Remotely Sensed Data." *Transactions of the ASAE* 48 (3): 1101–1113. https://doi.org/10.13031/2013.18520.

Oliver, M. A., and R. Webster. 1986. "Combining Nested and Linear Sampling for Determining the Scale and Form of Spatial Variation of Regionalized Variables." *Geographical Analysis* 18 (3): 227–242. https://doi.org/10.1111/j.1538-4632.1986.tb00095.x.

Olokotum, M., V. Mitroi, M. Troussellier, R. Semyalo, C. Bernard, B. Montuelle, W. Okello, C. Quiblier, and J.-F. Humbert. 2020. "A Review of the Socioecological Causes and Consequences of Cyanobacterial Blooms in Lake Victoria." *Harmful Algae* 96 (June): 101829. https://doi.org/10.1016/j.hal.2020.101829.

Reichle, R. H., G. J. M. De Lannoy, R. D. Koster, W. Crow, J. S. Kimball, Q. Liu, and M. Bechtold. 2022. "SMAP L4 Global 3-Hourly 9 Km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 7, Surface Soil Moisture." Boulder, Colorado USA: NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/EVKPQZ4AFC4D.

Reichle, R. H., G. J. M. De Lannoy, Q. Liu, J. V. Ardizzone, A. Colliander, A. Conaty, W. Crow, et al. 2017. "Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using in situ Measurements." *Journal of Hydrometeorology* 18 (10): 2621–2645. https://doi.org/10.1175/JHM-D-17-0063.1.

Shahrban, M., J. P. Walker, Q. J. Wang, and D. E. Robertson. 2018. "On the Importance of Soil Moisture in Calibration of Rainfall–Runoff Models: Two Case Studies." *Hydrological Sciences Journal* 63 (9): 1292–1312. https://doi.org/10.1080/02626667.2018.1487560.

Song, S., and W. Wang. 2019. "Impacts of Antecedent Soil Moisture on the Rainfall-Runoff Transformation Process Based on High-Resolution Observations in Soil Tank Experiments." *Water* 11 (2): 296. https://doi.org/10.3390/w11020296.

Ulaby, F. T., G. A. Bradley, and M. C. Dobson. 1979. "Microwave Backscatter Dependence on Surface Roughness, Soil Moisture, and Soil Texture: Part II-Vegetation-Covered Soil." *IEEE Transactions on Geoscience Electronics* 17 (2): 33–40. https://doi.org/10.1109/TGE.1979.294626.

Vanderlinden, K., H. Vereecken, H. Hardelauf, M. Herbst, G. Martínez, M. H. Cosh, and Y. A. Pachepsky. 2012. "Temporal Stability of Soil Water Contents: A Review of Data and Analyses." *Vadose Zone Journal* 11 (4): vzj2011–0178. https://doi.org/10.2136/vzj2011.0178.

Wakigari, S. A., and R. Leconte. 2023. "Assessing the Potential of Combined SMAP and in-Situ Soil Moisture for Improving Streamflow Forecast." *Hydrology* 10 (2): 31. https://doi.org/10.3390/hydrology10020031.

Wang, J.-F., A. Stein, B.-B. Gao, and Y. Ge. 2012. "A Review of Spatial Sampling." *Spatial Statistics* 2 (December): 1–14. https://doi.org/10.1016/j.spasta.2012.08.001.

Wang, J. R., P. E. O'Neill, T. J. Jackson, and E. T. Engman. 1983. "Multifrequency Measurements of the Effects of Soil Moisture, Soil Texture, and Surface Roughness." *IEEE Transactions on Geoscience & Remote Sensing* GE-21 (1): 44–51. https://doi.org/10.1109/TGRS.1983.350529.

Wang, L., and J. J. Qu. 2009. "Satellite Remote Sensing Applications for Surface Soil Moisture Monitoring: A Review." *Frontiers of Earth Science in China* 3 (2): 237–247. https://doi.org/10.1007/s11707-009-0023-7.

Wu, X., G. Lu, Z. Wu, H. He, T. Scanlon, and W. Dorigo. 2020. "Triple Collocation-Based Assessment of Satellite Soil Moisture Products with in situ Measurements in China: Understanding the Error Sources." *Remote Sensing* 12 (14): 2275. https://doi.org/10.3390/rs12142275.

Yu, D., P. Xie, X. Dong, X. Hu, J. Liu, Y. Li, T. Peng, H. Ma, K. Wang, and S. Xu. 2018. "Improvement of the SWAT Model for Event-Based Flood Simulation on a Sub-Daily Timescale." *Hydrology and Earth System Sciences* 22 (9): 5001–5019. https://doi.org/10.5194/hess-22-5001-2018 .

Zhang, L., C. He, and M. Zhang. 2017. "Multi-Scale Evaluation of the SMAP Product Using Sparse in-Situ Network Over a High Mountainous Watershed, Northwest China." *Remote Sensing* 9 (11): 1111. https://doi.org/10.3390/rs9111111.