Evaluation of soil moisture from CCAM-CABLE simulation, satellite-based models estimates and satellite observations: Skukuza and Malopeni flux towers region case study

Floyd V. Khosa[1,2], Mohau J. Mateyisi[1], Martina R. van der Merwe[1], Gregor T. Feig[1,3,4], Francois A. Engelbrecht[5,6], Michael J. Savage[2]

Correspondence to: Floyd Khosa (vukosikhosa@yahoo.com)

[1] CSIR, Natural Resources and the Environment – Global Change and Ecosystem Dynamics, P.O. Box 395, Pretoria 0001, South Africa

[2] Agrometeorology Discipline, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

[3] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa

[4] South African Environmental Observation Network (SAEON), P.O. Box 2600, Pretoria 0001, South Africa

[5] CSIR, Natural Resources and the Environment – Climate Studies, Modelling and Environmental Health, P.O. Box 395, Pretoria, 0001, South Africa

[6] Global Change Institute (GCI), University of the Witwatersrand, Johannesburg, 2050, South Africa

Abstract

Reliable estimates of daily, monthly and seasonal soil moisture are useful in a variety of disciplines. The availability of continuous in situ soil moisture observations in southern Africa barely exists hence, process-based simulation model outputs are a valuable source of climate information, needed for guiding farming practises and policy interventions at various spatio-temporal scales. The aim of this study is to evaluate soil moisture outputs from simulation and satellite-based soil moisture products, and to compare modelled soil moisture across different landscapes. The simulation model consists of a global circulation model known as the conformal-cubic atmospheric model (CCAM), coupled with the CSIRO Atmosphere Biosphere Land Exchange model (CABLE). The satellite-based soil moisture data products include satellite observations from the European Space Agency (ESA) and satellite observation-based model estimates from the Global Land Evaporation Amsterdam Model (GLEAM). The evaluation is done for both the surface (0-10 cm) and root zone (10-100 cm) using in situ soil moisture measurements collected from two study sites. Due to the differences in the spatial scales of the various datasets, the evaluation mainly focuses on phase agreement as opposed to the absolute values. The results indicate that both the simulation and satellite data derived models produce outputs that are higher in range compared to in situ soil moisture observations at the two study sites, especially at the surface. The correlation coefficient ranges between 0.7 to 0.8 (at the root zone) and 0.7 to 0.9 (at the surface), suggesting that models mostly agree at the surface than at the root zone, this was further confirmed by the root mean squared error and the standard deviation values. The models mostly show a bias towards overestimation of the observed soil moisture both at the surface and root zone, with the CCAM-CABLE showing the least bias. An analysis evaluating phase agreement using the cross-wavelet analysis has shown that the models are generally in phase with the in situ observations with some time lags in some instances. An analysis of soil moisture mutual information

40   (MI) between CCAM-CABLE and the GLEAM models has successfully revealed that both the simulation and model estimates have a high MI at the root zone as opposed to the surface. The MI mostly ranges between 0.5 and 1.5 both at the surface and root zone. The MI is predominantly high for low lying areas relative to high lying areas.

Keywords: atmospheric model, cross wavelet, flux tower, land surface model, soil moisture

45   1   Introduction

Accurate estimates[1] of daily, monthly and seasonal soil moisture is important in a number of fields including agriculture (McNally et al., 2016), water resources planning (Decker, 2015), weather forecasting (van den Hurk et al., 2012) and the quantification of the impacts of extreme weather events such as droughts (Sheffield and Wood, 2008), heat waves (Fischer et al., 2007; Lorenz et al.,
50   2010) and floods (Brocca et al., 2011). Soil moisture has been identified as one of the 50 essential climate variables (ECVs) by the Global Climate Observing System (GCOS) and the European Space Agency climate change initiative (ESA-CCI) (McNally et al., 2016). Available soil moisture affects the fluxes of heat and water at the surface and directly impacts local and regional weather patterns (Dorigo et al., 2015; Raoult et al., 2018; Yuan and Quiring, 2017).

55   Soil moisture is a key parameter to consider in the partitioning of precipitation and net radiation. The temporal and spatial variation in soil moisture is controlled by vegetation, topography, soil properties and climate variability (Xia et al., 2015). Root zone soil moisture plays a vital role in the transpiration process of evapotranspiration (ET) especially in arid and semi-arid regions, where most of the water loss is accounted for by transpiration during the dry period (Jovanovic et al., 2015; Palmer et al.,
60   2015). The dry period, which is constituted by months when the sites experience minimum rainfall, occurs during the austral winter season May to October. Regions where soil moisture strongly influences the atmosphere is at the transition between wet and dry climates. This is associated with the strong coupling between ET and soil moisture which is a characteristic of these regions (van den Hurk et al., 2012; Lorenz et al., 2010).

65   The model evaluation in this study is achieved through a qualitative and quantitative comparison of modelled and in situ soil moisture products. Modelled and Satellite data derived soil moisture fields can be derived at different temporal and spatial resolutions while in site observations are mainly point-based (Fang et al., 2016). Despite the in situ data being limited in coverage, they are very useful for the calibration and validation of modelled and satellite-derived soil moisture estimates (Xia et al.,
70   2015). Point-based in situ soil moisture data that are used as a reference in this study consist of surface and root zone measurements. The fact that the in situ data are point-based, poses significant challenges in the understanding of spatial patterns in soil moisture (Yuan and Quiring, 2017). Direct satellite observations, on the other hand, are presently only available for the surface. To obtain root zone estimates of soil moisture satellite-based surface soil moisture data are used in conjunction with
75   ground-based observations and model estimates. The modelled soil moisture data are largely dependent on accurate surface forcing data (e.g. air temperature, precipitation and radiation) and the parameterisation of the land surface schemes (Xia et al., 2015). This is done in the framework of physically-based models whose accuracy may vary depending on the response of the models to the forcing data.

---

[1] *Estimate* here refers to both process-based model simulation and satellite-derived data products thereafter, the term *simulation* will be used for process-based model outputs while *estimates* will be reserved for satellite-derived data.

80   The study is inspired by the notion that an understanding of soil moisture characteristic patterns, for the study region can be reliably obtained by looking at independent datasets from simulation experiments, theoretical or analytical models and in situ observations. In Africa, the evaluations of the soil moisture data products, from these various estimation approaches, are sparse mainly due to the lack of publicly available in situ observations (Sinclair and Pegram, 2010). The lack of publicly

85   available long term and complete in situ soil moisture measurements in most parts of the world leads to a reliance on global climate models (GCMs) to estimate the land surface states (Dirmeyer et al., 2013). The data produced by land surface models, hydrological models and GCMs have been widely evaluated for many continents and regions (Albergel et al., 2012; An et al., 2016; Dorigo et al., 2015; McNally et al., 2016; Yuan and Quiring, 2017). The available studies include those conducted by

90   McNally et al. (2016) and Dorigo et al. (2015) evaluating ESA-CCI satellite soil moisture products over East and West Africa respectively.

The aims of this study are twofold. Firstly, it is to evaluate the ability of the process-based simulation and satellite-derived soil moisture products to capture the observed variability in soil moisture at specific flux tower locations. Secondly, to understand how a coupled land-atmosphere model

95   simulated results of soil moisture compare against satellite-based estimates on broad landscape classes that belong to homogenous elevation and soil types. The evaluation is undertaken at two soil depths namely; surface (SSM, i.e., 0-10 cm) and root zone (RZSM, i.e. 10-100 cm), using long term in situ measurements with the objective of establishing if the respective soil moisture data products are representative of local conditions. This is done for two study sites whose data records are publicly

100   available, namely the Skukuza and Malopeni flux tower sites located in the Kruger National Park in South Africa. The two study sites receive summer rainfall and the colder winter months overlap with the dry period. Of these two sites, only the Skukuza site forms part of the global flux data network (FLUXNET). Other international flux observation networks, such as the International Soil Moisture Network (ISMN) have no affiliated data sites on the study region.

105   An investigation of how the CCAM-CABLE process-based simulation, satellite-derived and GLEAM model estimates compare with the in situ observations is undertaken. We look at the spatio-temporal variations in simulated soil moisture data from a coupled land-atmosphere model. The conformal cubic atmospheric model (CCAM) of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) coupled to the CSIRO Atmosphere Biosphere Land Exchange (CABLE)

110   model, three versions of the European Space Agency (ESA) satellite observations (i.e., active, passive and combined), and estimates from three versions of the global land evaporation Amsterdam Model (GLEAM) are evaluated. The central idea is to understand how processed based and satellite-based models spatial patterns compare at a regional level, with a focus on grid points that belong to specific landscapes classes. This is done for landscapes where the availability of in situ observations over

115   space and time presents a major challenge for climate model evaluation studies. We focus on the periodic patterns of soil moisture at a point. In particular, we investigate, both quantitatively and qualitatively, the agreement in phase and magnitude between the respective soil moisture data products with a view of establishing if they are representative of local conditions.

An understanding of the extent to which the climate model simulations and GLEAM model estimates

120   have similar patterns at a regional level within inter-annual time scales is achieved by looking at a measure of their mutual information (MI). Model correspondence in capturing dominating processes is investigated by looking at the modelled soil moisture signal mutual information (MI). This is done for different landscapes organised by dominating soil and vegetation types, as well as altitude ranges across the study region. The study seeks not only to uncover interesting patterns in the observed data,

125   for the study region but also to highlight the strengths as well as aspects of the climate model simulation and GLEAM estimates which may benefit from continuous testing and improvement.

The ability of models to capture seasonal cycles of terrestrial processes such as soil moisture is one indication of how well the physical processes that underlie the variability of soil moisture over space and time are represented. A comparison of satellite-derived products with in situ observations may
130   also yield useful insight into the strengths and weaknesses of various remote sensing techniques that are used. A climate models' ability to represent and capture the seasonality of a system under inter- and intra-annual climate variability could be considered more important than its agreement with observations in absolute values (Fang et al., 2016). The remainder of the study is structured as follows: Sect. 2 describes the datasets used, the study design and methods for analysing the datasets.
135   Section 3 presents the results and the discussion, followed by the conclusions in Sect. 4.

2   Materials, methods and data

2.1   Study sites and in situ observations

In situ soil moisture measurements from the Council for Scientific and Industrial Research (CSIR) operated network of eddy covariance flux towers, in the Lowveld region of the Mpumalanga
140   (Skukuza) and Limpopo (Malopeni) provinces are used. Several other soil moisture in situ measurements sites exist in the country. However, their data are not publicly available.

2.1.1   Skukuza

The Skukuza flux tower site is a long-term measurement site, located within the Kruger National Park conservation area in South Africa (25.0197° S, 31.4969° E; Fig. 1). The Skukuza flux tower has been
145   operational from 2000 to present. The site falls within a semi-arid savanna biome at an altitude of 370 m above sea level, with a mean rainfall of 547 mm year^{-1}, and average annual minimum (during the dry season) and maximum (during the wet season) temperatures of 14.5 and 29.5°C, respectively for the averaging period from 2001 to 2014. The vegetation is dominated by an overstory of *Combretum apiculatum* (Sond.), and *Sclerocarya birrea* (Hochst.) with a height of approximately 8-10 m, and a
150   tree cover of approximately 30% (Archibald et al., 2009). The understory is a grass layer dominated by *Panicum maximum* (Jacq.), *Digitaria eriantha* (Steud.), *Eragrostis rigidor* (Pilg.) and *Pogonarthria squarrosa* (Roem. and Schult.). The soil has a yellowish sandy loam texture and is of the Clovelly form (Feig et al., 2008), and the dominant soil type for the 25 km resolution grid cell where the flux tower is located is silty loam (Fig. 1). The Skukuza flux tower site is extensively
155   described in previous studies including those by Archibald et al. (2009), Scholes et al. (2001) and Khosa et al. (2019). In situ soil moisture data are collected 90 m north of the tower, and the measurements are taken at two profiles which are 8 m apart. The sensors are located at four different depths for both profiles i.e., 5, 15, 30 and 40 cm (Pinheiro and Tucker, 2001). Time-domain reflectometry (TDR) probes (Campbell Scientific CS615L) are used to measure soil moisture at a 30-
160   minute temporal resolution. These measurements were averaged to a daily time period (only done for days for which at least 80 % of the half-hourly measurements were available over a 24-hour period) in order to match the resolution of the other soil moisture products. For this study, the in situ data from the year 2001 to 2014 are used.

2.1.2   Malopeni

165   The Malopeni flux tower is located 130 km North west of the Skukuza flux tower (23.8325° S, 31.2145° E; Fig. 1), at an elevation of 384 m above sea level. The tower has been collecting data since

2008 to present, however, data was not collected between January of 2010 and January of 2012 due to equipment failure. The site has a mean rainfall of 472 mm year^{-1}, and annual average minimum and maximum air temperatures of 12.4 and 30.5°C respectively, for the averaging period from 2008 to 2014. The site is dominated by broadleaf *Colophospermum mopane*, which characterises a hot and dry savanna (Ramoelo et al., 2014), *Combretum apiculatum* and *Acacia nigrescens* are also abundant at the site. The grass layer is dominated by *Schmidtia pappophoroides* and *Panicum maximum*. The soil at the site is predominantly of the shallow sandy loam texture, and the dominant soil type for the 25 km resolution grid cell where the flux tower is located is silty loam (Fig. 1). The soil moisture probes are located at four different profiles and depths. The sensor types and depths positioning are the same for the Malopeni and Skukuza flux tower sites. Soil moisture is collected at four different profiles (i.e., 16 sensors at four depths per site) and averaged to represent surface and root zone soil moisture at the site, for Skukuza only sensors at two profiles are working (i.e., 8 sensors).

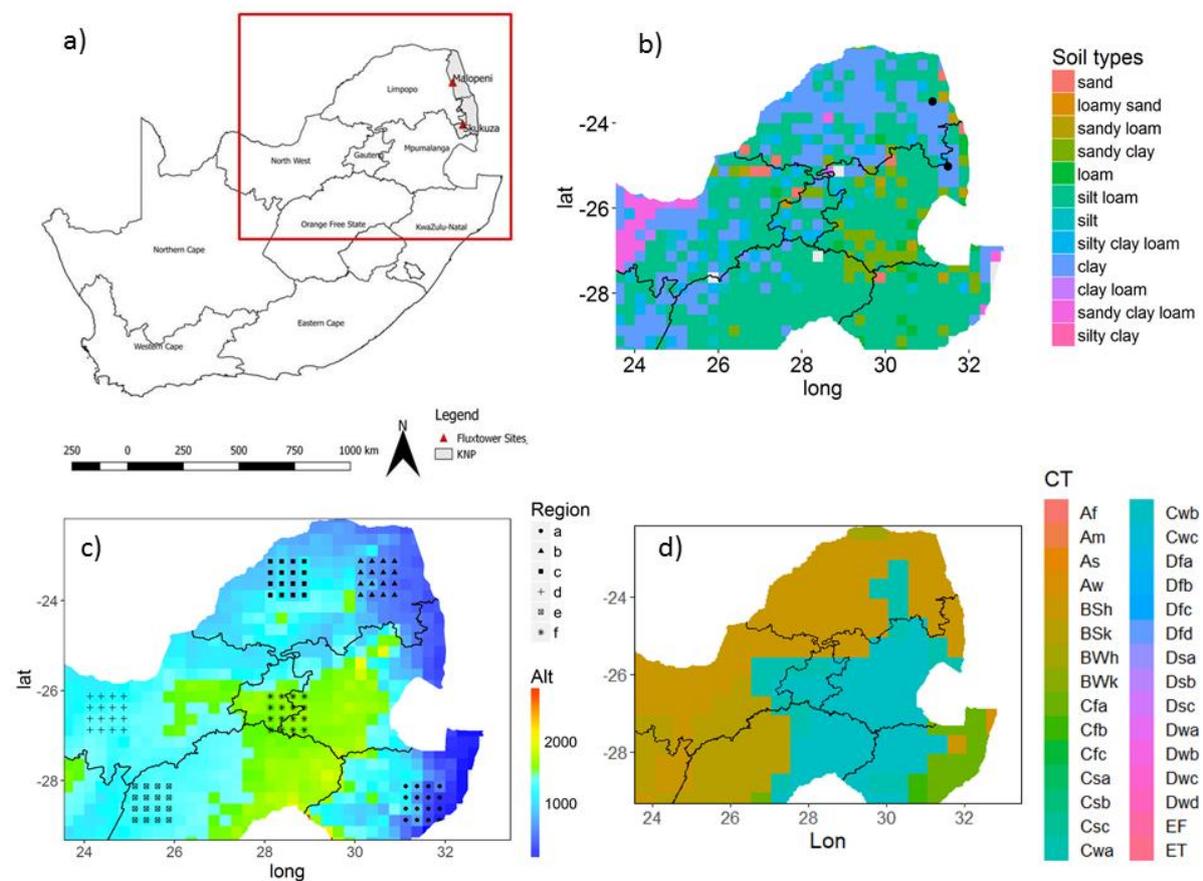



Figure 1. Maps indicating (a) South Africa, Kruger national park (KNP), flux tower sites (Skukuza and Malopeni) and the area considered for grid inter-comparison (red box), (b) dominant soil types (https://soilgrids.org) per grid cell, at a resolution of 25 km, (c) the altitude (Alt, m) at the study region at a 25 km resolution, with the selected grid points clusters (i.e., a-f) demarcating landscape classes that belong to specific elevation categories labelled in the order of the increasing altitude, the elevation dataset is obtained from the national centers for environmental information (https://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html) (d) Köeppen-Geiger climate types (CT) across the study region at a 50 km resolution (http://koeppen-geiger.vu-wien.ac.at/present.htm).

2.2 Datasets

2.2.1 Soils texture data

190 The "SoilGrids" dataset from the international soil reference information centre (ISRIC) was used in this study to map soil types (Fig. 1b). The data are available online (https://soilgrids.org), and is described in detail in the study by Hengl et al. (2017). The dataset has a spatial resolution of 250 m and is resampled to 25 km, firstly by resampling to 1 km and then to 25 km, using the nearest neighbour method to match the resolution of the soil moisture products. We acknowledge that

195 resampling from fine to coarse resolution might introduce a bias towards certain soil types. However, the nearest neighbour method is suitable for resampling categorical data. Soils were classified into 12 dominant types ranging between sand and silty clay as shown in Fig. 1d. The soil type data are available at various depths, here we only consider the data representing the surface (i.e., 0-5 cm).

2.2.2 Satellite observations

200 The European Space Agency climate change initiative (ESA-CCI) satellite-derived soil moisture datasets are used in this study (Liu et al., 2012; Yuan and Quiring, 2017). These global datasets are based on passive and active satellite microwave sensors, and provide surface soil moisture estimates at a resolution of ~25 km (i.e., 0.25˚) (Fang et al., 2016; Yuan and Quiring, 2017). The ESA-CCI merges soil moisture estimates from the active and passive satellite microwave sensors into one

205 dataset (http://www.esa-soilmoisture-cci.org/), using the backward propagating cumulative distribution function method (Dorigo et al., 2015; Fang et al., 2016). A detailed description of the merged active and passive sensors and their functioning is provided by Fang et al. (2016), Dorigo et al. (2015) and Liu et al. (2012). The merging of active and passive sensors is based on their sensitivity to vegetation density, as the accuracy of these products varies as a function of vegetation cover (Liu et

210 al., 2012). In this study, version 3.2 (v3.2) of the ESA-CCI soil moisture data is used. The merged data product is used in this study as it has better data coverage compared to the individual products. Missing data in satellite products are not unusual since retrievals are normally at an interval of 2-3 days (Albergel et al., 2012). However, data from each of the different sensor types are also considered for the evaluation of long-term seasonal cycles.

215 ## 2.3 Models for simulating soil moisture

2.3.1 CCAM-CABLE

The variable-resolution atmospheric model CCAM developed by the CSIRO in Australia (McGregor, 2005; McGregor and Dix, 2001, 2008) was used to dynamically downscale ERA reanalysis data to 8 km resolution over north-eastern South Africa (Fig. 1a) for the period 1979-2014. Similar

220 downscaling of reanalysis data obtained over southern Africa using CCAM are described by Engelbrecht et al. (2011), Dedekind et al. (2016) and Horowitz et al. (2017). The ability of the CCAM model to realistically simulate present-day southern African climate has been extensively demonstrated (e.g. Engelbrecht et al., 2015, 2009, 2011; Malherbe et al., 2013; Winsemius et al., 2014). The CABLE soil sub-model expresses soil as a heterogeneous system consisting of three

225 constituent phases namely water, air and solid (Kowalczyk et al., 2006; Wang et al., 2011). Air and water compete for the same pore space, and the change in their volume fractions is due to drainage, precipitation, ET and snowmelt. In this model, there is no heat exchange between the moisture and the soil due to the vertical movement of water, as soil moisture is assumed to be at ground temperature. The soil is partitioned into six layers, with the layer thickness of 0.022 m, 0.058 m, 0.154 m, 1.085 m

230 and 2.875 m from the top layer. Only the top layer contributes to evaporation while plant roots extract water from all layers depending on the soil water availability and the fraction of plant roots in each layer (Wang et al., 2011). Soil moisture is solved numerically using the Richard's equation (Kowalczyk et al., 2006).

2.3.2 GLEAM

235 The Global Land Evaporation Amsterdam Model (GLEAM) version 3.1 is a set of algorithms used to estimate surface, root-zone soil moisture and terrestrial evaporation using satellite forcing data (Martens et al., 2017). The method is based on the use of the Priestley and Taylor (1972) evaporation model, stress module, and the rainfall interception model (Miralles et al., 2011). Three data sets from the GLEAM namely v3a, v3b and v3c were used in this study. The data are freely available at
240 www.gleam.eu. Version 3a is based on satellite observed soil moisture, snow water equivalent and vegetation optical depth, reanalysis radiation and air temperature, and a multi-source precipitation product. Versions 3b and 3c are satellite-based with common forcing data excluding soil moisture and vegetation optical depth, these are based on different passive and active microwave sensors, i.e., ESA CCI for v3b and Soil Moisture and Ocean Salinity (SMOS) for v3c (Martens et al., 2017).

245 The different components of terrestrial processes (i.e., transpiration, open-water evaporation, bare soil evaporation, sublimation and water loss) are separately driven in GLEAM (Martens et al., 2017). Each grid cell in GLEAM contains fractions of four different land cover types namely: open water (e.g. dam, lake), short vegetation (i.e., grass), tall vegetation (i.e., trees) and bare soil. These fractions are based on the global vegetation continuous field product (MOD44B) except for the fraction of open
250 water. The MOD44B product is based on the moderate resolution image spectroradiometer (MODIS) observations (Martens et al., 2017). Soil moisture is estimated separately for each of these fractions and then aggregated to the scale of the pixel based on the fractional cover of each land cover type. Root zone soil moisture is calculated using a multi-layered water balance equation which uses snowmelt and net precipitation as inputs, and drainage and evaporation as outputs (Miralles et al.,
255 2011). The depth of soil moisture is a function of land-cover type comprising one layer of bare soil (0-10 cm), two layers for short vegetation (0-10, 10-100 cm) and three layers for tall vegetation (0-10, 10-100, and 100-250 cm) (Martens et al., 2017).

2.4   Analysis approach and data processing

2.4.1   Statistical analysis

260 The first part of the analysis focuses on evaluating the monthly time series data of soil moisture products at the site level using observations. The monthly temporal scale is considered because, on very short time scales such as daily and hourly, local effects can lead to a pronounced noise in the observations. Such noise, however, is anticipated to be filtered through a long term average. At a monthly time scale, the soil moisture seasonal cycle is well developed. A data threshold of 80 %, i.e.
265 daily values are available for at least 80% of the total number of days in a particular month, was used to average daily data to monthly. Months that did not meet the 80 % threshold were excluded from the analysis. Time series data for the evaluation sites, were extracted from the soil moisture products, using the flux tower's geographical coordinates. The satellite products present averaged soil moisture data per grid cell. A distance-weighted average (DWA) technique was used to interpolate the CCAM-
270 CABLE model simulations to estimate soil moisture values representative of observational sites. The DWA method proved to be more representative than the nearest neighbour (NN) method, as the DWA

method interpolates to the exact location of the tower by considering simulated values at grid points surrounding the location.

The soil moisture products were first converted to the percentage of volumetric soil moisture amounts for comparison purposes. As in Yuan and Quiring (2017), we assume that the soil moisture measurements at the 5 cm depth are representative of the depth range 0–10 cm. In situ data at depths 15, 30 and 40 cm were combined using the depth weighted average method to represent the 10-100 cm depth using Eq. (1):

$$SM_{10-100} = \sum_{i=1}^{n} \frac{LT}{SD} \times SM(i)$$

(1)

Where SM_{10-100} is the weighted soil moisture, n is the number of layers, LT is the layer thickness calculated as the difference between the soil depths, SD is the total soil depth of the soil profile and $SM(i)$ the daily in situ soil moisture values at the ith layer. The depth weighted average method as presented in this study (Eq. 1) has been used in other studies such as that by Yuan and Quiring (2017). Similarly, the data at depths 2.2 and 5.8 cm, and 15.4 and 40.9 cm from CCAM-CABLE are averaged to represent 0-10 and 10-100 cm respectively using Eq. (1).

Table 1. Overview of soil moisture datasets; satellite (grey) in percentage, modelled (blue), simulation (pink) and in situ observations (green) presented as a ratio ($m^3\,m^{-3}$) of soil to moisture per unit area.

Soil moisture product	Spatial resolution (km)	Spatial coverage	Soil depth (cm)	Period
ESA-Combined	25	Global	0-10	1978-2015
ESA-Active	25	Global	0-10	1991-2015
ESA-Passive	25	Global	0-10	1978-2015
CCAM-CABLE	8	Regional	2.2, 5.8, 15.4, 40.9, 108. 5, 287.2 (bedrock)	2000-2014
Skukuza	Point data	Point	5, 15, 30, 40	2000-2017
Malopeni	Point data	Point	5, 15, 30, 40	2008-2017
GLEAM v3a	25	Global	0-10, 10-100	1980-2016
GLEAM v3b	25	Global	0-10, 10-100	2003-2015
GLEAM v3c	25	Global	0-10, 10-100	2011-2015

The soil moisture products used in this study (Table. 1) are under the same latitude and longitude projection. All the soil moisture projections are at the same spatial resolution of 25 km, except for the CCAM-CABLE model with a resolution of 8 km. The bilinear interpolation method was used to resample the CCAM-CABLE simulations from 8 to 25 km to match the resolution of the other soil

moisture products. To evaluate how close the modelled soil moisture estimates are to in situ measurements using the Taylor plots (Taylor, 2001) as well as the cross-wavelet analysis.

2.4.2 Cross-wavelet analysis

The cross-wavelet method analyse the frequency structure of a bivariate time series using the Morlet wavelet (Veleda et al., 2012). The wavelet method is suitable for analysing periodic phenomena of time series data, especially in situations where there is potential for frequency changes over time (Rosch and Schmidbauer, 2018; Torrence and Compo, 1998). The cross-wavelet analysis provides suitable tools to compare the frequency components of two time-series, thereby concluding their synchronicity at a given period and time. In this study, the cross-wavelet analysis is used to qualitatively compare the cyclic patters of the observations and the models' estimates. In particular, it is used to assess if there exist phase differences between dominating periodic features of the in situ observations and the models' estimates. The cross-wavelet analysis algorithm used is described in Rosch and Schmidbauer (2018) and is implemented within the "WaveletComp" package in the R software. This method has been used in other studies, such as that by Koirala and Gentry (2012), for investigating the climate change impacts on hydrologic response.

The cross-wavelet analysis only applies to complete datasets (i.e., without missing values). Since the in situ observations have missing data, the multiple imputations method as discussed in studies by Rubin (1987) and Rubin (1996) has been used to gap-fill the in situ time series. The multiple imputation procedure is implemented in the "Amelia" package also available in the standard repository for R packages. The number of imputed datasets was set to five and combined using Rubin's rules (Rubin, 1996). The multiple imputations method is only applied to the Skukuza dataset both the surface (Fig. C1.a) and root zone (Fig. C1.b). This is because the Skukuza data has fewer gaps compared to Malopeni (Fig. A1.b). The imputed soil moisture observations are shown in Appendix C together with the statistics of the measures of the distribution for both the gap filled and non-gap filled datasets. The cross-wavelet analysis is applied to non-stationary data using the default method (i.e., white noise) with the simulations repeated ten times.

2.4.3 Mutual information

The second part of the analysis inter-compares model simulations and satellite estimates of soil moisture at a regional scale. The MI is calculated between the residuals of the de-trended and de-seasonalised time series at a regional scale between the CCAM-CABLE simulations and GLEAM estimates. The de-trended and de-seasonalised procedure before the calculation of the MI is done to ensure that the computed MI is not attributed to the similarities in the trend and cyclic components of the signal, which may be correlated, but it is based on the residual components which are the uncorrelated features of the soil moisture signal. In this way, the MI calculation presents a comparison matrix for inter-model soil moisture spatial pattern comparison. In particular, the MI gives a sense of similarity between the models indicating the level of coincidence or overlap in the distribution of the residuals between a pair of CCAM-CABLE simulations and each of the GLEAM model estimates per grid point. In the case that MI values between models are low the inter-model reflects uncertainty in how the models capture the modelled processes.

The de-trending and de-seasonalising of the time series removes the systematic components of the signal including bias. This is achieved through an approach reported in a study by Cleveland et al. (1990) where the "stl" package, available in the standard package repository in R, is used to de-trend the time series into its components. The MI calculation is described in Kraskov et al. (2004) and is

applied in this study using the "varrank" package which is also available in the R CRAN repository. The MI measure calculated from the residual components of the respective soil moisture signals presents a robust way of assessing if the respective models have a correspondence in spatial patterns of soil moisture across landscapes. In this paper, the MI is used as an index for classification of the models according to the coincidence in the distribution of residuals at regional level. The MI is calculated for the time series ranging between 2011 and 2014 using a de-trended and de-seasonalised time series with no systematic components including bias.

3 Results and discussion

3.1 Evaluation of the satellite-estimated and model-simulated seasonal cycle soil moisture

In this section, we discuss how the respective outputs reflect the key features of the observed soil moisture. As highlighted in the introduction, the variability of the simulation output, satellite-derived data and satellite-based model estimates are studied relative to the observations. Much focus is placed on investigating how well the periodic features of the soil moisture are reflected by the respective soil moisture datasets. The patterns of soil moisture at the study sites are mainly driven by rainfall, which is generally higher during the summer season, and low in winter as shown in Fig. 2. The long term surface soil moisture for both the sites follows a pattern comparable to that of rainfall as can be seen in Fig. 2.

3.1.1 Long term seasonal cycles

The soil moisture patterns presented in Fig. 2 show that the study sites mainly contain higher soil moisture at the surface than at the root zone, this is shown by both the modelled soil moisture and the observation. This is indicative that, water at these study sites is predominantly lost through runoff and ET, and only a small fraction infiltrates the soil as is stored at the root zone. In general, there is an agreement on the seasonal cycle of soil moisture (Fig. 2) between the various product outputs and the observations in terms of phase, especially at the surface. Notably, the observed soil moisture seasonal cycle at the surface at both Skukuza and Malopeni surface displays a local maximum in April and shows and increases from September to December and January. The general features of the observed signal are captured by all the models. The soil moisture amplitudes are less pronounced in the root zone, but with November and October maxima at Skukuza and Malopeni respectively. In some instances, there is a lag such as the one presented by GLEAM v3c (i.e., maxima in October instead of November) at the surface, both at Skukuza and Malopeni. The soil moisture patterns are consistent with the observed rainfall cycle which undergoes an onset in October and a cessation in April. The root-level soil-moisture pattern displays a signature of soil moisture retention, which relates to the persistence of dry and wet periods at various soil depths (Seneviratne et al., 2006). In light of this, it would be interesting to see how both the CCAM-CABLE simulation and the GLEAM soil moisture products depict the onset and cessation of the wet season, this will be discussed in Section 3.2. The CCAM-CABLE model outputs reflect that soil-moisture reaches its highest values in March rather than April for Skukuza at the surface. The output does not reproduce the recorded elevated soil moisture for Malopeni in April at the surface. This is probably since the CABLE soil-moisture scheme does not take into account soil resistance (Whitley et al., 2016). Despite this, the long term CCAM-CABLE monthly means of soil moisture are relatively comparable to the observation even in terms of magnitude (Fig. 2).

GLEAM v3c, agrees with in situ measurements on the existence of an April soil moisture maximum, but it reflects the observed soil moisture increase, in November, a month earlier (i.e., in October). The satellite observations and GLEAM models (Fig. 2) display the same soil moisture signal as observed at the respective sites, indicating that the April maximum, in particular, is not an artefact of the point observations. We can safely deduce that the bias in GLEAM v3c is not induced by satellite-based forcing data. However, this calls for further investigations on the sensitivity of the model to its driving data at a high resolution. We anticipate that at high temporal resolution there is a strong variability in the in situ soil moisture signal which may not entirely be captured by both CCAM-CABLE and GLEAM, possibly due to their relatively low spatial resolution. The relatively low resolution (8 km in the horizontal) in the case of CCAM-CABLE, in particular, potentially has strong implications on how representative the effective drivers of soil moisture such as soil texture and vegetation covers are in terms of observations at specific sites.

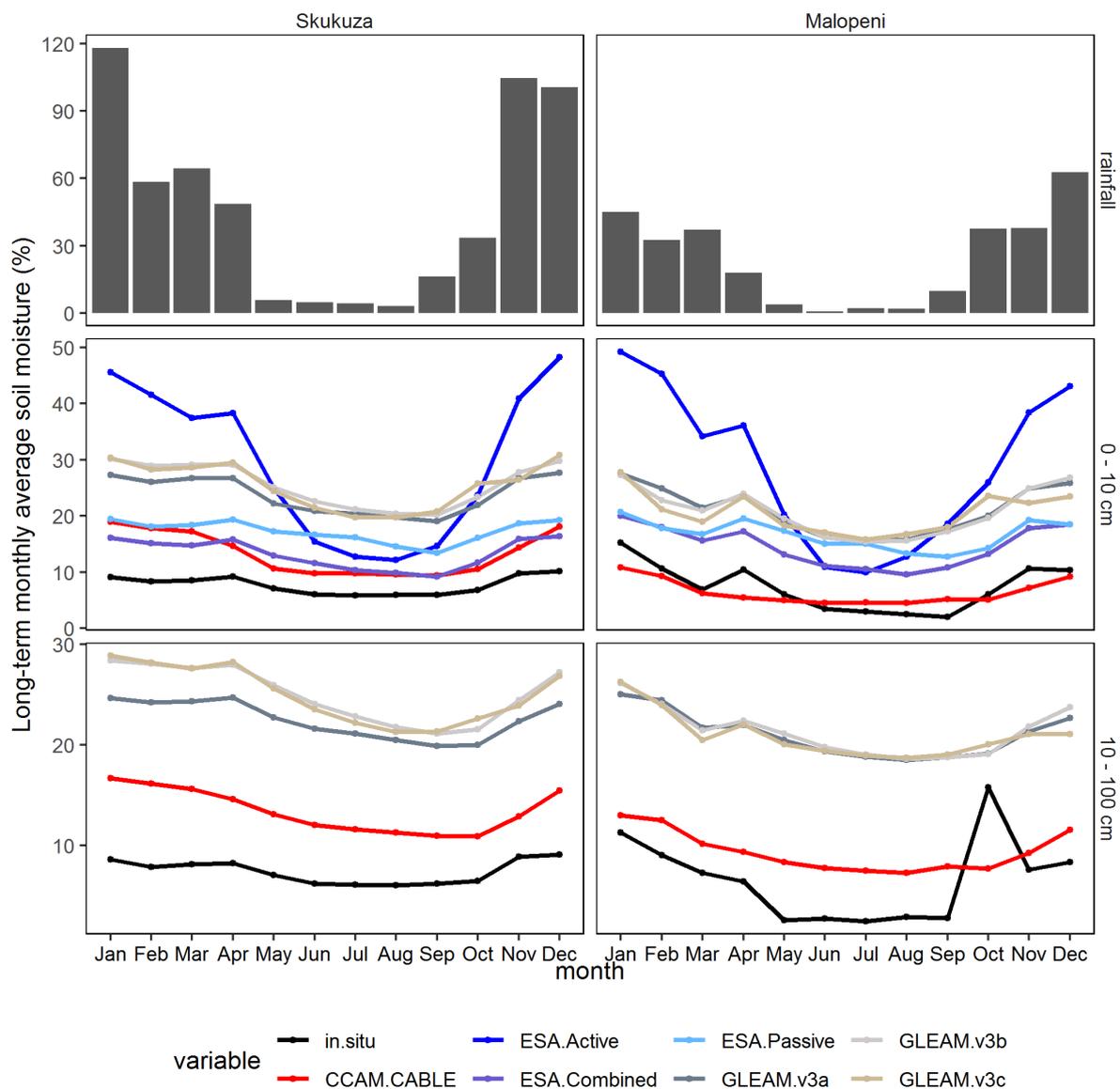



Figure 2. Seasonal variation in the long term mean monthly rainfall (mm), surface (i.e., 0-10 cm) and root zone (i.e., 10-100 cm) soil moisture, based on in situ observations and a variety of soil moisture products. The in situ data is collected from two sites, namely Skukuza (2001-2014) and Malopeni (2008-2013).

The GLEAM models (Fig. 2) are generally consistent with in situ measurements in estimating soil
moisture both in terms of phase, both at the surface and root zone. The magnitude of GLEAM v3a
root zone estimates is lower than those of the other GLEAM models at the Skukuza site. This can be
attributed to the unique multi-source weighted ensemble precipitation (MSWEP) data used to force
GLEAM v3a (Martens et al., 2017), which is different to the precipitation forcing data used in
GLEAM v3b and v3c. We further observe that the GLEAM models, ESA and in situ observations
have the same length of the dry period (i.e., about 4 months), except for the ESA-Active observation
which has a shorter dry period (i.e., about 3 months).

The ESA-Active satellite product is known to work best for moderate to densely vegetated areas as
opposed to savanna sites such as Skukuza and Malopeni where tree cover is sparse (Dorigo et al.,
2015). There is a minimal difference between the ESA-Passive and ESA-Combined satellite products
both in terms of phase and magnitude. Generally, the ESA-Combined and ESA-Passive datasets have
the least difference during the dry period for all sites. A number of studies evaluated the ESA
products at a regional and global scale using in situ data and concluded that passive sensors displayed
improved performance over bare to sparsely vegetated regions, whereas the active sensors perform
better in moderately vegetated regions (Al-Yaari et al., 2014; Dorigo et al., 2015; Liu et al., 2012;
McNally et al., 2016).

Using long term monthly averages, both the CCAM-CABLE and GLEAM models can capture the
intrinsic seasonality of the soil moisture signal for the sites as reflected by both the in situ and satellite
observations. This is despite their being different both in the forcing data and model structure. Studies
by Wang and Franz (2017) and Seneviratne et al. (2010) suggest that local factors (e.g., vegetation,
soil and topography) mostly control soil moisture variability at spatial scales less than 20 km, rather
than meteorological forcing. For a fourteen-year averaging period, undoubtedly the monthly means
are sensitive to anomalously high precipitation, and hence soil moisture in some months. It is
therefore instructive to investigate how well the simulated and estimated patterns of soil moisture
compared with the in situ data monthly for the respective years.

3.1.2 Intra- and inter-annual variability in soil moisture

This section presents a quantitative evaluation of the soil moisture time-series from the soil moisture
products at a monthly time-resolution. The agreement of the short term seasonal cycles between the
various outputs and observations are quantified in Fig. 3 using the Taylor plot. The Taylor plot
present three evaluation metrics namely; 1) the standard evaluation, which evaluates the amplitudes of
the modelled soil moisture relative to the observations, 2) the centred root mean squared error
(RMSE) measuring the distance in magnitude between the various products and the observations, and
3) the correlation coefficient measuring the agreement in phase. For the Skukuza site, we learn that
the standard deviation for the surface soil moisture observation is around 4.5 %. The standard
deviation values for the surface time series of the various soil moisture products are mostly within a
range that is close to that of in situ observations. In particular, the GLEAM products mainly present
similar standard deviations with the observations, while the CCAM-CABLE and ESA-Combined
products show standard deviation values slightly lower than that of the observations indicating a slight
underestimation by these products. At the root zone the soil moisture standard deviation is relatively
lower (i.e., about 1.5 %) for the observations while all other soil moisture projects reflect much higher
standard deviation indicating an overestimation of the root zones soil moisture by these products.
Based on the correlation coefficient, we learn that there is a higher correlation between the observed
and modelled soil moisture products at the surface ranging between 0.7 and 0.9. At the root zone, the

correlation coefficients, for the site, range between 0.6 and 0.8. This indicates that there is more agreement in phase at the surface compared to the root zone.

440   At Malopeni, we learn that the standard deviation for observed soil moisture values is about 4.7 % at the surface and 3.2 % at the root zone. In both cases the models present a standard deviation with a range closer to that of the observed root zone values. For this particular site, the agreement between the various products and the observations is more pronounced at the root zone (RMSE ranges between 1.8 and 2.3 %) than at the surface (RMSE ranges between 2.1 and 3.5 %).

445   On the basis of comparisons of standard deviations, we can conclude that the pattern variations for different soil moisture products are of the right amplitude at the surface for the two respective sites. The amplitude of the pattern of variation among observations is relatively lower at root zone, this is consistent with that of the models at Malopeni but not Skukuza. The disparity in the amplitude of variation at Skukuza, is indicative that there is a need for further investigation on the homogeneity

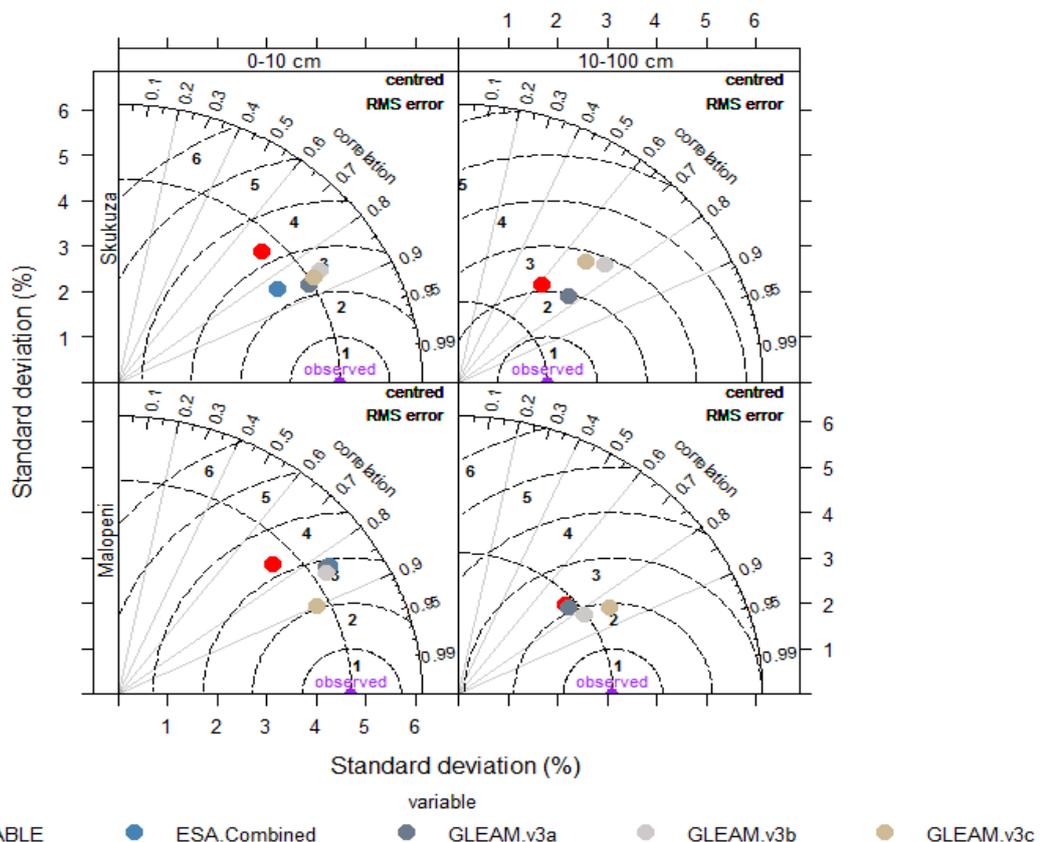
450   assumptions for the site when it comes to the dominating drivers of soil moisture. The highest agreement with in situ data is observed for the GLEAM and ESA products at the surface. Higher RMSE values for the CCAM-CABLE model at the surface, might be indicative that in some instances the modelled signal is shifted relative to the observed signal in some years. This will be investigated in detail in the next section.

455

Figure 3. Taylor plots quantitatively comparing monthly modelled soil moisture to observations at Skukuza (2001–2014) and Malopeni (2008–2013), both at the surface (0-10 cm) and root zone (10-100 cm). The vertical solid grey lines represent the correlation coefficient. The broken black line facing the clock-wise direction represents the standard deviation of the in situ observation. While, the semi-circle broken lines represent the

460   centred root mean squared error.

The ESA-combined satellite product presents a similar performance with the GLEAM products at Skukuza and Malopeni. The ESA data has been shown to generally capture soil moisture in different regions and climatic zones of the world (Loew et al., 2013; McNally et al., 2016; Wang et al., 2016; Zeng et al., 2015). Our study confirmed (Fig. 3) that the ESA-combined product captures local (i.e., South African semi-arid) conditions within an acceptable amount of certainty. A study conducted by Yuan and Quiring (2017) assessing the performance of CMIP5 models both at the surface and root zone, concluded that the models performed better at the root zone relative to the surface. These results contradict the findings of this study, where we generally observe better agreement between soil moisture products and in situ measurements at the surface than at the root zone. Based on the general picture of the extent to which the soil moisture products proved to be representative of the quantitative features of the soil moisture signal, as driven by precipitation at the site, it is compelling to further resolve qualitatively, for each year, how the various soil moisture outputs compare with the in situ observations. To this effect, the next section will present the results from a cross-wavelet analysis of the soil moisture output and the in situ observations.

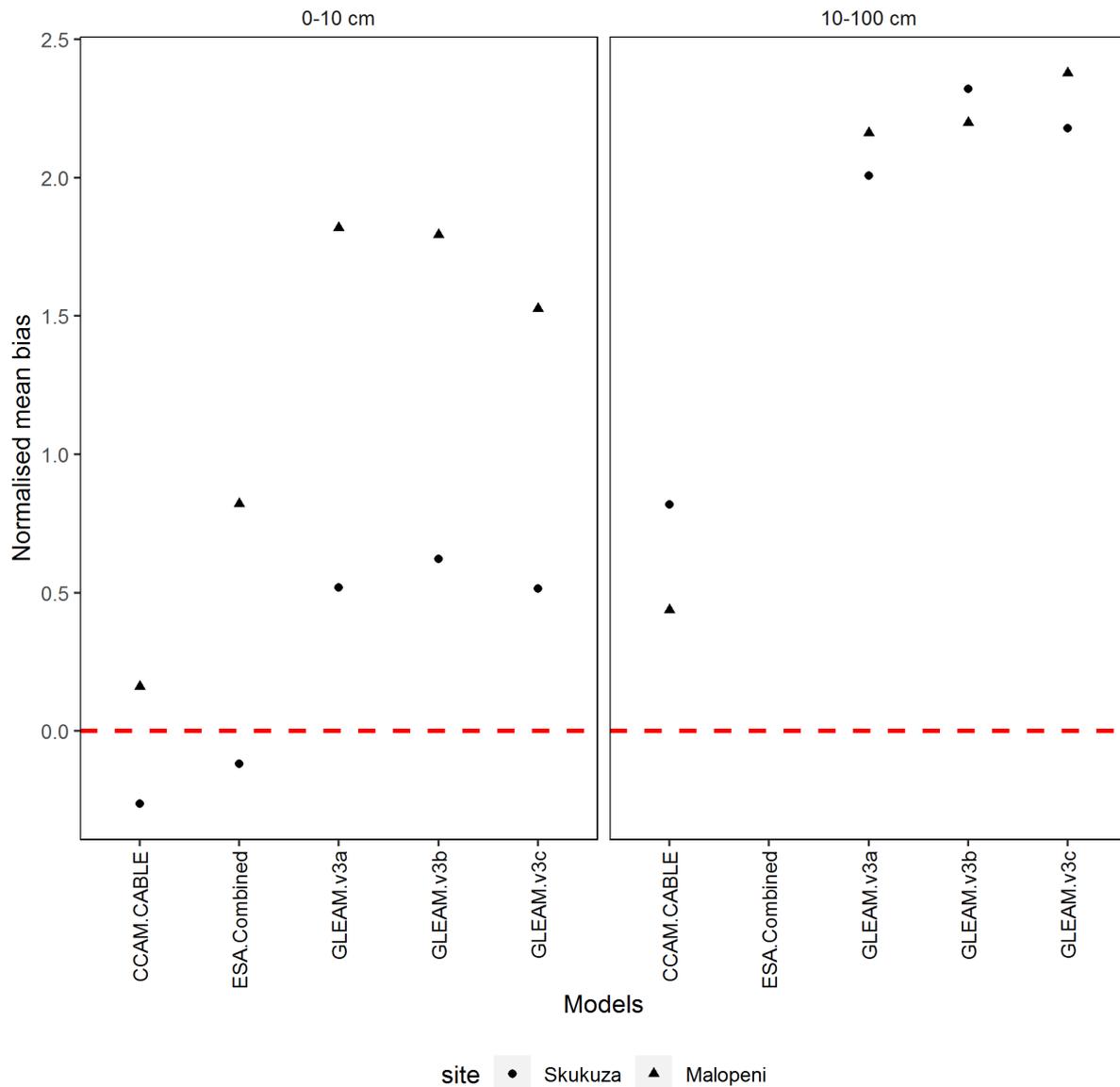



Figure 4. Normalised mean bias (NMB) of surface (0-10 cm) and root zone (10-100 cm) soil moisture, computed between the various soil moisture products and the in situ observations at Skukuza and Malopeni.

Figure 4 shows the normalised mean bias (NMB) computed between the various soil moisture products and the in situ observations. We learn from Fig. 4 that the models are mostly biased towards an overestimation (i.e., values above the horizontal line) of the observed soil moisture. The overestimation is more pronounced at the root zone relative to the surface. This is mostly true both at Skukuza and Malopeni. We also learn that the models mainly present higher overestimation bias at Malopeni compared to Skukuza. The GLEAM and ESA-Combined products predominantly show higher bias towards overestimation compared to the CCAM-CABLE model. The CCAM-CABLE model shows the least bias relative to the other soil moisture products both at the surface and the root zone. At the Skukuza site, the CCAM-CABLE and ESA-Combined products show and underestimation of the observe soil moisture at the surface. These results are consistent with those of the standard deviation presented in Fig. 3.

3.1.3 Cross-wavelet analysis

In this section, a cross-wavelet transform (CWT) constructed from two continuous wavelet transforms (CWT) applied to the modelled and observed time series respectively is studied. The CWT is instrumental in depicting the relationship in time and frequency space between two time-series. This is achieved by analysing localised intermittent oscillations in the respective time series. By looking at the regions in time and frequency space with relatively large common power (represented by red colours; Fig. 5) and a consistent phase relationship (depicted by arrows), we gain a sense of whether there is a physical relationship between the observed and modelled soil moisture fields. Looking at Fig. 5 we learn that the soil moisture signals components with a common power are immediately identifiable and are portrayed as having periods (y-axis values) that lie between 8 and 15 months. This is depicted by strong red colour regions bound by white lines, which marks the region with 10% significance level (i.e., 90 % confidence level). On comparing the surface and root zone cross-wavelets, we can conclude that the statistically significance cyclic components with the dominating common power are generally between the periods of 8 and 15 months. This can be associated with seasonal soil moisture variation as driven by meteorological drivers, most of which have a return period of about a year.

From Fig. 5a we can see based on the alignment of the arrow (Fig. B1) that the most common high power signals between modelled and observed data are in phase, in some instances with a time lag. This is identified by the direction of the arrows which are inclined either upwards or downwards, See Fig. B1 in Appendix B for an interpretation of the direction of the arrows. From the graph of the phase difference, we can see that there is an interchange of years in which the modelled field are leading or lagging in phase however, the phase difference is mostly very small. There is a time lag of two days on average between CCAM-CABLE simulations and in situ observations at the period of about 12 months, and a lag of about six days on average between GLEAM v3a and the in situ observations at the surface. At the root zone, we observe a wider lag of between 14 and 24 days between the soil moisture products (i.e. CCAM-CABLE and GLEAM-v3a) and the observations. This further confirms that there is a better agreement between the soil moisture products and the observation at the surface than at the root zone.

In all models, precipitation is a source of soil moisture at the surface while heat and wind are sinks of moisture from the surface. As mentioned earlier the models introduce different assumptions about dominating drivers of root zone soil moisture for instance, which may potentially explain the

520 existence of broader time lags at the root zone. We further observe, in Fig. 5, that there is an agreement between the models and observations on the seasonal and intra-annual signal of soil moisture at Skukuza, this is shown by orange depicted regions on the cross-wavelet graphs. These are the signal components mainly ranging between the periods of two to six months. This could be associated with anomalous years where the transition periods between the austral winter and summer

525 may have months with below (dry) or above (wet) normal soil moisture conditions. Despite these periods having a relatively high common power, they are not demarcated as statistically significant.

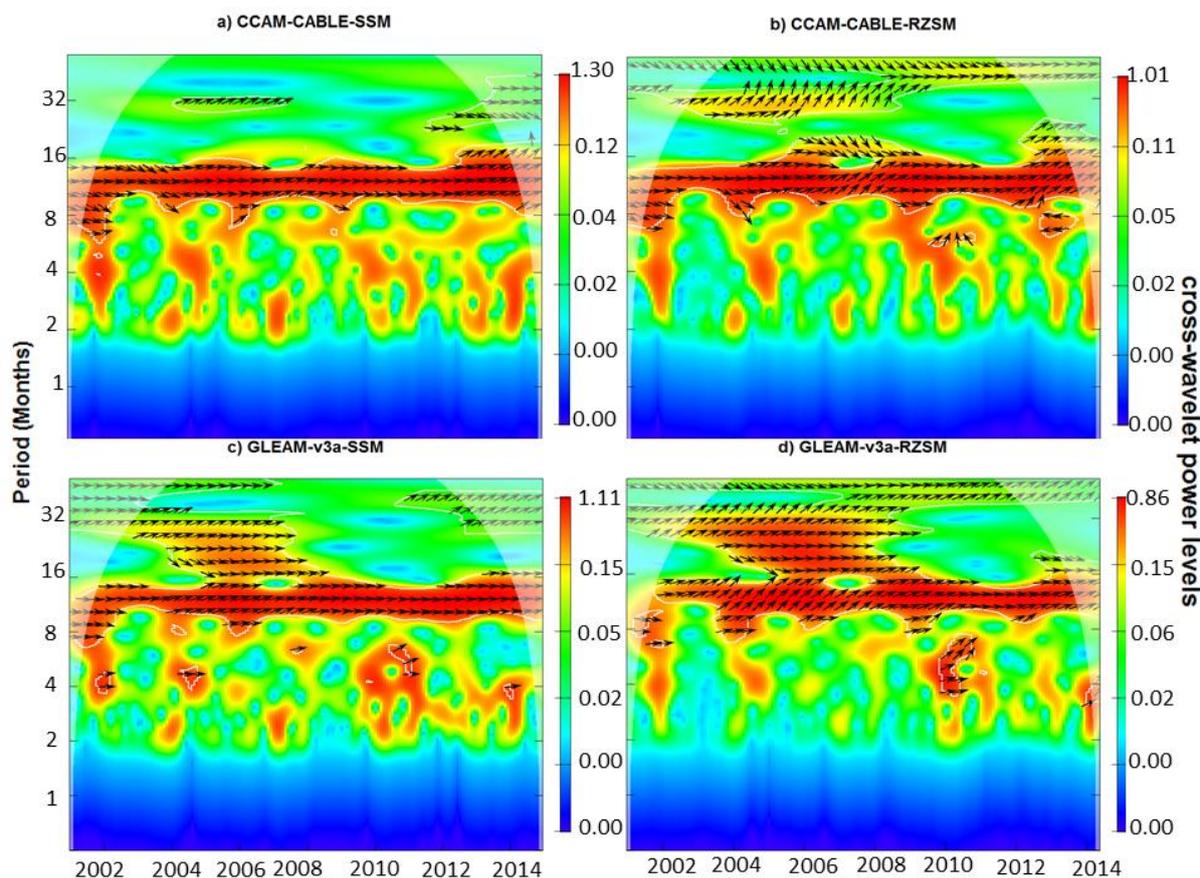



Figure 5. Cross wavelet power spectrum of surface (SSM, 0-10 cm) and root zone (RZSM, 10-100 cm) soil moisture between in situ observations, CCAM-CABLE (a, b) and GLEAM v3a (c, d) at Skukuza respectively.

530 The white contour lines indicate periods of significance at 10 %. The arrows pointing to the right indicates that the models and in situ observation are in phase while arrows point left reflecting that the models are anti-phase. The case where in situ observations are leading either CCAM-CABLE or GLEAM v3a is indicated by arrows pointing straight down. The dome shape (shaded areas) represents the cone of influence between 2001 and 2014. The red colour indicates weak variation while blue indicates the strong variation between the respective time

535 series.

It would be interesting to establish how the qualitative insight gained in understanding the models' ability to capture the observed soil moisture signal at the two respective sites will translate to a regional level. An upscaling of the evaluation done at a point is not possible in the absence of site observation at a regional level. The rest of the discussion in this paper is dedicated to an inter-

540 comparison of process-based model outputs and satellite-derived model outputs. The idea is to discuss the model outputs in connection with the broader landscapes classes within the region.

3.2   Linking soil moisture patterns to landscapes

 So far we have investigated the capabilities of the models in capturing the temporal features of soil
moisture at the flux tower sites. An interesting question to address is, to what extent do the respective
models compare in capturing soil moisture organisation across different landscapes as characterised
by altitude range, climatic zone, dominant soil, biome types and slope aspect within the considered
25km resolution. In the case where there are no in situ soil moisture fields, we may not reliably tell
which product is the most representative of the soil moisture organisation, however, we can classify
the models on the basis of their shared patterns at the selected landscapes. We consider six sub-
regions which are numbered (a) to (f) as depicted in (Fig. 1c), whose characteristics soil, vegetation
and climate features are fully described in Table 2. The vegetation types for the study area used here
are presented in a study by Khosa et al. (2019)

Table 2. A detailed description of the selected sub-regions (Fig. 1c) indicating elevation
(https://www.ngdc.noaa.gov/mgg/fliers/06mgg01.html), climate (http://koeppen-geiger.vu-
wien.ac.at/present.htm), vegetation and soil types (https://soilgrids.org).

Zone	Elevation range (m)	Climate type	Dominant vegetation type(s)	Dominant soil types
a	0-823	Oceanic	Savanna	Clay
b	824-1298	Humid	Savanna	Clay
c	1299-1317	Hot semi-arid	Savanna	Clay
d	1318-1322	Hot semi-arid	Grassland	Silt-loam
e	1323-1658	Cold semi-arid	Grassland	Silt-loam
f	1659-2440	Cold semi-arid	Grassland	Silt-loam

In Fig. 6 we see moisture pattern for austral summer (DJF) for the year 2011 to 2014. The violin plots
summarise daily data points that are separated by a five days interval throughout the season for the
respective models. From Fig. 6 we learn that the distribution of soil moisture values for three GLEAM
models fall within an overlapping range. This suggests that the spread of soil moisture values for the
GLEAM models is relatively close.

By comparing the spread and the mode (i.e., thicker parts of the violin) of soil moisture values across
models, we can conclude that for the region (a), which is characterised by predominance of clay soil
and relatively low elevation range, there is no clear variation of soil moisture spread that could be
associated with models or the respective south- and east-facing slopes. In summary, the distribution of
soil moisture values is comparable amongst the models and slopes. On the humid (b) and hot semi-
arid (c, d) regions, we learn that the soil moisture spread is comparable between CCAM-CABLE and
ESA but relatively lower to that of the GLEAM models. It is worth reiterating at this point that
GLEAM models also show higher soil moisture values relative to the in situ observations at the
Malopeni at Skukuza flux tower sites, which share the same elevation range and climate type as
region (b). For the three landscapes, there is no clear pattern which distinguishes the organization of
soil moisture according to slope direction. In the case where a landscape has slopes facing more than
one direction, highly overlapping distributions indicate that topographic and thermal drivers of soil
moisture do not lead to dominant or identifiable soil moisture patterns among the respective models.

For the cold and high lying semi-arid region, (e) and (f), CCAM-CABLE shows a noticeable variation
in soil moisture with slope aspect, in which case north facing slopes turns out to have lower soil
moisture than South- and west-facing ones. For the north-facing slopes of the two regions, the
relatively lower soil moisture values for CCAM-CABLE are corroborated by that of the ESA-

combined model, which generally portray comparatively low soil moisture values for the two high-lying cold semi-arid regions. It is a well-known fact that along the Drakensberg range, which is close the regions (e) and (f), north- and east-facing slopes have more sunshine exposure than the south- and west-facing slopes (Bristow, n.d.). Notably, the CCAM-CABLE, ESA-Combined and GLEAM models reflect contrasting patterns with slope-aspect for the high-lying areas. This calls for model evaluation against observations. Such an evaluation can potentially yield valuable information on which model assumptions or schemes can benefit from further refinements, taking into account dominant thermal and slope dependent soil moisture processes for the landscape.

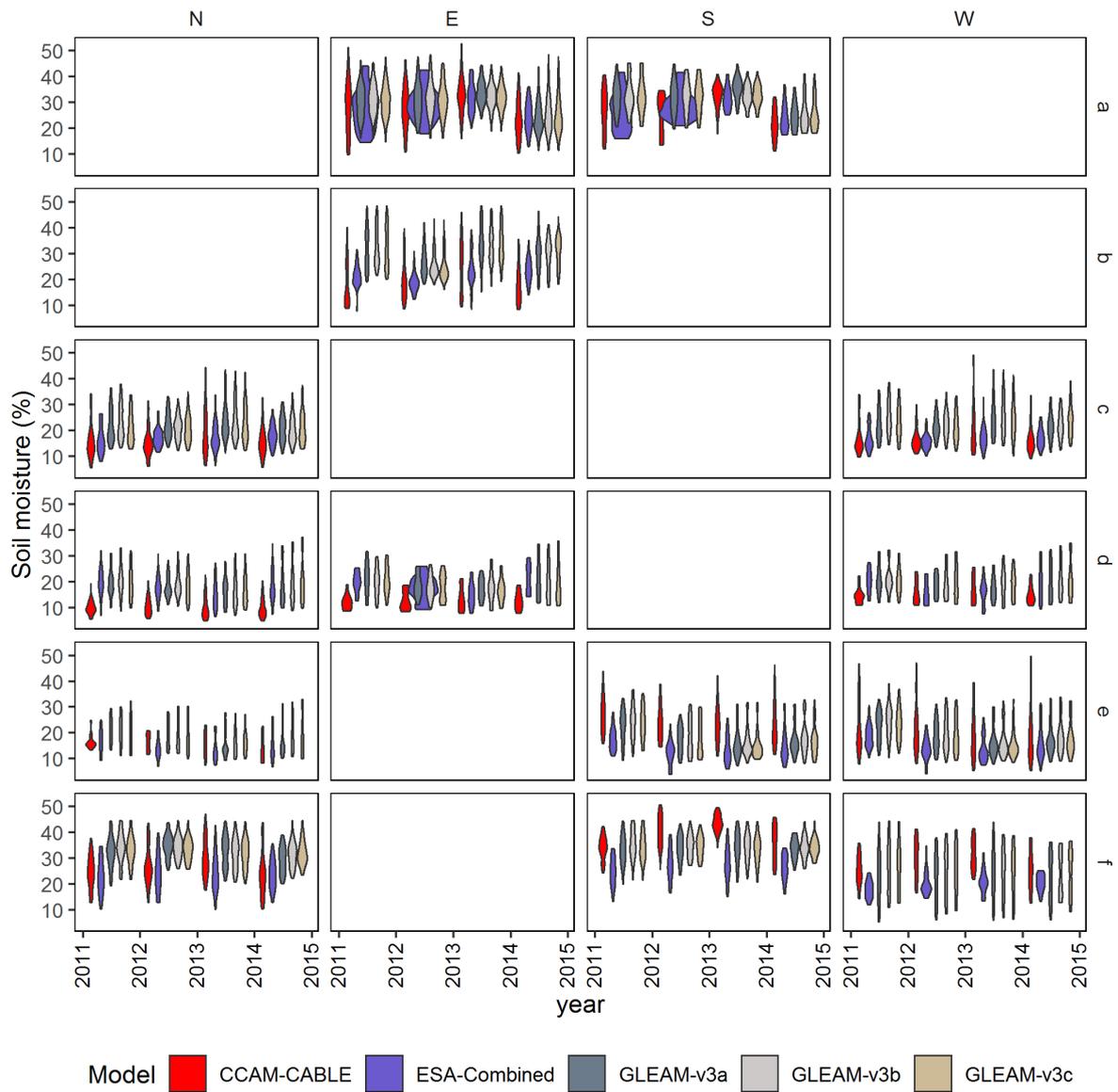



Figure 6. Comparison of modelled soil moisture patterns across sub-regions (a-f) of increasing altitude, slopes of different aspect i.e., north (N), east (E), south (S) and west (W). The violin plots show the distribution shape of the soil moisture values per model within the respective regions.

For the selected landscapes, we have learn that the three GLEAM models mostly reflect a spread of soil moisture values which is largely overlapping, while CCAM-CABLE shows the existence of distinct moisture distributions that can be associated with slope aspect especially on high lying regions. A clear continuous picture of how CCAM-CABLE compare with GLEAM models across the

595   entire study domain can be obtained by investigating how different joint distributions of a pair of
       CCAM-CABLE and GLEAM residuals per grid point compare to a product of their marginal
       distributions. This is best quantified by the MI, which is an information theory function that can be
       used as a measure of similarity between a pair of time series of residuals. The compared time series
       are computed on a common grid point for the respective models. The MI is equal to zero when the
600   joint distribution of the pair coincides with the product of the marginal for the respective models. This
       suggests that the respective models are portraying independent signals. For the studied datasets we expect
       that the MI values should be greater or equal to 2 in the extreme case when the two pairs are identical.
       Figure 6 depicts the MI which is calculated from a pair of de-trended and de-seasonalised time series
       of monthly averaged soil moisture for CCAM-CABLE and each of the three versions of GLEAM.
605   The de-trending and de-seasonalising of each pair also lead to a removal of systematic biases. The
       obtained MI values are mostly equal to or greater than 0.5. This is true for both the surface and root
       zone. It is desirable to have the MI for all satellite data derived products, however, the ESA products
       did not have enough spatial data points to yield a fair comparison.

       We can also see in Fig. 7 that the MI at the root zone is higher than at the surface, this could be
610   suggestive that, the sensitivity of soil moisture to the driving processes is comparable between both
       GLEAM and CCAM-CABLE models at the root zone. The MI pattern for both the surface and root
       zone compliment the box-and-whisker plot, indicating that the coincidence in the soil moisture values
       is highest in the proximity of the lowest lying oceanic savanna region (a) which is dominated by the
       clay soil. For this region, the MI values mainly ranges between 1 and 2. CCAM-CABLE has been
615   depicted as having low soil moisture values relative to all versions of GLEAM, on part of the humid
       savanna region (b) for the surface. We can also see that, on the humid savanna which include region
       (b), that the models predominantly have low MI values ranging between 0 and 1 at the surface. The
       lowest MI values at the surface are also noticeable on the cold semi-arid high-lying grasslands in the
       neighbourhood of regions (e) and (f). From Fig. 7, we can conclude that the study region is dominated
620   by grid points with relatively high MI values that fall within the range [0.5 - 2]. Lower MI values for
       the high-lying regions are indicative that, there is a pronounced model uncertainty when it comes to
       the models response to processes that drive soil moisture for the region. While higher MI values, as
       seen on the rest of the regions, gives an indication that the respective models comparably responds to
       the dominating processes that drive soil moisture variation. This is the case at least qualitatively

625

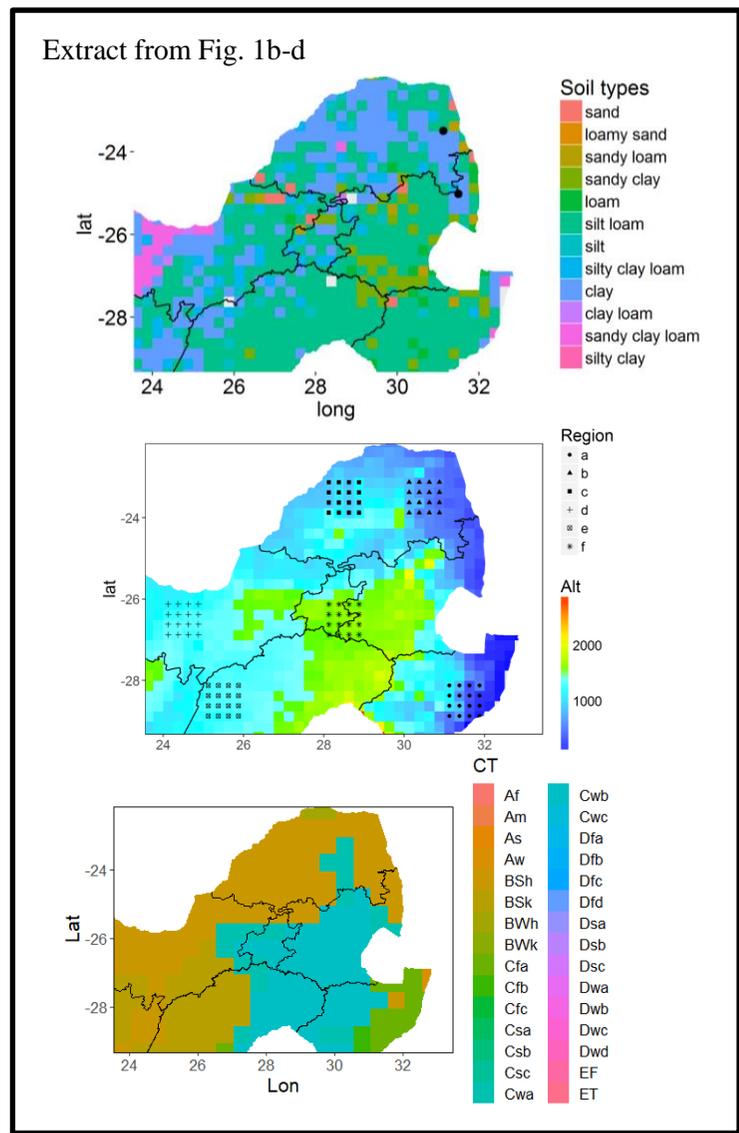
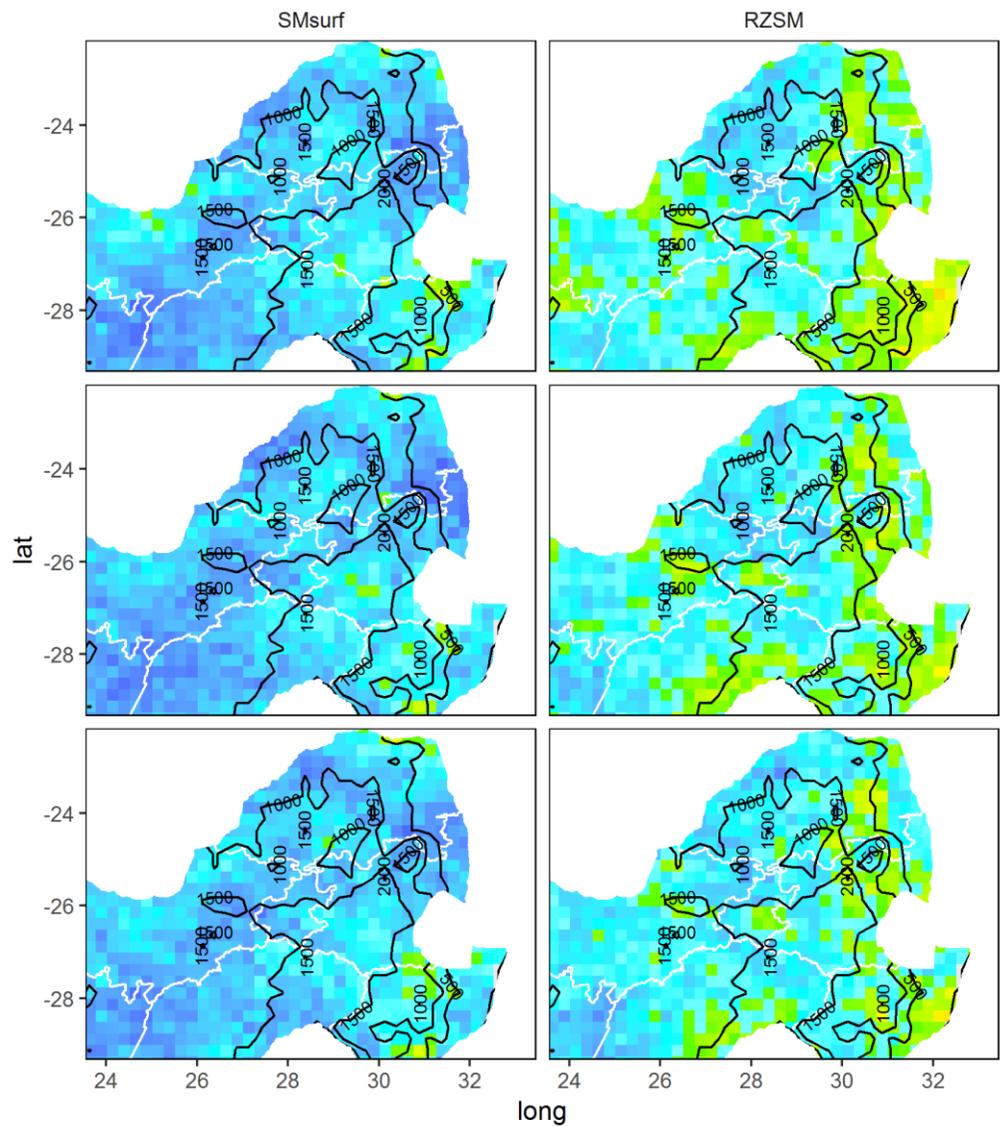



Figure 7. Mutual information (MI) computed on the residuals of monthly time series (2011-2014) of surface (SMsurf, 0-10 cm) and root zone (RZSM, 10-100 cm) soil moisture, between CCAM-CABLE simulations and GLEAM models estimates. The contour lines represent the elevation (Fig. 1c) of the study area, while the white lines represent the provincial boundaries. An extract from Fig. 1b-d is shown at the right hand side of the figure to aid with the interpretation of the MI.

4   Conclusions

In this study, the ability of a process-based simulation model (CCAM-CABLE), satellite data-driven model estimates (GLEAM) and satellite observations (ESA-Active, -Passive and -Combined) are evaluated against site-specific in situ observations from two flux tower sites namely, Skukuza and Malopeni. The evaluation was done for two soil depths namely the surface (i.e., 0-10 cm) and root zone soil moisture (i.e. 10-100 cm), to understand how the respective data products capture the characteristic patterns of soil moisture. The evaluation included an assessment of qualitative features of long term (i.e. multi-year), and short term (i.e., monthly) averages of the soil moisture signal relative to the in situ measurements. We generally learn that all GLEAM soil moisture products at all depths presented higher soil moisture magnitude range compared to observations while CCAM-CABLE and ESA-combined outputs turn out to be relatively closer in magnitude to the observation both at Malopeni and Skukuza. The systematic difference in magnitude between the model output and observation may emanate from the difference in spatial scale between in situ measurements and the rest of the products. We also learn from this study that all GLEAM models compare well with the in situ observations in reflecting the seasonality of soil moisture. This is despite the noted systematic bias of the soil moisture magnitudes in the GLEAM products. The models mostly show a bias towards overestimation of the observed soil moisture both at the surface and root zone, with the CCAM-CABLE showing the least bias.

A wavelet analysis was used to reveal, at a qualitative level, how periodic features, compare between the CCAM-CABLE, GLEAM models and in situ observations. We learned that at the surface, high power common features of the surface soil moisture signal are in phase with observations and come at a periodicity of about 12 months. We also learned that high power common soil moisture signals at the root zone have a relatively pronounced time lag. The time lag is of a time scale not exceeding a month at all soil depths (i.e., it lies between 5 and 20 days) for the periods ranging between 2001 and 2014) between CCAM-CABLE and GLEAM v3a.

The study also investigated through the use of mutual information (MI), how different joint distributions of pairs of grid points, among CCAM-CABLE and the respective GLEAM models compare, with a product of their marginal distributions. This gave a basis for classifying the models according to their similarity or dependence in capturing soil moisture responses to the underlying drivers. In this case, the emphasis is on evaluating the extent to which both approaches have a joint variation or shared MI. The analysis has successfully revealed that both the simulation and model estimates have a high similarity at the root zone as opposed to the surface for all GLEAM model outputs. The difference in the surface soil moisture between the CCAM-CABLE simulation and GLEAM models outputs at high lying areas, opens-up interesting questions relating to the extent to which the influence of different drivers of soil moisture is represented by the two approaches. To understand this, future research will benefit from investigating the sensitivity of the models to changes in soil moisture drivers, particularly change in vegetation cover and soil type on soil moisture memory. It would also be interesting to unearth the soil moisture organisation for the respective models, at much higher spatial resolution where processes that drive soil moisture may reliably be attributed to the patterns on the soil moisture signal. Despite CCAM-CABLE and GLEAM having relatively high MI for the majority of landscapes, application of these model outputs should take into

account that systematic biases do exist, and that there is a high model uncertainty particularly at high lying areas.

Team list and Author contribution:

- Floyd – Developed research questions, analysed the data and compiled the manuscript.
- Marna and Mohau –suggested datasets to be explored, reviewed the manuscript, and made inputs on data analysis approaches and research questions formulations.
- Gregor – Inputs into the formulation of research questions, provision of in situ data, and critical discussion and review of the manuscript.
- Francois – Led the CCAM-CABLE model simulations and introduced the lead author to the model structure and the dynamical downscaling methods.
- Michael – Supervisory role and manuscript review.

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgements: This work was funded by the EEGC030 project of the CSIR. The authors wish to acknowledge Ms Humbelani Thenga and Dr Marc Pienaar for their contributions.

Data availability:

In situ data

Data from the CSIR owned flux tower (i.e. Skukuza and Malopeni) may be requested from Ms Humbelani Thenga (HThenga@csir.co.za).

CCAM-CABLE

The CCAM-CABLE simulations data may be requested from Dr Rebecca Garland (RGarland@csir.co.za).

ESA CCI

The data are available online from www.esa-soilmoisture-cci.org/

GLEAM

The data are available online from www.gleam.eu

Analysis scripts:

https://github.com/vkhosa/Floyd-Vukosi-Khosa-HESS-2018.git

5 References

Al-Yaari, A., Wigneron, J. P., Ducharne, A., Kerr, Y. H., Wagner, W., De Lannoy, G., Reichle, R., Al Bitar, A., Dorigo, W., Richaume, P. and Mialon, A.: Global-scale comparison of passive (SMOS) and active (ASCAT) satellite based microwave soil moisture retrievals with soil moisture simulations (MERRA-Land), Remote Sens. Environ., 152, 614–626, doi:10.1016/j.rse.2014.07.013, 2014.

Albergel, C., de Rosnay, P., Gruhier, C., Munoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y. and Wagner, W.: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, Remote Sens. Environ., 118, 215–226, doi:10.1016/j.rse.2011.11.017, 2012.

An, R., Zhang, L., Wang, Z., Quaye-Ballard, J. A., You, J., Shen, X., Gao, W., Huang, L. J., Zhao, Y. and Ke, Z.: Validation of the ESA CCI soil moisture product in China, Int. J. Appl. Earth Obs. Geoinf., 48, 28–36, doi:10.1016/j.jag.2015.09.009, 2016.

Archibald, S. A., Kirton, A., van der Merwe, M. R., Scholes, R. J., Williams, C. A. and Hanan, N.: Drivers of inter-annual variability in net ecosystem exchange in a semi-arid savanna ecosystem, South Africa, Biogeosciences, 6(2), 251–266, doi:10.5194/bg-6-251-2009, 2009.

Bristow, D.: Vegetation of the Drakensberg, [online] Available from: http://southafrica.co.za/vegetation-drakensberg.html (Accessed 30 September 2019), n.d.

Brocca, L., Melone, F. and Moramarco, T.: Distributed rainfall-runoff modelling for flood frequency estimation and flood forecasting, Hydrol. Process., 25(18), 2801–2813, doi:10.1002/hyp.8042, 2011.

Cleveland, R. B., Cleveland, W. S., McRae, J. E. and Terpenning, I.: STL: A seasonal-trend decomposition procedure based on loess, J. Off. Stat., 6(1), 3–73, doi:citeulike-article-id:1435502, 1990.

Decker, M.: Development and evaluation of a new soilmoisture and runoff parameterization for the CABLE LSM including subgrid-scale processes, J. Adv. Model. Earth Syst., 7, 513–526, doi:10.1002/2015MS000507, 2015.

Dedekind, Z., Engelbrecht, F. A. and Van Der Merwe, J.: Model simulations of rainfall over southern africa and its eastern escarpment, Water SA, doi:10.4314/wsa.v42i1.13, 2016.

Dirmeyer, P. A., Jin, Y., Singh, B. and Yan, X.: Trends in land–atmosphere interactions from CMIP5 simulations, J. Hydrometeorol., 14(3), 829–849, doi:10.1175/JHM-D-12-0107.1, 2013.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M. and Kidd, R.: Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sens. Environ., 162, 380–395, doi:10.1016/j.rse.2014.07.023, 2015.

Engelbrecht, F., Landman, W., Engelbrecht, C., Landman, S., Bopape, M., Roux, B., McGregor, J. and Thatcher, M.: Multi-scale climate modelling over Southern Africa using a variable-resolution global model, Water SA, 37(5), 647–658, doi:10.4314/wsa.v37i5.2, 2011.

Engelbrecht, F., Adegoke, J., Bopape, M.-J., Naidoo, M., Garland, R., Thatcher, M., McGregor, J., Katzfey, J., Werner, M., Ichoku, C. and Gatebe, C.: Projections of rapidly rising surface temperatures over Africa under low mitigation, Environ. Res. Lett., 10(8), 085004, doi:10.1088/1748-9326/10/8/085004, 2015.

Engelbrecht, F. A., McGregor, J. L. and Engelbrecht, C. J.: Dynamics of the conformal-cubic atmospheric model projected climate-change signal over southern Africa, Int. J. Climatol., 29(7), 1013–1033, doi:10.1002/joc.1742, 2009.

Fang, L., Hain, C. R., Zhan, X. and Anderson, M. C.: An inter-comparison of soil moisture data products from satellite remote sensing and a land surface model, Int. J. Appl. Earth Obs. Geoinf., 48, 37–50, doi:10.1016/j.jag.2015.10.006, 2016.

Feig, G. T., Mamtimin, B. and Meixner, F. X.: Soil biogenic emissions of nitric oxide from a semi-arid savanna in South Africa, Biogeosciences, 5(6), 1723–1738, doi:10.5194/bg-5-1723-2008, 2008.

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D. and Schär, C.: Soil moisture-atmosphere interactions during the 2003 European summer heat wave, J. Clim., 20(20), 5081–5099, doi:10.1175/JCLI4288.1, 2007.

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A.,

Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLoS One, 12(2), 1–40, doi:10.1371/journal.pone.0169748, 2017.

Horowitz, H. M., Garland, R. M., Thatcher, M., Landman, W. A., Dedekind, Z., van der Merwe, J. and Engelbrecht, F. A.: Evaluation of climate model aerosol seasonal and spatial variability over Africa using AERONET, Atmos. Chem. Phys., 17(22), 13999–14023, doi:10.5194/acp-17-13999-2017, 2017.

van den Hurk, B., Doblas-Reyes, F., Balsamo, G., Koster, R. D., Seneviratne, S. I. and Camargo, H.: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe, Clim. Dyn., 38(1–2), 349–362, doi:10.1007/s00382-010-0956-2, 2012.

Jovanovic, N., Mu, Q., Bugan, R. D. H. and Zhao, M.: Dynamics of MODIS evapotranspiration in South Africa, Water SA, 41(1), 79–91, doi:10.4314/wsa.v41i1.11, 2015.

Khosa, F. V., Feig, G. T., van der Merwe, M. R., Mateyisi, M. J., Mudau, A. E. and Savage, M. J.: Evaluation of modeled actual evapotranspiration estimates from a land surface, empirical and satellite-based models using in situ observations from a South African semi-arid savanna ecosystem, Agric. For. Meteorol., 279(July), 107706, doi:10.1016/j.agrformet.2019.107706, 2019.

Kowalczyk, E. A., Wang, Y. P., Law, R. M., Davies, H. L., McGregor, J. L. and Abramowitz, G.: The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model., 2006.

Kraskov, A., Stögbauer, H. and Grassberger, P.: Estimating mutual information, Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top., 69(6), 16, doi:10.1103/PhysRevE.69.066138, 2004.

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., Evans, J. P. and van Dijk, A. I. J. M.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sens. Environ., 123, 280–297, doi:10.1016/j.rse.2012.03.014, 2012.

Loew, A., Stacke, T., Dorigo, W., De Jeu, R. and Hagemann, S.: Potential and limitations of multidecadal satellite soil moisture observations for selected climate model evaluation studies, Hydrol. Earth Syst. Sci., 17(9), 3523–3542, doi:10.5194/hess-17-3523-2013, 2013.

Lorenz, R., Jaeger, E. B. and Seneviratne, S. I.: Persistence of heat waves and its link to soil moisture memory, Geophys. Res. Lett., 37(9), 1–5, doi:10.1029/2010GL042764, 2010.

Malherbe, J., Engelbrecht, F. A. and Landman, W. A.: Projected changes in tropical cyclone climatology and landfall in the Southwest Indian ocean region under enhanced anthropogenic forcing, Clim. Dyn., 40(11–12), 2867–2886, doi:10.1007/s00382-012-1635-2, 2013.

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A. M. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., Verhoest, N. E. C. C., Férnandez-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C. C.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev. Discuss., 10(5), 1903–1925, doi:10.5194/gmd-10-1903-2017, 2017.

McGregor, J. L.: C-CAM geometric aspects and dynamical formulation, Australia., 2005.

McGregor, J. L. and Dix, M. R.: The CSIRO conformal-cubic atmospheric GCM, Fluid Mech. its Appl., 61, 197–202, doi:10.1007/978-94-010-0792-4_25, 2001.

McGregor, J. L. and Dix, M. R.: An updated description of the conformal-cubic atmospheric model, High Resolut. Numer. Model. Atmos. Ocean, (2001), 51–75, doi:10.1007/978-0-387-49791-4_4, 2008.

McNally, A., Shukla, S., Arsenault, K. R., Wang, S., Peters-Lidard, C. D. and Verdin, J. P.: Evaluating ESA CCI soil moisture in East Africa, Int. J. Appl. Earth Obs. Geoinf., 48, 96–109, doi:10.1016/j.jag.2016.01.001, 2016.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A. and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15(2), 453–469, doi:10.5194/hess-15-453-2011, 2011.

Palmer, A. R., Weideman, C., Finca, A., Everson, C. S., Hanan, N. and Ellery, W.: Modelling annual evapotranspiration in a semi-arid, African savanna: functional convergence theory, MODIS LAI and the Penman–Monteith equation, African J. Range Forage Sci., 32(1), 33–39, doi:10.2989/10220119.2014.931305, 2015.

Pinheiro, A. C. and Tucker, C. J.: Assessing the relationship between surface temperature and soil moisture in southern Africa, Remote Sens. Hydrol., 2000(267), 296–301, 2001.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., 100(2), 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Raj Koirala, S. and W. Gentry, R.: SWAT and wavelet analysis for understanding the climate change impact on hydrologic response, Open J. Mod. Hydrol., 02(02), 41–48, doi:10.4236/ojmh.2012.22006, 2012.

Ramoelo, A., Majozi, N., Mathieu, R., Jovanovic, N., Nickless, A. and Dzikiti, S.: Validation of global evapotranspiration product (MOD16) using flux tower data in the African savanna, South Africa, Remote Sens., 6(8), 7406–7423, doi:10.3390/rs6087406, 2014.

Raoult, N., Delorme, B., Ottlé, C., Peylin, P., Bastrikov, V., Maugis, P. and Polcher, J.: Confronting Soil Moisture Dynamics from the ORCHIDEE Land Surface Model With the ESA-CCI Product: Perspectives for Data Assimilation, Remote Sens., 10(11), 1786, doi:10.3390/rs10111786, 2018.

Rosch, A. and Schmidbauer, H.: WaveletComp : A guided tour through the R-package, , 1–38 [online] Available from: http://www.hs-stat.com/projects/WaveletComp/WaveletComp_guided_tour.pdf, 2018.

Rubin, D. B.: Multiple Imputation for Nonresponse in Surveys, edited by D. B. Rubin, John Wiley & Sons, Inc., Hoboken, NJ, USA., 1987.

Rubin, D. B.: Multiple Imputation after 18+ Years, J. Am. Stat. Assoc., 91(434), 473–489, doi:10.1080/01621459.1996.10476908, 1996.

Scholes, R. J., Gureja, N., Giannecchinni, M., Dovie, D., Wilson, B., Davidson, N., Piggott, K., McLoughlin, C., Van der Velde, K., Freeman, A., Bradley, S., Smart, R. and Ndala, S.: The environment and vegetation of the flux measurement site near Skukuza, Kruger National Park, Koedoe, 44(1), 73–84, doi:10.4102/koedoe.v44i1.187, 2001.

Seneviratne, S. I., Koster, R. D., Guo, Z., Dirmeyer, P. A., Kowalczyk, E., Lawrence, D., Liu, P., Mocko, D., Lu, C.-H., Oleson, K. W. and Verseghy, D.: Soil moisture memory in AGCM simulations: Analysis of global land–atmosphere coupling experiment (GLACE) data, J. Hydrometeorol., 7(5), 1090–1112, doi:10.1175/JHM533.1, 2006.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B. and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, Earth-Science Rev., 99(3–4), 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010.

Sheffield, J. and Wood, E. F.: Global trends and variability in soil moisture and drought characteristics, 1950-2000, from observation-driven simulations of the terrestrial hydrologic cycle, J.

840   Clim., 21(3), 432–458, doi:10.1175/2007JCLI1822.1, 2008.

Sinclair, S. and Pegram, G. G. S.: A comparison of ASCAT and modeled soil moisture over South Africa, using TOPKAPI in land surface mode, Hydrol. Earth Syst. Sci. Discuss., 6, 7439–7482, 2010.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106(D7), 7183–7192, doi:10.1029/2000JD900719, 2001.

845   Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, Bull. Am. Meteorol. Soc., 79(1), 61–78, doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2, 1998.

Veleda, D., Montagne, R. and Araujo, M.: Cross-wavelet bias corrected by normalizing scales, J. Atmos. Ocean. Technol., 29(9), 1401–1408, doi:10.1175/JTECH-D-11-00140.1, 2012.

Wang, S., Mo, X., Liu, S., Lin, Z. and Hu, S.: Validation and trend analysis of ECV soil moisture data
850   on cropland in North China Plain during 1981–2010, Int. J. Appl. Earth Obs. Geoinf., 48, 110–121, doi:10.1016/j.jag.2015.10.010, 2016.

Wang, T. and Franz, T. E.: Evaluating climate and soil effects on regional soilmoisture spatial variability using EOFs, Water Resour. Res., (1), 5375–5377, doi:10.1002/2013WR014979.Reply, 2017.

855   Wang, Y. P., Kowalczyk, E., Leuning, R., Abramowitz, G., Raupach, M. R., Pak, B., Van Gorsel, E. and Luhar, A.: Diagnosing errors in a land surface model (CABLE) in the time and frequency domains, J. Geophys. Res. Biogeosciences, 116(1), 1–18, doi:10.1029/2010JG001385, 2011.

Whitley, R., Beringer, J., Hutley, L. B., Abramowitz, G., De Kauwe, M. G., Duursma, R., Evans, B., Haverd, V., Li, L., Ryu, Y., Smith, B., Wang, Y. P., Williams, M. and Yu, Q.: A model inter-
860   comparison study to examine limiting factors in modelling Australian tropical savannas, Biogeosciences, 13(11), 3245–3265, doi:10.5194/bg-13-3245-2016, 2016.

Winsemius, H. C., Dutra, E., Engelbrecht, F. A., Archer Van Garderen, E., Wetterhall, F., Pappenberger, F. and Werner, M. G. F.: The potential value of seasonal forecasts in a changing climate in southern Africa, Hydrol. Earth Syst. Sci., 18(4), 1525–1538, doi:10.5194/hess-18-1525-
865   2014, 2014.

Xia, Y., Ek, M. B., Wu, Y., Ford, T. and Quiring, S. M.: Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: comparison and analysis, J. Hydrometeorol., 16(5), 1962–1980, doi:10.1175/JHM-D-14-0096.1, 2015.

Yuan, S. and Quiring, S. M.: Evaluation of soil moisture in CMIP5 simulations over the contiguous
870   United States using in situ and satellite observations, Hydrol. Earth Syst. Sci., 21(4), 2203–2218, doi:10.5194/hess-21-2203-2017, 2017.

Zeng, J., Li, Z., Chen, Q., Bi, H., Qiu, J. and Zou, P.: Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations, Remote Sens. Environ., 163, 91–110, doi:10.1016/j.rse.2015.03.008, 2015.

875

Appendix

5.1 Appendix B – Cross wavelet analysis

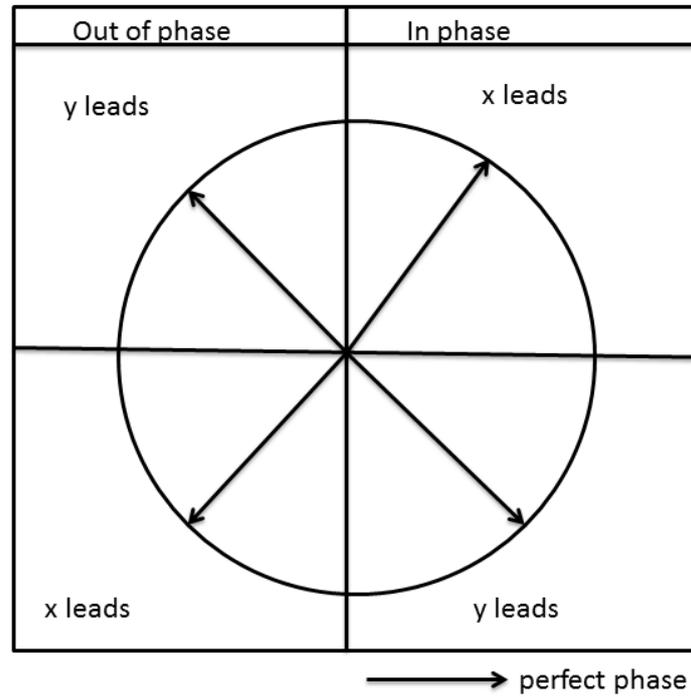



Figure B1. Phase interpretation between two time series x and y. When series x leads, y lags and vice versa. This figure is inspired by a study by (Rosch and Schmidbauer, 2018).

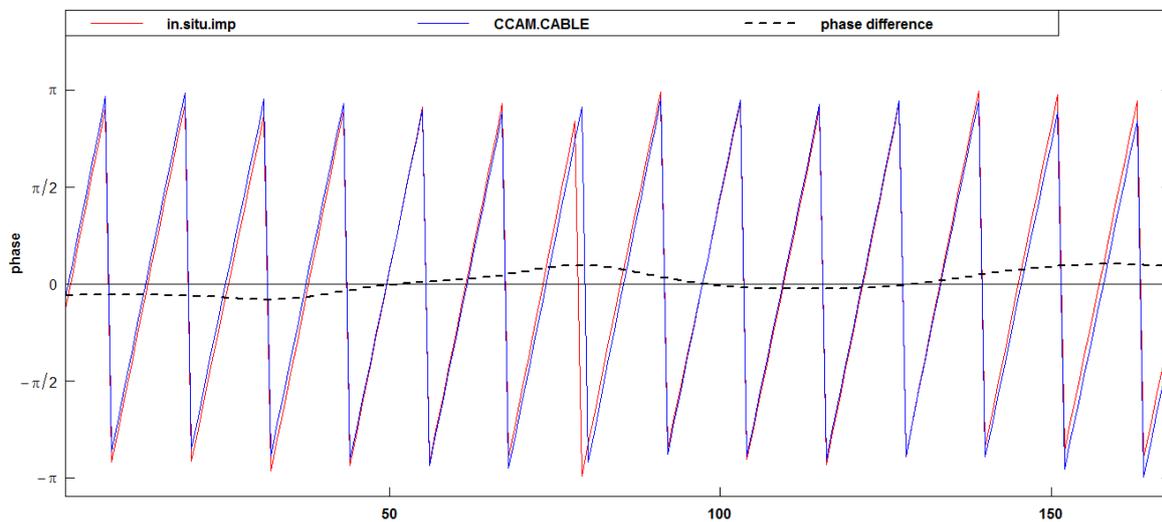



Figure B2. Phase difference between surface soil moisture simulated using CCAM-CABLE, and GLEAM v3a at Skukuza between 2001, and 2014 at period 12 at the surface.

 5.2 Appendix C – Multiple imputation

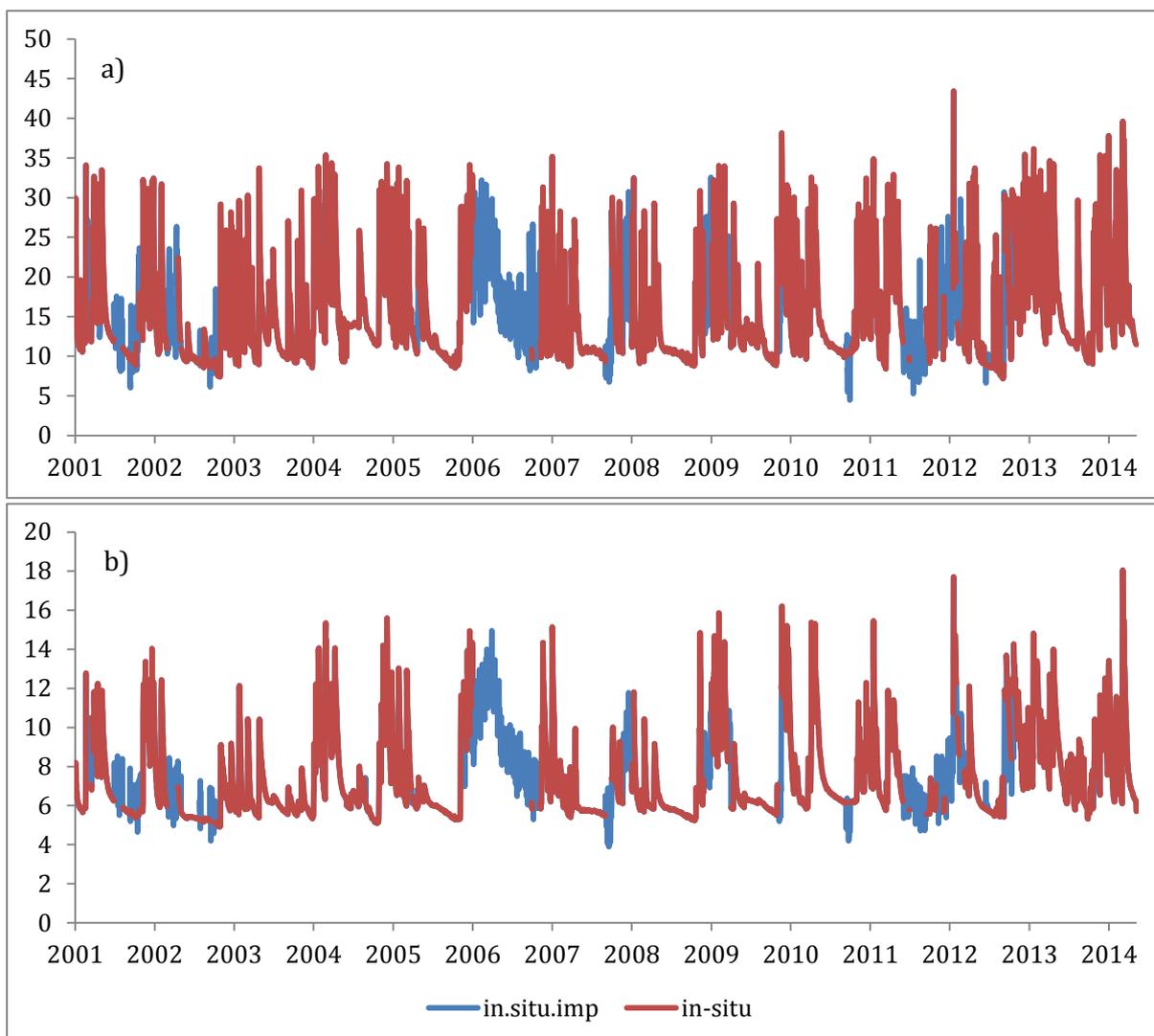



Figure C1. Daily a) surface and b) root zone soil moisture time series at Skukuza showing the imputed parts (blue) of the time series and the observed parts (red).

Table C1. Statistics of the distribution of the imputed and observed time series of surface and rootzone soil
 moisture at the Skukuza site.

Surface soil moisture	Original data	Imputed data
Mean	15.59	15.76
Median	13.33	13.83
Standard deviation	6.21	6.10
Variance	38.68	37.22
Root zone soil moisture		
Mean	7.45	7.55
Median	6.49	6.69
Standard deviation	2.18	2.17
Variance	4.76	4.74