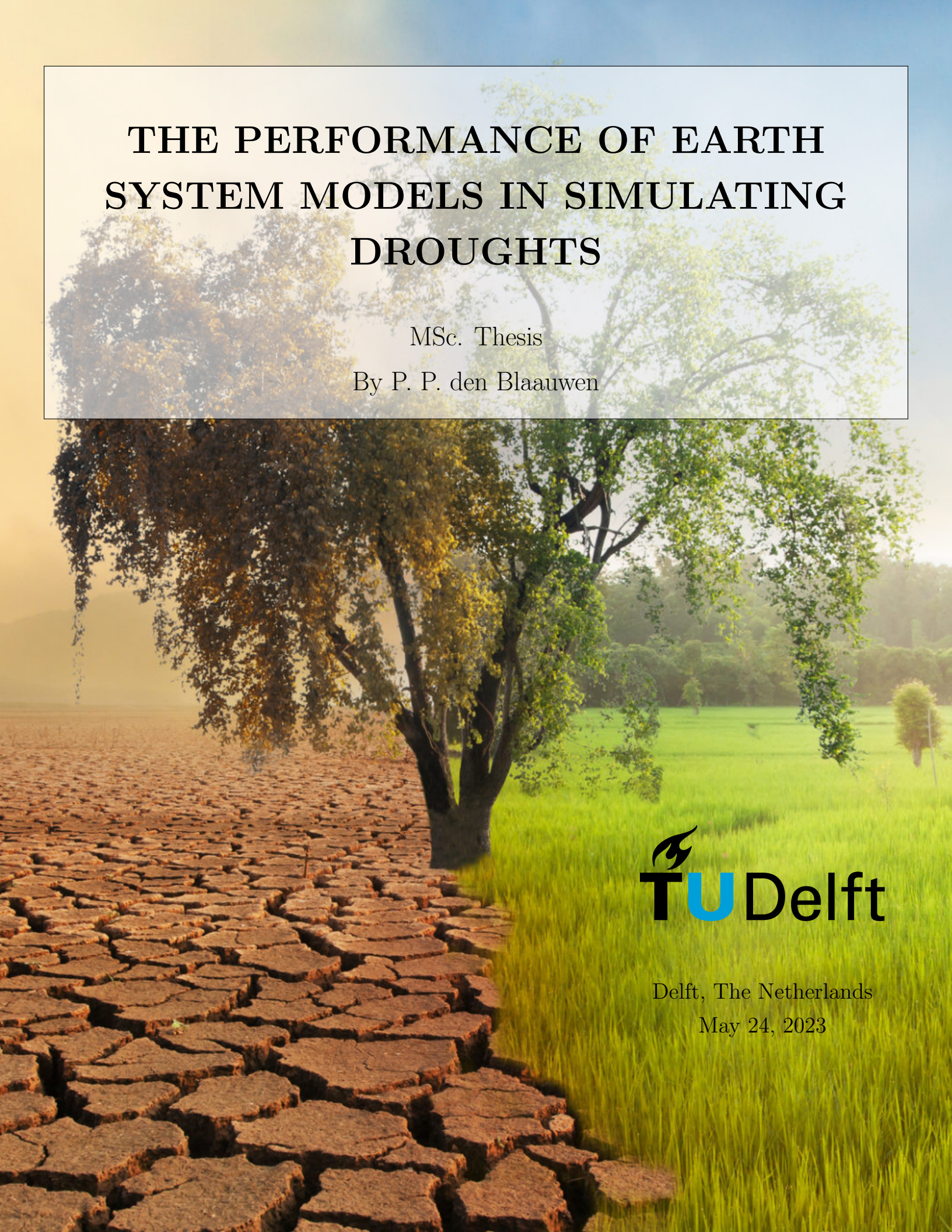
THE PERFORMANCE OF EARTH SYSTEM MODELS IN SIMULATING DROUGHTS

MSc. Thesis

By P. P. den Blaauwen



TU Delft

Delft, The Netherlands

May 24, 2023

THE PERFORMANCE OF EARTH SYSTEM MODELS IN SIMULATING DROUGHTS

MSc. Thesis

By Pepijn Pelle den Blaauwen

Student number:

4585941

Project duration:

September 2022 - May 2023

Assessment committee:

Ir. F. van Oorschot
Dr. ir. R. van der Ent
Dr. ir. R. Hut

Department:

Water Management

Faculty:

Civil Engineering and Geosciences

A digital version is available at https://repository.tudelft.nl/.

On the cover: Pieters H., 13 July 2020 media release: Understanding drought, the frequency of it, vulnerability and how the Garden Route as a region can adapt to become more resilient 2021

Preface

Dear reader,

In front of you lies the master thesis *'The performance of Earth System Models in simulating droughts'*. This research was the final requirement before obtaining my Master of Science degree in civil engineering at the Delft University of Technology. Within civil engineering, I chose the water management track and specifically in the area of water resources engineering. Besides this specialization, I had the freedom to further explore personal interests on the part of meteorology, climate modelling and hydrology which all align nicely within this specific research.

I would like to thank my daily supervisor, Fransje van Oorschot, for her guidance and support throughout this project. She always took the time to help me when I needed it. In general, she helped me the most with setting up the ICT systems needed to examine the Earth System model data. I would also like to thank Ruud van der Ent, the chair of my assessment committee, and Rolf Hut for their helpful input and feedback.

The project duration was from September 2022 until May 2023, during which I dived into the world of climate modelling, data analysis and programming. Personally I enjoyed this experience and I would not mind to continue in this field in my future career.

Thank you for taking the time to read my thesis.

Sincerely,

Pepijn den Blaauwen

Delft, The Netherlands
May 24, 2023

Abstract

This research evaluated the performance of Land Surface Models (LSMs) in simulating droughts, examining Land-Hist offline simulations from the Land, Surface, Snow and Soil Moisture Intercomparison Project (LS3MIP). It is well known that LSMs possess uncertainties and biases due to oversimplifications or the absence of certain physical processes (e.g., groundwater interactions and lateral connectivity). Therefore, the objective of this research was to identify the strengths and weaknesses of various LSMs and how this relates to the performances in simulating soil moisture droughts. To address this objective, eight LSMs were evaluated: CESM2, CMCC-ESM2, E3SM-1-1, EC-Earth3-Veg, HadGEM3-GC31-LL, IPSL-CM6A-LR, MIROC6, and UKESM1-0-LL. Two reference evaporation data sets (DOLCE V3 and an ensemble of FLUXCOM-RS, BESS and PML) and a reference soil moisture data set (SoMo.ml) were utilized for the evaluation. After a global analysis on the LSM evaporation characteristics, six climate diverse study areas were selected for further investigation. A long-term analysis was performed by examining the water balance and implementing the LSMs into the Budyko Framework. Subsequently, soil moisture deficits were calculated for the driest periods in time, and the resulting accumulated deficits were compared with the reference evaporation data. The timing and progression of the deficits were evaluated utilizing the reference soil moisture data. Finally, the sensitivity of the model was evaluated by examining the response of evaporation anomalies to precipitation anomalies and comparing this with the reference evaporation data. The results showed that there was a large spread in output and performance among the LSMs across all parts of the evaluation. The greatest contrasts among the LSMs were found in the dry to wet transition zones within the tropics. In this latitudinal range, the worst performing LSMs overestimated the accumulation of soil moisture deficits and the severity of droughts, while the opposite was found for the extratropical regions. Additionally, the models showed, in general, to be overly sensitive to precipitation anomalies. When ranking the implemented model bases in the LSMs based on their performance during droughts, the findings showed that the Community Land Model (implemented in CMCC-ESM2, E3SM-1-1 and CESM2) was predominantly the best performing, followed by ORCHIDEE (IPSL-CM6A-LR) and HTESSEL (EC-Earth3-Veg). MATSIRO (MIROC6) and JULES (HadGEM3-GC31-LL and UKESM1-0-LL) were the least performing model bases. From a hydrological perspective, the findings of this research could be linked to some known limitations of LSMs. Oversimplified soil and vegetation dynamics could contribute to the LSMs being overly sensitive to precipitation anomalies while the contrasts between the tropical and extratropical regions could be attributed to the representation of the soil moisture-evaporation coupling, which plays a greater role in the tropical study areas. Ultimately, this research could contribute to LS3MIP and the Land Surface Modeling community, as the results highlight the strengths and weaknesses of the LSMs in simulating soil moisture droughts. From there, this research could contribute to improving LSMs, understanding drought mechanisms, and addressing climate change impacts, especially in drought-prone regions.

List of Figures

List of Tables

Acronyms

C2 CESM2

CE CMCC-ESM2

CMIP Coupled Model Intercomparison Project

DO DOLCE-V3

E3 E3SM-1-1

EE EC-Earth3-Veg

ESMs Earth System Models

FBP FLUXCOM-RS BESS and PML

GSWP3 Global Soil Wetness Project Phase 3

HG HadGEM3-GC31-LL

IP IPSL-CM6A-LR

IPCC Intergovernmental Panel on Climate Change

ITCZ Intertropical Convergence Zone

LS3MIP Land Surface Snow and Soil Moisture Intercomparison Project

M6 MIROC6

NetCDF Network Common Data Form

SDGs Sustainable Development Goals

UK UKESM1-0-LL

Contents

1 — Introduction

1.1 Background

Recent global challenges on the part of climate change are leading to the demand and development of more sophisticated and complex Earth System Models (ESMs). Within these models, the biophysical processes of our Earth system are implemented to simulate the response of land-ocean-atmosphere interactions [Seneviratne et al., 2010]. As various modeling groups are working on a unique and rather complex model, it remains a challenge to compare their performance. Therefore, for an accurate implementation on a global scale, these models are compared in the Coupled Model Intercomparison Project (CMIP). This project aims to diagnose systematic biases and uncertainties through a series of experiments and simulations. In addition to this objective, the project provides recommendations for the climate reports of the Intergovernmental Panel on Climate Change (IPCC) [Eyring et al., 2016].

Within the CMIP, of which the sixth phase is the most recent, a specific group of researchers contributes to Land Surface Snow and Soil Moisture Intercomparison Project (LS3MIP) through the development and improvement of Land Surface Models (LSMs), which are a part of the ESMs. Among other things, they focus on modeling water and energy fluxes and validating them against certain uncertainties and biases. Part of the focus within LS3MIP lies on the representation of land-atmosphere feedbacks (LFMIP), while the other part focuses on the representation of the land processes (LMIP). For the latter, off-line simulations are used in which the land surface does not interact with other components of the model and in which pre-set conditions are equal for each LSM [van den Hurk et al., 2016]. This set-up offers the possibility to compare and evaluate the performance of the models in terms of the representation of underlying processes of the land surface. However, studies have shown that despite the efforts of LS3MIP, there are still some major challenges to overcome in the field of LSMs (e. g.: [Fisher and Koven, 2020], [Mueller and Seneviratne, 2014]).

These studies have shown that uncertainties and biases exist in the hydrological parameterizations and representations within LSMs and that much of this is related to groundwater and soil moisture dynamics [Clark et al., 2015]. For example, most LSMs used in Earth system modeling do not account for any lateral flow and only account for vertical fluxes on a grid scale [Kim and Mohanty, 2016]. This means that certain water fluxes such as soil moisture and groundwater dynamics are generally not well incorporated into the model and that this, for instance, also affects the representation of river discharges [Zampieri et al., 2012]. Furthermore, research has shown that the root zone storage capacity is often not well represented in LSMs [van Oorschot et al., 2021]. This parameter describes the available soil moisture for vegetation that can be taken up by its roots and contributes, among others, to the hydrological cycle in terms of the available water for transpiration [de Boer-Euser et al., 2016]. In addition to the challenges in soil moisture and groundwater, LSMs are known to be very limited in terms of the representation of vegetation phenology [Zeng et al., 2018]. Vegetation plays a major role in the water cycle and is linked to various water fluxes, such as evaporation, through a variety of feedback. All these uncertainties and biases related to the water cycle, in turn, affect the simulation of other essential aspects of our Earth system, such as the magnitude of the temperature and the carbon uptake. This underscores the importance of intercomparison studies and progress in this field of research.

1.2 Research questions and objectives

The objective of this research is to evaluate the performance of the LSMs, used in LS3MIP, in the simulation of droughts. For this research, it is chosen to consider soil moisture droughts. These are caused by a deficit in the precipitation minus evaporation flux. Other categories of droughts, such as hydrological and socioeconomic droughts, may become the result of a soil moisture drought but are beyond the scope of this research [Van Loon, 2015]. As the atmospheric forcing input for the LMIP simulations is constant for all LSMs, the focus is on the evaporation output, which is the main driver of the soil moisture drought in this case.

The following research question is constructed for this research:

What is the performance of Land Surface Models considered in LS3MIP when simulating soil moisture droughts?

In addition to evaluating the individual performance of the LSMs considered in LS3MIP, this research aims to investigate the impact of potential errors and biases within the models on the simulation of soil moisture droughts. This analysis is supported by examining potential causes for certain performances from a hydrological perspective and based on previous conducted research.

In order to answer the research question, the following three subquestions are constructed:

1. Do the Land Surface Models satisfy the physical constraints that are involved in the water balance?

This subquestion looks into how a Land Surface Model simulates the water balance, which is a crucial part of the Earth's climate system. If there are significant errors in this, it leads to a misrepresentation of reality in the simulations followed by less accurate future predictions. Added to this is the fact that within LSMs biophysical processes are intertwined and a deviation in the water balance has an effect on, for example, the carbon and energy cycles [Seneviratne et al., 2010]. However, in this research, only the water balance is considered, as the investigation of feedback with other cycles is beyond the scope of this research. This subquestion checks whether the water balance is closed at all times and how the distribution of different water fluxes differs between the LSMs. This analysis is a first indication of LSM performance and is not yet focused on drought conditions.

2. What is the performance of the Land Surface Models in simulating soil moisture deficits during droughts?

In the first part of this subquestion, the LSM output is compared with reference evaporation data to determine the model accuracy of the simulated absolute value of the SMD, build up during a dry period in time. During droughts, it is important to estimate when a soil moisture deficit may become significant in terms of, for example, food security and water availability. Any biases or errors within the LSMs could lead to, for instance, incorrect simulations of the wilting point for a specific crop, preventing appropriate measures from being taken. By highlighting errors and biases compared to these reference data sets, this first part of the subquestion contributes to the eventual answering of the research question.

The second part of this subquestion only focuses on the timing and progression of the simulated drought conditions by comparing the SMDs to the reference soil moisture data. In addition to the total decrease in soil moisture during a drought, it is crucial for LSMs to accurately simulate the pattern of soil moisture decline over time. Comparing this with the reference soil moisture data shows the ability of the LSMs to capture key physical processes during droughts. In addition to capturing the temporal characteristics, it is crucial for the LSMs to simulate the start and end of the dry period at the right time. This is to avoid overestimating or underestimating the severity of the drought. This subquestion contributes to the overarching research question by evaluating the performance of LSMs based on their correlation with reference soil moisture data and shining light on how well they represent the underlying physical processes.

3. What is the sensitivity of the modeled evaporation anomalies to precipitation anomalies?

In this subquestion the sensitivity of the LSM evaporation anomalies to precipitation anomalies is considered. Since this research focuses on the simulation of soil moisture droughts, it is important to investigate the influence of precipitation anomalies as input on the subsequent output of evaporation. This analysis shows which LSMs are the most water limited in terms of their evaporation output. To create an overall picture of the water fluxes within the LSMs, runoff and soil moisture output are also considered. This subquestion contributes to the overarching research question by shining light on the performance and representation of physical processes of the LSMs during drier periods in time.

Relevance of this research

In the current era, in which climate change and food insecurity are critical issues, droughts are becoming an increasingly significant problem. Improving the simulation of these events could benefit the reliability of food and water supplies and contribute to Sustainable Development Goals (SDGs) 2 and 6 (Zero hunger and Clean water and sanitation). Furthermore, this research could contribute to the research fields of land surface and Earth system modeling by identifying potential errors and biases within the LSMs. The ultimate goal is that this will assist in improving the performance of these models and to bridge the gap between the hydrology and Earth system modeling communities. Future research will hopefully build on the results of this study and help to close the gap between LSMs and reality even more.

2 — Methodology

2.1 Data description

All data used in this research were of the Network Common Data Form (NetCDF) file type. This is a file type that can store multidimensional data allowing it to incorporate variables, such as evaporation and precipitation, over a spatial grid and over a given time series. NetCDF is a widely used data type in the climate model research community. All analyses were performed on the DelftBlue supercomputer from the Delft University of Technology [Delft High Performance Computing Centre (DHPC), 2022].

2.1.1 Land Surface Models in LS3MIP

For this research, from LS3MIP, the LMIP experiments were the most relevant, as the land-only experiments provided insight into the representation of key physical processes within the model. The strength of the LMIP is that the experiments are run under the same pre-set conditions and all LSMs experience the same Global Soil Wetness Project Phase 3 (GSWP3) atmospheric forcing, making the outcome comparable and errors and biases easier to identify. In addition, LMIP does simulations of the past (Land-Hist) and predictions for the future (Land-Future). For this research the focus was on the Land-Hist simulations as this created the possibility to compare the output to various reference data. Enhancing past simulations is likely to also improve future predictions. The ensemble includes a variety of LSMs, each with its own characteristics, strengths and weaknesses. As a result, different models may emerge in different parts of the world that best represent the hydrologic cycle [van den Hurk et al., 2016].

Figure 2.1 shows the pathway from CMIP6 until the Land-Hist simulations that were evaluated in this research. In the remainder of this research, the LSM, which is a part of the overarching ESM, was referred to by the same name as the overarching model. From LS3MIP the following LSMs were considered in this research: CESM2 (C2), CMCC-ESM2 (CE), E3SM-1-1 (E3), EC-Earth3-Veg (EE), HadGEM3-GC31-LL (HG), IPSL-CM6A-LR (IP), MIROC6 (M6) and UKESM1-0-LL (UK). Table 2.1 shows these eight LSMs and the outputs that were evaluated in this research. It should be noted that these experiments from LS3MIP include around 30 LSMs, provide more output and a longer time scale than investigated in this research and shown in Table 2.1. Unfortunately, the mrsol (soil moisture) output was not available for E3 and EE in this research.

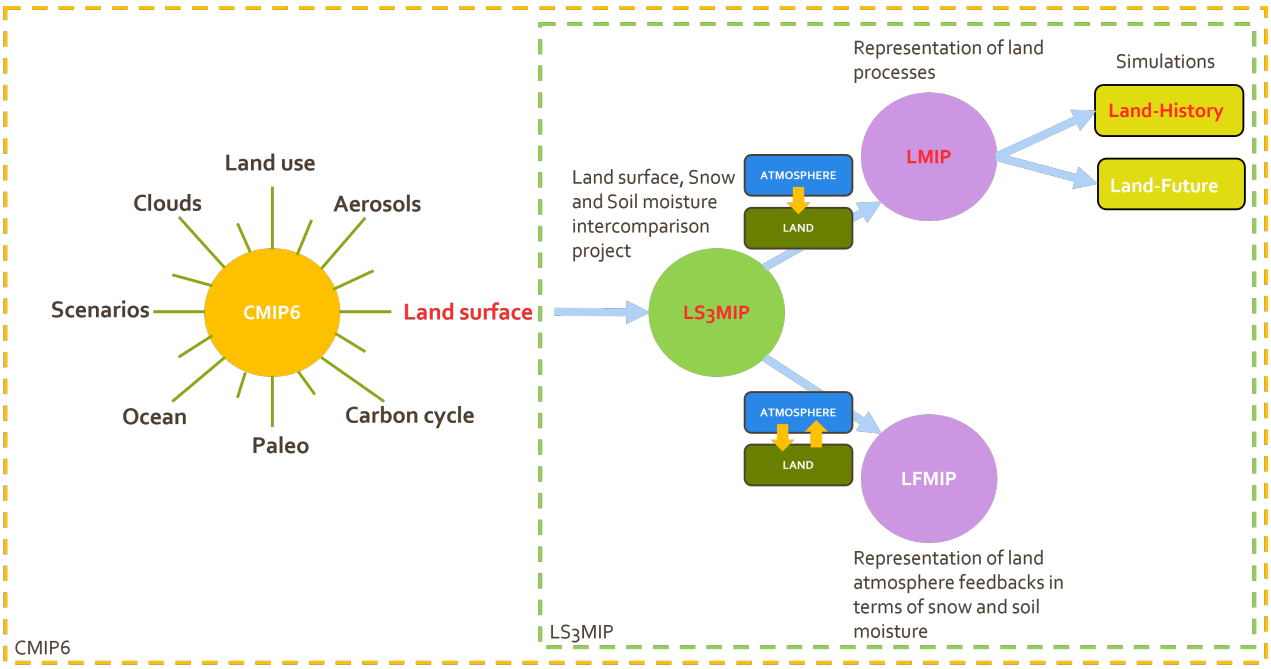


Figure 2.1: *Pathway, shown in red text, from the overarching CMIP6 ensemble towards the Land-History simulations evaluated in this research*

LSM	OUTPUT	TIMESCALE	RESOLUTION	REFERENCE
CESM2	E, SM and R	2001 -2010	288° x 192°	[Danabasoglu et al., 2020]
CMCC-ESM2	E, SM and R	2001 -2010	288° x 192°	[Lovato et al., 2022]
E3SM-1-1	E and R	2001 -2010	720° x 360°	[Bogenschutz et al., 2020]
EC-Earth3-Veg	E and R	2001 -2010	512° x 256°	[Döscher et al., 2021]
HadGEM3-GC31-LL	E, SM and R	2001 -2010	192° x 144°	[Andrews et al., 2020]
IPSL-CM6A-LR	E, SM and R	2001 -2010	144° x 143°	[Boucher et al., 2020]
MIROC6	E, SM and R	2001 -2010	256° x 128°	[Watanabe et al., 2011]
UKESM1-0-LL	E, SM and R	2001 -2010	192° x 144°	[Sellar et al., 2019]

Table 2.1: *LSMs included in LS3MIP and their output and timescales evaluated in this research. E is the evaporation, SM is the total soil moisture and R is the total run-off output of the LSM.*

2.1.2 SoMo.ml reference soil moisture data

The SoMo.ml (SO) dataset is a collection of worldwide in-situ soil moisture measurements that extrapolates daily data over the land surface area of the world by making use of machine learning [O. and Orth, 2021]. Over a thousand measurement points were used for this purpose (Figure 2.2). Meteorological variables were used as training data for the machine learning algorithm to simulate the relationship between input and determined soil moisture. Most soil moisture data sets make use of satellite data or model simulations, which makes this data set quite unique (e.g.: [Chen et al., 2021]).

A drawback of this product is that not every continent has the same density in terms of data points. It was therefore important to keep this distribution of points in mind during this research. As can be seen in Figure 2.2, to minimize biases due to the machine learning method, this research only selected study areas near at least two measurement points. Furthermore, the data set was scaled to match the mean and variability of the ERA5 data set which implies that it is not a completely independent product. Another constraint of this dataset is that it simulated the soil moisture storage up to a depth of 0.5 meters, divided over three layers. Although it is already an improvement compared to the top few centimeters included in most satellite observation-based products, soil moisture storage changes are likely to occur until greater depths in the soil layer [O. and Orth, 2021]. Moreover, there are some uncertainties in the measurements as soil moisture can be quite heterogeneous spatially, which is hard to capture making use of point data.

However, since this product primarily consists of in situ measurements, it is assumed that this product is more reliable compared to the benchmark evaporation products described in section 2.1.3. For this study, the three layers (0 - 0.1, 0.1 - 0.3 and 0.3 - 0.5 meter depth) were combined and used as reference data. Since the accounted depth of 0.5 meter is a limitation to compare the absolute value to the computed LSM soil moisture deficit, this data set was primarily used as a reference for the timing and progression of the droughts.

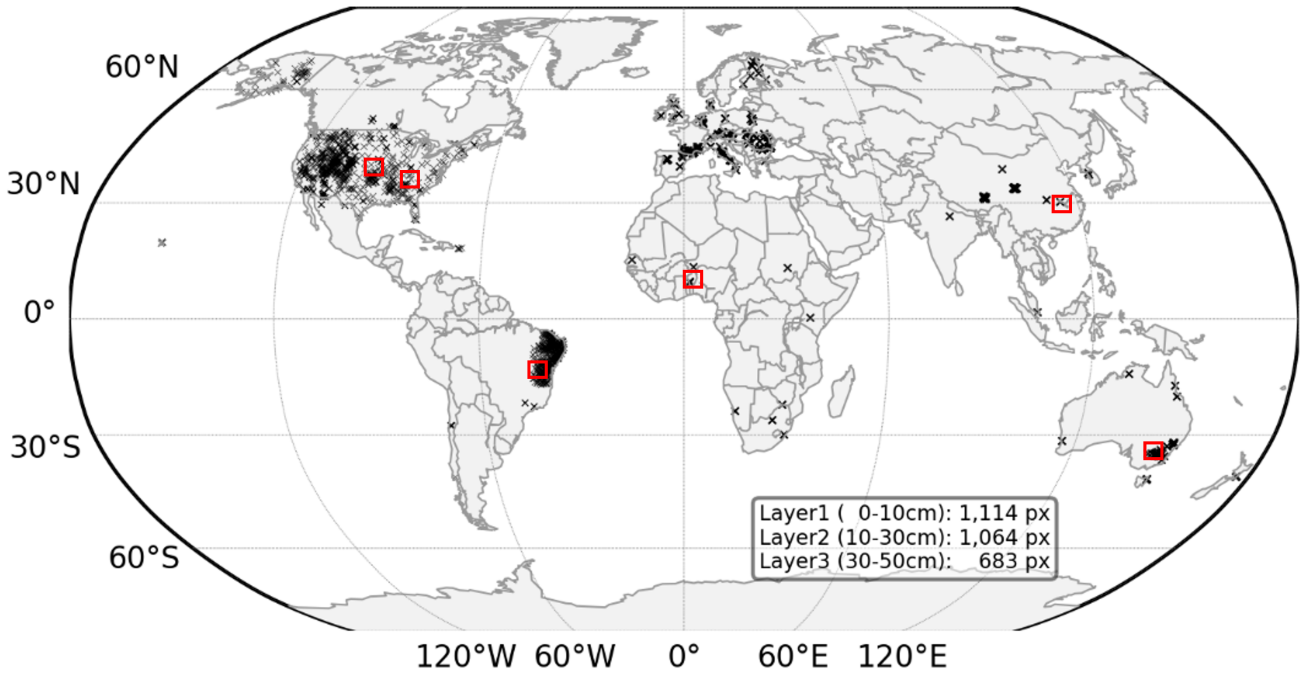


Figure 2.2: *Global distribution of the SoMo.ml in situ measurement points [O. and Orth, 2021]. The selected study areas, which are discussed in section 2.4, are enclosed by red squares.*

2.1.3 Benchmark evaporation products

Evaporation is an important process for the climate system of the Earth and its hydrological cycle [Pielke et al., 1998]. To provide reference data that best describes this, several benchmark evaporation products have been generated over the years (e.g. Table 2.2). These products are used as reference data by researchers to investigate, for example, the water cycle and land surface interactions [Mueller et al., 2013]. Furthermore, the products can be used to evaluate model outputs like the LSMs considered in this research (Table 2.1). However, it is known that these products, on their turn, have a number of uncertainties and biases in them. For example, it remains a challenge to accurately measure evaporation and the global network of measurements that exists (e.g. FLUXNET), represents point data, and is not well distributed on a global scale [Jiang and Ryu, 2016, Mueller et al., 2013].

Within the benchmark evaporation products, there are differences in how the evaporation is simulated. There are products that make use of a combination of all types of other products and models (reanalysis), there are products that make use of satellite data (remote sensing) and there are products that make use of in situ measurements (diagnostic) [Hobeichi et al., 2018]. Besides these three different types, there are products that use a combination of aforementioned methods. The benchmark evaporation products considered in this research are shown in Table 2.2.

PRODUCT	DERIVATION	TIMESCALE	RESOLUTION	REFERENCE
LANDFLUX-EVAL	Reanalysis	1989 - 2005	1.0°	[Mueller et al., 2013]
BESS	Remote sensing	Same as input	5 km	[Jiang and Ryu, 2016]
FLUXCOM RS METEO	Diagnostic, remote sensing and meteorological data	Same as input	0.5°	[Jung et al., 2019]
PML (V2)	Remote sensing and meteorological data	2000 - 2017	0.5 km	[Zhang et al., 2016]
GLEAM (V3)	Remote sensing	1980 - 2015	Varies per dataset	[Martens et al., 2017]
DOLCE (V1)	Reanalysis	2000 - 2009	0.25°	[Hobeichi et al., 2018]
DOLCE (V2-1)	Reanalysis	1980 - 2018	0.25°	[Hobeichi et al., 2021]
DOLCE (V3)	Reanalysis	1980 - 2018	0.25°	[Hobeichi et al., 2021]
CLASS	Reanalysis	2003 - 2009	0.5°	[Wouters et al., 2019]

Table 2.2: *Considered benchmark evaporation products in this research*

For this research, two benchmark evaporation products were chosen that served as reference data in the comparison with the LS3MIP model output. The requirements for these products were:

- The products must have an overlapping time series of at least 10 years. Otherwise, the time series studied is too short.
- The products must be as independent as possible of each other in terms of the models and methods used. This is to avoid overlapping biases and errors.
- The products should have as little missing data as possible. This with the exception of areas with almost no actual evaporation (e.g. the Sahara and Antarctica).

After reviewing all products in Table 2.2 for the aforementioned requirements ,the two products were chosen. The first one being DOLCE-V3 (DO) which is a reanalysis product and was derived from four global evaporation data sets. ERA5-land, FLUXCOM and two versions of GLEAM were used for this purpose [Hobeichi et al., 2021]. The second product being a combination of FLUXCOM-RS BESS and PML (FBP) as described by Singh et al. (2020).

2.1.4 GSWP3 atmospheric data

The Global Soil Wetness Project Phase 3 (GSWP3) is an initiative that aims to provide accurate reference data for various variables involved in the Earth's climate system [Dirmeyer, 2011]. This dataset was used as atmospheric forcing in the Land-Hist offline simulations conducted in the LS3MIP, and were held constant for all LSMs involved. The project combines both satellite measurements and ground-based observations and includes future scenario predictions alongside historical data. GSWP3 offers variables related to the water cycle, such as precipitation and relative humidity, as well as temperature and wind speed. The project has the goal of accelerating model development by providing reliable reference data for model input and validation [Dirmeyer, 2011]. Since its introduction, GSWP data has been applied in various studies (e.g.: [Al-Yaari et al., 2021]; [Essery et al., 2020]). For this research, the precipitation and temperature data from GSWP3 were used.

2.1.5 Data interactions

Figure 2.3 shows a how all data sets, discussed in sections 2.1.1, 2.1.2, 2.1.3 and 2.1.4, interacted in this research. It is worth noting that the primary objective was not to fully understand the functioning and documentation of individual LSMs, but rather to concentrate on their inputs and outputs. This research focused on the evaluation of the LSMs by investigating the simulated water fluxes during droughts and by utilizing three reference data sets, leading to a conclusion on the research question.

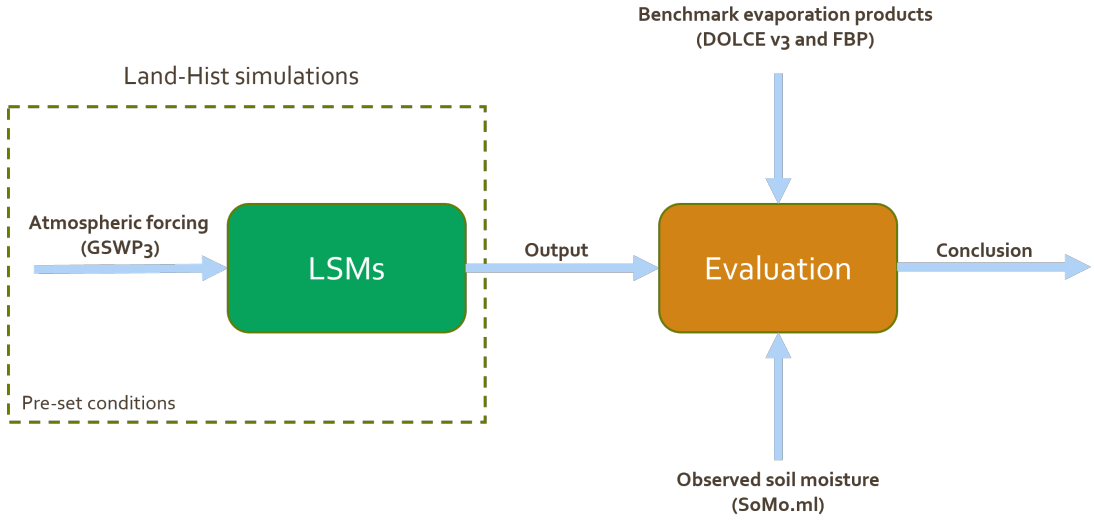


Figure 2.3: *General interactions between the GSWP3, reference data sets, LSMs and the evaluation done in this research*

2.2 Hydrological aspects

2.2.1 Evaporation

The term "evaporation" carries various meanings in different fields of research, therefore, it is important to specify the definition used in this study. For instance, certain studies differentiate between evapotranspiration and evaporation. The physical meaning of evaporation is the change in phase from a liquid to the gas phase and plays a major role in the hydrological cycle of the Earth system. In this research and within the LSMs, evaporation is defined as the combination of interception (E_I), transpiration (E_T), soil evaporation (E_S) and open water evaporation (E_O), as can be seen in Equation: 2.1. Interception is the fraction of precipitation that is intercepted and evaporated before reaching the soil, often due to the presence of vegetation. Transpiration is the water that is absorbed from the soil by vegetation and returned to the atmosphere by evaporation from the leaves. Soil evaporation is the process by which water evaporates directly from the soil. Lastly, open water evaporation is the water that is evaporated from a free water surface. It may vary which of these processes are dominant for the magnitude of the total evaporation.

$$E = E_I + E_T + E_S + E_O \tag{2.1}$$

2.2.2 Water balance

For this research, the water balance (Equation 2.2) was used to assess the performance of the model in capturing the conservation of mass principle. The water balance assumes that the sum of the incoming precipitation minus the outgoing evaporation and run-off equals zero ($\frac{\Delta S}{\Delta t} = 0$) over an extended time period and that changes in storage as well as interactions with groundwater are negligible. For this equation, the ten year mean evaporation ($\overline{E}$) and runoff ($\overline{R}$) output from the LSMs as well as the ten year mean GSWP3 precipitation ($\overline{P}$) were used. If the $\frac{\Delta S}{\Delta t}$ is far away from zero and negative, it indicates problems with the conservation of mass principle in the LSM, accompanied by a persistent decrease in storage. A large positive $\frac{\Delta S}{\Delta t}$ suggests a numerical problem in the LSMs and an imbalance between the inputs and outputs of the system and a persistent increase in storage.

$$\frac{\Delta S}{\Delta t} = \overline{P} - \overline{E} - \overline{R} \tag{2.2}$$

2.2.3 Budyko framework

The Budyko framework was used to examine the relationships between hydrologic variables such as precipitation and actual evaporation. This framework visualizes the relation between the potential and actual evaporation and checks whether the modelled evaporation exceeds the energy or water limit, which is in theory not possible [Chen and Sivapalan, 2020]. The Budyko framework assumes no changes in water storage over a time series of multiple years. Furthermore, the framework provides an empirically determined relation (Equation: 2.3) between the dryness index and the evaporation index ($\frac{E}{P}$). The ϕ represents the dryness index (Section: 2.2.6) and the ω is a free parameter which is calibrated to be equal to 2.6 [-] in the original Budyko curve [Greve et al., 2020]. For this research the ω was calibrated based on the reference evaporation data as discussed by Greve et al. (2020).

$$\frac{E}{P} = (1 + \phi) - (1 + \phi^{\omega})^{\frac{1}{\omega}} \tag{2.3}$$

2.2.4 Droughts and soil moisture deficits

Droughts are defined as the lack of water in a certain hydrological system beyond the annual variability of the water cycle. For this research, the focus was on soil moisture droughts which are caused by a deficit in the precipitation minus evaporation flux. These droughts develop over time and can have numerous negative consequences for agriculture and ecosystems. Over time, a soil moisture drought could potentially develop into a hydrological or socio-economic drought [Van Loon, 2015].

To be able to say something about the LSM performance during droughts, this research focused on soil moisture deficits (SMDs) over time. The SMD represents the cumulative difference between the incoming and outgoing water fluxes in the vertical direction. It is a simplification of reality, as it does not take into account any interactions with groundwater or horizontal water fluxes. The SMDs were calculated by the negative cumulative difference between precipitation and evaporation over time, as can be seen in Equation 2.4. To avoid a persistent decrease in SMD (i.e. negative deficit) when annual precipitation exceeds annual evaporation, the value was kept at greater than or equal to zero. The assumption being that at SMD = 0, the soil is at field capacity and all extra water will infiltrate or flow away as run-off. When the value exceeds zero, there is a difference between the field capacity and the actual soil moisture which represents the SMD. In equation 2.4, t1 equals the start and t2 the end of the investigated dry period in time, as shown in Table 2.4.

$$\mathrm{SMD(t)} = \max(0, -\int_{t1}^{t2} (P(t) - E(t))dt) \tag{2.4}$$

2.2.5 Potential evaporation

Potential evaporation is an important concept within the hydrology. It represents the amount of water that would evaporate based on the atmospheric conditions and has no limit on water that is actually available for evaporation. To determine the potential evaporation E_p, the Hargreaves-Samani method was used. This method uses the principle that the E_p depends on the amount of available energy. By means of temperature and radiation, an estimate can be computed of the magnitude of potential evaporation [Hargreaves and Samani, 1985]. Both the temperature and the radiation of the GSWP3 data sets were used for this. Since other methods, such as the Penman-Monteith method, require additional meteorological data variables as input this was a suitable method for this research [Chiew et al., 1995]. Variables used in this method, such as soil heat flux and actual vapor pressure, were not included in the GSWP3 data and the LSM forcing.

2.2.6 Dryness index

The dryness index (DI) (Equation 2.5) was a relevant hydrologic index for this research as it describes the Dryness of a region. It calculates the ratio between the precipitation and the potential evaporation. Low DI values (i.e. DI < 0.8) represent very wet regions while high DI values (i.e. DI > 4.0) represent very dry and arid regions. For this research the DI was used as a climate classification tool.

The $\overline{P}$ in Equation 2.5 refers to the annual precipitation of GSWP3 and $\overline{E_p}$ is the annual potential evaporation calculated as described in Section 2.2.5.

$$\mathrm{DI} = \frac{\overline{E_p}}{\overline{P}} \tag{2.5}$$

2.2.7 Seasonality index

The seasonality index was used to describe the annual seasonality of precipitation in the selected study areas. The index varies between zero and 1.83. A value of zero means that all precipitation is divided equally over the 12 months. A value of 1.83 means that all precipitation falls in a particular month [Guhathakurta and Saji, 2013].

In Equation 2.6, P_{an} overline is equal to the average annual precipitation and P_n is the average precipitation for month n.

$$\mathrm{SI} = \frac{1}{\overline{P_{an}}} \sum_{n=1}^{12} \left| P_n - \frac{\overline{P_{an}}}{12} \right| \tag{2.6}$$

2.3 Global analysis

To examine the overall differences between the various LSMs, a global analysis was carried out. Looking into the global mean and the spatial and temporal characteristics of the evaporation. Based on this, the research proceeded by specifying study areas to investigate in further depth.

2.3.1 Seasonality

The first step of the global analysis was to examine the seasonality of the global mean evaporation of the LSMs. The aim was to determine which deviations were caused by general biases and whether variations could be observed over different seasons. The monthly time series between 2001 and 2010 was used to compare the global monthly mean of the LSMs. It was important to note that all data sets were somewhat different in resolution and missing data and were therefore not entirely comparable. To make the models comparable, Antarctica was completely removed from all data sets, and the resolution was set equal. The data sets were set at the same resolution of 360° x 180°. The LSMs that were modified from a fine to a coarse resolution used the bilinear interpolation method, and from coarse to fine the nearest neighbor method was used. These modifications were performed by Climate Data Operators on the command line [Schulzweida, 2022]. Besides the individual LSMs, this step also considered the ensemble of the eight LSMs, as research has shown that the ensemble often has a lower deviation to reference data compared to individual LSMs [Wang et al., 2020].

To account for the influence of the curvature of the earth on the global mean, the entire grid was weighted based on the cosine of the latitude, as shown in Equations 2.7 and 2.8.

$$\phi = \frac{\text{latitude}}{180} * \pi \tag{2.7}$$

$$w = \cos(\phi) \tag{2.8}$$

To determine whether there was a statistically significant difference between the evaporation means of the LSMs, a Kruskal-Wallis test was performed. The complete time series between 2001 and 2010 was considered. This test is similar to the one-way ANOVA, with the exception that it can be used when the data sets are not normally distributed. The test is based on the assumption that the variances are approximately equal, the distributions have a similar shape, and the data sets are independent of each other [Zach, 2022]. Testing the assumption of similar variances was performed with the Levene's test [van den Berg, 2023]. For the Kruskal - Wallis test, the null hypothesis that all means were equal is rejected with a p-value of less than .05. The alternative hypothesis was that at least one of the LSM means was different from the others. Subsequently, t-tests were performed to determine the statistical differences among the individual LSMs and whether these differences were significant.

2.3.2 Spatial and temporal characteristics

After the analysis of the global mean seasonality of the LSMs, the spatial and temporal characteristics of the LSM evaporation were considered. The northern hemisphere summer months (**J**une, **J**uly and **A**ugust) and southern hemisphere summer months (**D**ecember, **J**anuary and **F**ebruary) were coupled and averaged over the investigated time series. Since this research focused on droughts, it was decided that the relatively hot summer months were the most relevant seasons to investigate in more detail. This analysis examined the spatial differences among the LSMs and compared the individual models to the LSM mean for the JJA and DJF months. In addition, the differences between the LSM ensemble mean and the DOLCE V3 reference data were examined. As mentioned in section 2.3.1 all data must be set equal in terms of resolution to perform this step.

2.3.3 Specifying study areas

The third step of the global analysis was the identification of study areas that were relevant to further investigate in this research. To determine these areas, the results of the analysis described in 2.3.2 were considered. In addition, a global plot was created showing the spatial distribution of the dryness index (Figure 2.4). This plot assisted in specifying the areas of interest by using the dryness index as a climate classification tool. Regions with a high dryness index (i.e. : DI > 5) experience almost neglectable actual evaporation and were therefore less relevant for this research. Similarly, areas with a very low Dryness index (i.e.. : DI < 0.8) are considered very wet and less prone to droughts, and hence, were also less relevant for this research. Another constraint for the identification of the study areas, as discussed in section 2.1.2, was the spatial distribution of the in situ measurement points of the SO reference data. For this research it was decided that the chosen study areas must cover at least two data points from this data set. This, to minimize certain biases and errors from the machine learning based extrapolation of the model [O. and Orth, 2021].

The most relevant water limited areas for this research were determined to be on the transition zone between wet and arid regions ($1.5 <$ DI < 5.0). These wet to dry transitional zones experience a strong seasonality in terms of precipitation, often depend on agriculture and are therefore sensitive to weather changes and droughts [Gamo et al., 2013]. In addition, these transition zones are on average relatively sensitive to climate change and more prone to change in biotopes and desertification ([Oliveras and Malhi, 2016]; [Vieira et al., 2020]; [Singh et al., 2020]). As this research focused on the performance of LSMs in simulating droughts, these dry to wet transition zones were especially interesting to investigate more thoroughly.

For this research, six study areas were specified with a size of 4° x 4° (latitude x longitude). The specified locations were divided into three types: water limited locations in the tropics (2x), water limited locations in the extratropics (2x) and energy limited locations (2x), all in terms of evaporation. For water limited locations the DI is greater than 1.0 [-] and energy limited location have a DI of less than 1.0 [-]. In the areas that are water limited, the soil moisture storage likely plays a significant role in the magnitude of the actual evaporation, since it is a potential limiting factor. For energy limited areas the role of the soil moisture storage likely differs as the actual evaporation is limited by the amount of energy that reaches the earth surface by solar radiation. By differentiating between energy and water limited study areas, the performance of the models could potentially be linked to the representation of underlying hydrologic processes such as the role of the soil moisture and vegetation.

2.4 Study areas

The study areas were determined as discussed in Section 2.3.3. Figure 2.4 shows a global plot of the dryness index, as well as the six specified study areas. Table 2.3 elaborates on the study areas and provides information on the coordinates, climate characteristics and elevation. Appendices A, B, C and G elaborate on the land cover, precipitation and temperature seasonality, the spatial characteristics of the soil moisture, and the Köppen-Geiger climate classification, respectively.

The six study areas could be grouped in pairs of two. The water limited locations in the tropics (A and B) experience a strong precipitation seasonality with a fairly constant temperature throughout the year and are located on a dry to wet transition zone with dryness indices of 1.82 and 1.72, respectively. The water limited locations in the extratropics (C and D) experience a lower seasonality in terms of precipitation but a higher annual temperature variation with dryness indices of 2.13 and 2.79. Therefore, on an annual basis, these locations are considered drier compared to locations A and B. Locations E and F are considered energy limited in terms of their annual evaporation as the dryness index is below 1.0 (0.96 and 0.71). This is mainly due to a high value for the annual precipitation. Combined with a significant annual temperature variation, this makes these locations relatively wet.

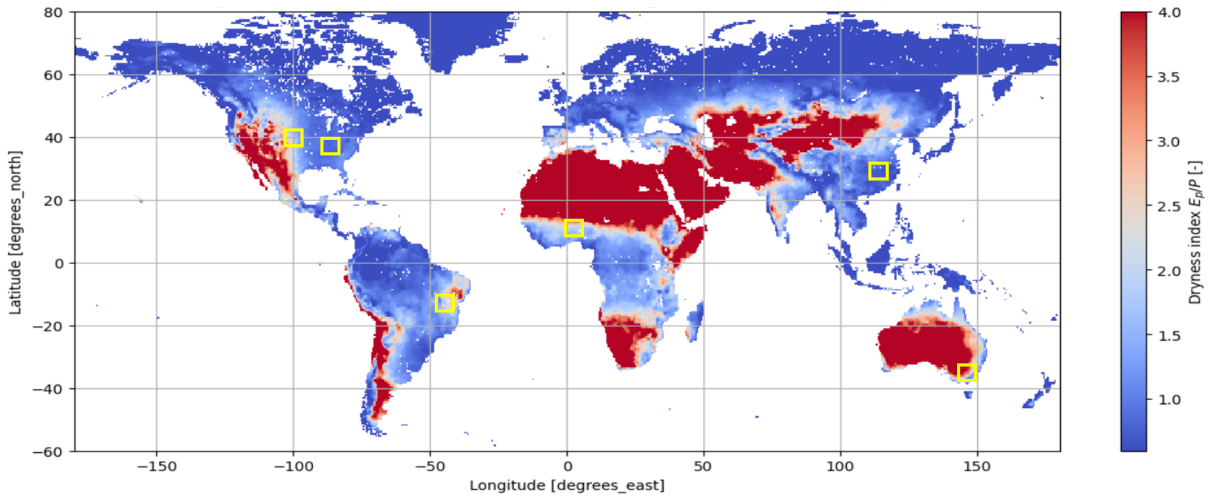


Figure 2.4: *Global plot of the mean Dryness index between 2001 and 2010 including the six specified study areas enclosed by a yellow square. The size of the squares is slightly exaggerated for clarity. The potential evaporation is computed as discussed in section 2.2.5 and the precipitation is obtained from the GSWP3 data.*

	Coordinates [lat , lon]	Annual P [mm]	Seasonality index [-]	Köppen-Geiger climate classification	Dryness index [-]	Elevation range [m]	Energy or water limited E
Location A: East Brazil	-16 until -12 , -46 until -42	∼900	0.77	As, BSh	1.82	300 - 1100	Water
Location B: West Africa	8 until 12 , 0 until 4	∼1100	0.83	Aw, BSh	1.72	60 - 650	Water
Location C: Central US	38 until 42, -102 until -98	∼600	0.51	Dfa, BSk	2.13	400 - 1200	Water
Location D: SE Australia	-37 until -33, 145 until 149	∼550	0.23	Cfb Cfa	2.79	50 - 1200	Water
Location E: East US	36 until 40, -89 until -85	∼1300	0.10	Cfa	0.96	100 - 800	Energy
Location F: East China	27 until 31, 114 until 118	∼1600	0.43	Cfa	0.71	30 - 1000	Energy

Table 2.3: *The specified study areas and additional information on some location specific characteristics relevant for this research. The dryness and seasonality indices were calculated as shown in Equations 2.5 and 2.6.*

2.5 Evaluation metrics

The performance of the LSMs was evaluated in three main parts. The first subquestion focused on the physical principles involved in the water balance (section 2.5.1). The second subquestion focused on the ability of the LSMs in simulating soil moisture deficits during droughts (sections 2.5.2 and 2.5.3). Lastly, the third subquestion focused on the sensitivity of the LSMs to precipitation anomalies (section 2.5.4). Section 2.5.5 focuses on how to combine these subquestions into an overall performance and answer to the research question.

2.5.1 Physical principles

First, the LSMs were evaluated on the basis of their ability to represent key physical processes involved in the water balance. This conservation of mass principle showed the LSM performance in terms of keeping the incoming and outgoing water fluxes in equilibrium. A significant negative $\frac{\Delta S}{\Delta t}$ value equaled an issue with the conservation of mass principle within the LSM and a persistent decrease in total soil moisture storage. A large positive $\frac{\Delta S}{\Delta t}$ equaled a potential numerical error and a persistent increase in total soil moisture storage. On the basis of this, a classification was possible for the LSM performance in terms of the ability of closing the water balance.

In addition, the Budyko framework showed the magnitude of the actual evaporation as a fraction of the precipitation. By comparing this magnitude to the DO and FB benchmark evaporation products, the LSMs could be ranked based on their root mean squared error (RMSE) to this reference data. Evaluating the LSMs this way and by looking at the ability to close the water balance, provided findings that could be related to certain biases, uncertainties or oversimplifications of the LSMs.

2.5.2 Water fluxes for the driest period in time

Second, to compare the accuracy of the accumulated SMDs of the LSMs with the SMDs of the reference products, the relative error was determined and compared per location during the driest period in time. The SMDs were computed as described in 2.2.4 for both the LSMs and benchmark evaporation products. The relative error was used since the SMD magnitude, and therefore also the absolute error, could vary greatly from one location to another. To determine the driest period, the combination of subsequent months with the largest decrease in SoMo.ml soil moisture was identified, which was bound between a local maximum and minimum. Table 2.4 shows the driest period for all study areas.

In addition to comparing the SMDs, which are based on the evaporation output of the LSMs, the soil moisture and runoff output during this driest period were considered. The LSM soil moisture for the first half meter of soil were compared with the SoMo.ml data, using the relative error as the evaluation metric. Furthermore, values for total soil moisture decline and runoff during this period were included. However, no reference data was available for these variables and comparisons could only be done between the LSMs themselves. Examination of all these water flows, in addition to evaporation, provided an overview of how the LSMs operate during droughts.

Study area	Start of dry period	End of dry period	Number of months
A: East Brazil	02-2007	10-2007	8
B: West Africa	09-2001	04-2002	7
C: Central US	05-2003	11-2003	6
D: Southeast Australia	07-2005	03-2006	8
E: East US	01-2007	07-2007	7
F: East China	05-2003	10-2003	5

Table 2.4: *The driest period in time based on the SO soil moisture dataset (see section 2.1.2).*

2.5.3 Correlation with reference soil moisture

Third, the performance of the LSMs in simulating the correct timing and progression of the soil moisture deficits was examined. This was done by computing the correlation between the SMDs of the LSMs and the reference soil moisture data during the dry periods in time. In this research, it was assumed that the reference soil moisture data set was a more superior product compared to the evaporation reference data sets. Consequently, this evaluation step was performed by computing the correlation with the reference soil moisture and not with the simulated SMDs of the reference evaporation data.

As the change in the reference soil moisture data is negative and the SMDs are greater than or equal to zero, a minus sign was placed in front of the SMDs in the figures of this research, to make the two comparable.
In this step, the assumption was that the temporal changes of the negative SMD equaled the timing and progression of the changes in observational soil moisture data. This as research states that the seasonality and therefore the correlation between the two is barely affected by the presence of run-off [Qiao et al., 2022].

For this analysis, the Pearson correlation was used, which describes the linear association between two variables. A correlation of 1.0 means a perfect positive correlation, while -1.0 stands for a perfect negative correlation. Zero is equal to no correlation at all. The Pearson correlation r is defined in Equation 2.9.

$$r = \frac{n * \Sigma(x * y) - (\Sigma x)(\Sigma y)}{\sqrt{(n * \Sigma x^2 - (\Sigma x)^2)(n * \Sigma y^2 - (\Sigma y)^2)}} \tag{2.9}$$

$$n = \text{Number of data points} \tag{2.10}$$

$$x = \text{Computed negative LSM soil moisture deficit} \tag{2.11}$$

$$y = \text{Reference soil moisture data} \tag{2.12}$$

The correlation was computed based on all dry periods within the investigated time series. For locations A and B, this was an annual returning period in time. While for study areas C, D, E and F not every year had a clear dry period in time. Therefore, for the latter locations only three dry periods within the time series were considered for this evaluation step. In addition to the numerical comparison of the Pearson correlation, a visual comparison was performed by plotting the time series during the dry periods in time for the LSM soil moisture deficits compared to the reference data. Two y-axes were used to clarify the differences between soil moisture and SMD.

2.5.4 Model sensitivity

The fourth and last step of the evaluation was to assess the sensitivity of the LSMs by looking at the relation between the GSWP3 precipitation anomalies and the LSM evaporation anomalies over the same time steps. This was done by subtracting the mean seasonality from both datasets and considering values that deviate from the seasonal pattern. To do so, the monthly mean of the 2001-2010 time series was determined and subtracted from all the data per time step. The data was then sorted by increasing precipitation anomaly. Subsequently, the corresponding evaporation anomalies were plotted over the increasing precipitation anomalies, and a linear line through the points was plotted by means of linear regression. For the fitted line, the "b" in $E = a * P + b$ equals zero, which means that the slope "a" describes the entire response of the E to P anomaly. The model sensitivities were visualized by a bar plot showing the slope of this fitted line. The two benchmark evaporation products (DOLCE V3 and FBP) were used as a reference for the evaporation anomalies. The same steps were taken for them as for the LSMs in this evaluation step. The relative error described the LSM performance with respect to this reference data.

Since DOLCE V3 and FBP are a combination of various other products, the standard deviation could be influenced by the more products that are included. Subsequently, this would influence the model sensitivity and therefore the magnitude of the slope in this analysis. Because of this, a comparison was done in the standard deviation of the anomalies from DOLCE V3 compared to the variability of ERA5, FLUXCOM and GLEAM, which DOLCE v3 consists of. This provided insight into whether combining these products influenced the model sensitivity. It was assumed that a lower standard deviation leads to a lower model sensitivity and vice versa.

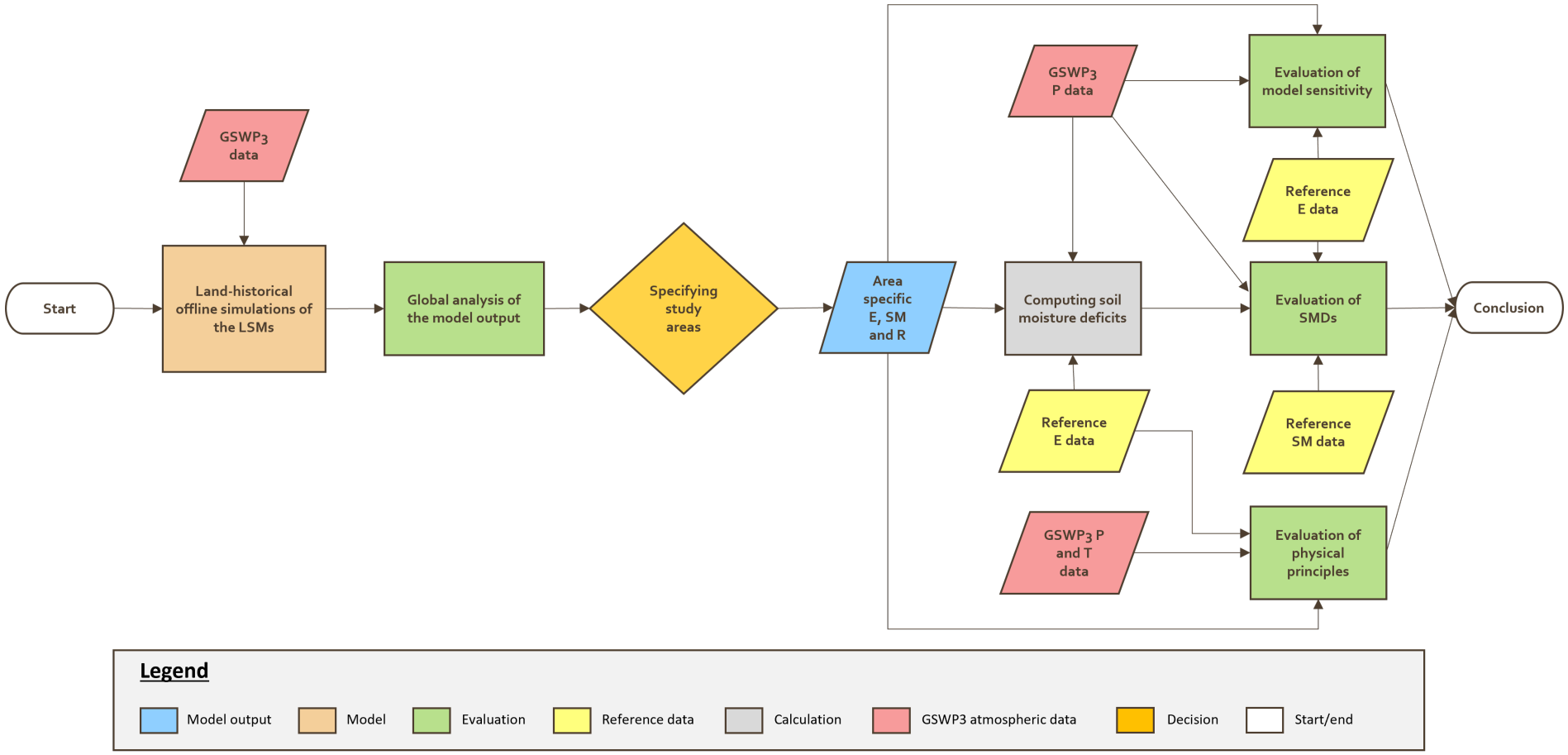
The steps taken for evaluating the evaporation anomalies in relation to the precipitation anomalies were repeated for the run-off and soil moisture anomalies. However, reference data for these variables was not considered, so their main purpose was to provide a general understanding of the model behavior under changing precipitation conditions. Additionally, the combination of these three variables provided a conclusion on the combined anomaly response caused by precipitation anomalies.

2.5.5 Overall performance

To answer the research question, it was important to evaluate the overall performance of the LSMs across all analyses. For this, their relative performance compared to each other per evaluation step was considered. The three best-performing LSMs received a '+', while numbers 4 and 5 received a '+/−', and the worst three LSMs received a '−'. For the water balance, this did not apply as it is a binary classification. A '+' in this case referred to no issues, while a '−' indicated problems with closing the water balance. It is important to keep in mind that this method was an oversimplification to clarify what the overall performances are and that they were no numerical metrics involved. Furthermore, this method considered relative performances and it did not indicate how good the best model performed per evaluation step.

The evaluation of the water balance and the Budyko framework was based on annual mean water fluxes, whereas the other analyses focused on drought periods. To rank the LSMs, their performances during droughts were given priority. The number of '+' received the highest weight, followed by the number of '+/−'. In case of a tie between LSMs for the drought analyses, their annual mean water fluxes were considered to determine the final ranking.

2.6 General workflow



3 — Results

The results discussed in this chapter had the aim of answering the research question: *What is the performance of Land Surface Models considered in LS3MIP when simulating soil moisture droughts?* Due to the nature of a soil moisture drought, the results focused on the performance of the evaporation output of the LSMs. The run-off and soil moisture were also considered in the results to provide more clarity on the working of the model. However, soil moisture was not available for all models and no reference dataset was used to evaluate the run-off.

The results focused on what the LSM evaporation characteristics were on a global scale in Section 3.1, how the model performed on physical principles in sections 3.2.1 and 3.2.2, how the model performed in terms of accumulated soil moisture deficits, change in soil moisture and run-off for the driest period in time in Section 3.2.3, how well the LSM captured the timing and progression of droughts in Section 3.2.4 and what the sensitivity was of the LSMs to precipitation anomalies in Section 3.2.5.

For clarity, in the remainder of this report the LSMs are abbreviated as discussed in section 2.1.1. For the reference products, DOLCE V3 is abbreviated as DO, FLUXCOM BESS PML as FB and SoMo.ml as SO.

3.1 Global analysis

3.1.1 Seasonality

In the comparison of the global seasonality (Figure 3.1), significant differences were observed in the overall mean among the LSMs. The UK and HG models consistently showed values well below the LSM mean, while EE and M6 exhibited values well above this value. However, this figure showed a clear difference between UK and HG in the months from November until March. HG showed a lower value in these months compared to UK while they showed a fairly similar response in the remaining months. The remaining four LSMs demonstrated similar responses to each other and the LSM mean.

The reference products DO and FB showed a slightly higher but comparable response to C2, CE, E3 and IP and the LSM mean. FB had the most data-missing areas, which led to the fact that this data set was used as a clipping mask for the LSMs and the reference product DO.

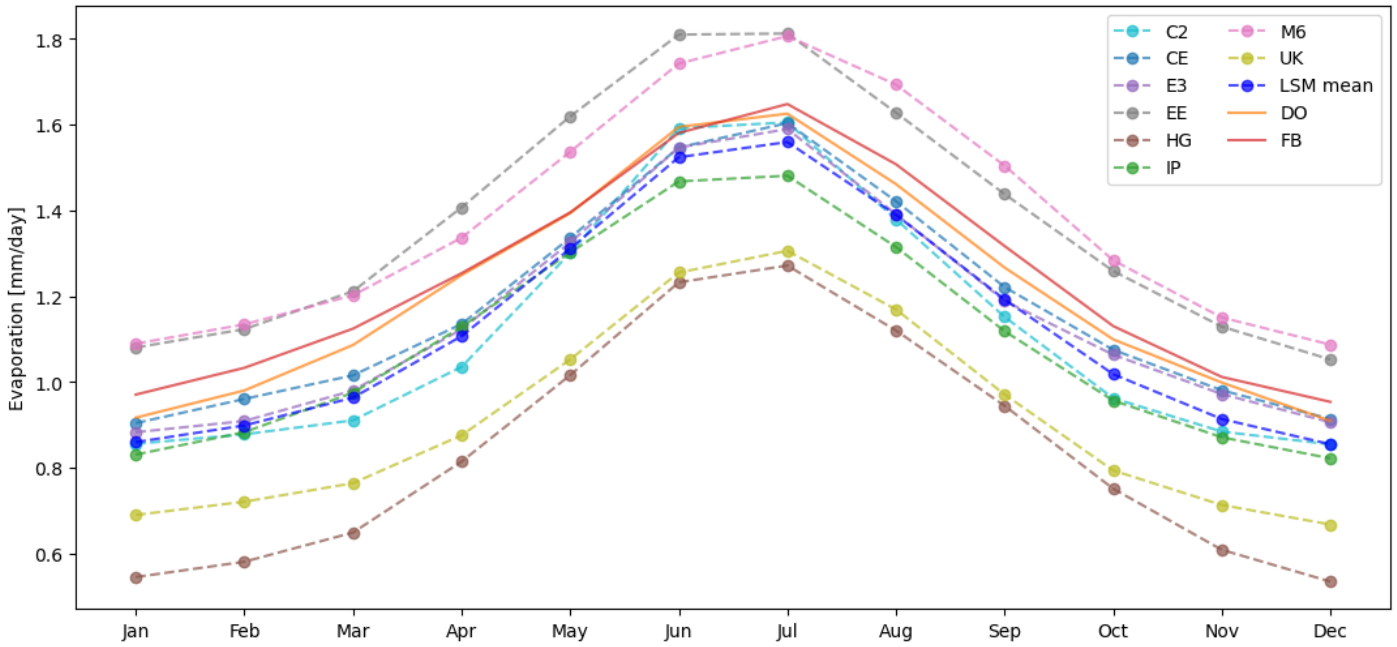


Figure 3.1: *Mean global monthly evaporation of the time series between 2001 and 2010 for the LSMs, LSM mean and benchmark evaporation products DO and FB. The LSM mean is the combination of the eights models considered in this research.*

Kruskal-Wallis statistical test

As discussed in section 2.3.1, to evaluate whether the means of the LSMs were statistically different, a Kruskal-Wallis test was performed. The data sets were independent of each other and the distributions had a similar shape, which were two important assumptions for this test. The third assumption was that the variances should be roughly equal. This was tested with the Levene's test which led to a Levene's statistic of 1.45 [-] and a p-value of .18. Therefore, since the p-value exceeds .05, the third assumption was also verified. The null hypothesis that the means were equal was rejected, as $p < .001$ with a Kruskal-Wallis statistic of 312.12 [-]. This showed that the LSM evaporation means were statistically different. Subsequently, t-tests were performed to examine the differences among the individual LSMs and whether these individual differences were significant, which can be seen in Appendix I. From this, the same similarities emerged as can be seen in Figure 3.1. Namely, that EE and M6 had the most in common, since no statistically significant difference was found. This also applied to HG and UK and the four remaining LSMs.

3.1.2 Spatial and temporal characteristics

JJA mean

The JJA mean per model compared to the LSM mean was determined as discussed in section 2.3.2. Appendix F shows similar plots, with the difference that the absolute values of the LSM evaporation are displayed.

The plots in Figure 3.2 were in line with the results found for the global seasonality (Figure 3.1). The EE and M6 showed a predominantly positive relationship while the UK and HG models had a predominantly negative relationship with the LSM mean. Since these plots show the absolute error, this was especially visible in the water rich regions (e.g. Amazon rain forest). E3 and IP showed the least differences compared to the LSM mean while the C2 and CE models have some clear positive and negative relations compared to the LSM mean. Besides the similarities with Figure 3.1, these plots showed some contrasting areas for EE and UK compared to M6 and HG. East Brazil, around the Democratic Republic of the Congo and Northern Australia were areas in which the minus sign reversed. The same areas stood out for the C2 and CE plots, since C2 had a positive and CE a negative relation to the LSM mean for these locations.

The reference product was higher than the LSM mean around Colombia, the Democratic Republic of the Congo, the Tibetan plateau and Myanmar. While it showed a negative relation in the Southeast of Brazil, the Sahel and the Eastern part of the African continent.

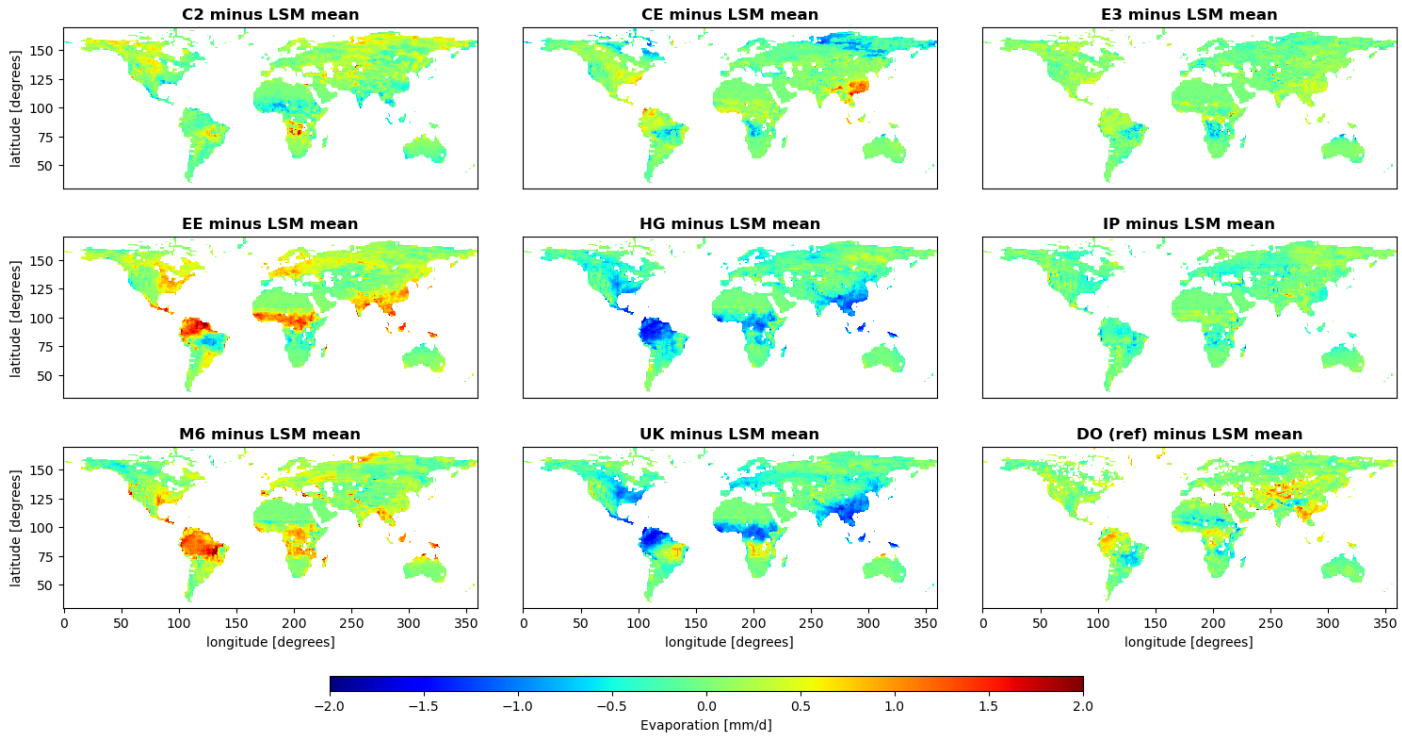


Figure 3.2: *Spatial characteristics of the individual mean LSM evaporation output as well as the DO reference product compared to the combined LSM mean for the* ***J****une,* ***J****uly and* ***A****ugust months between 2001 and 2010 in mm per day*

DJF mean

The DJF mean per model compared to the LSM mean was determined as discussed in section 2.3.2. Appendix F shows similar plots with the difference that the absolute value is displayed.

Just as for the JJA plots, the DJF mean plots in Figure 3.3 were in line with the results found for the global seasonality (Figure 3.1). Again, the EE and M6 showed a predominantly positive relationship while the UK and HG models had a predominantly negative relationship with the LSM mean. Just as for the JJA mean, the E3 and IP models showed the least differences with the LSM mean. C2 was clearly lower than the LSM mean around the Democratic Republic of the Congo and Northern Australia while CE primarily showed a positive relation in the Southeast of Brazil. Just like for the JJA plots, the UK model had some areas in which the model was higher compared to the LSM mean. Again, a clear difference could be seen with HG in these regions. This was mostly the case for the area just below the Sahel region and around the Bay of Bengal. Although somewhat less clear, the same areas had a negative relation compared to the LSM mean for EE, which therefore followed the same pattern as the results found in section 3.1.2. Namely, that the LSMs with a similar global mean bias differ from each other in certain regions around the equator.

The DO reference data was clearly higher than the LSM mean over the Amazon, around the Democratic Republic of the Congo and Indonesia. It showed a clear negative relation to the mean over the Northern part of Australia.

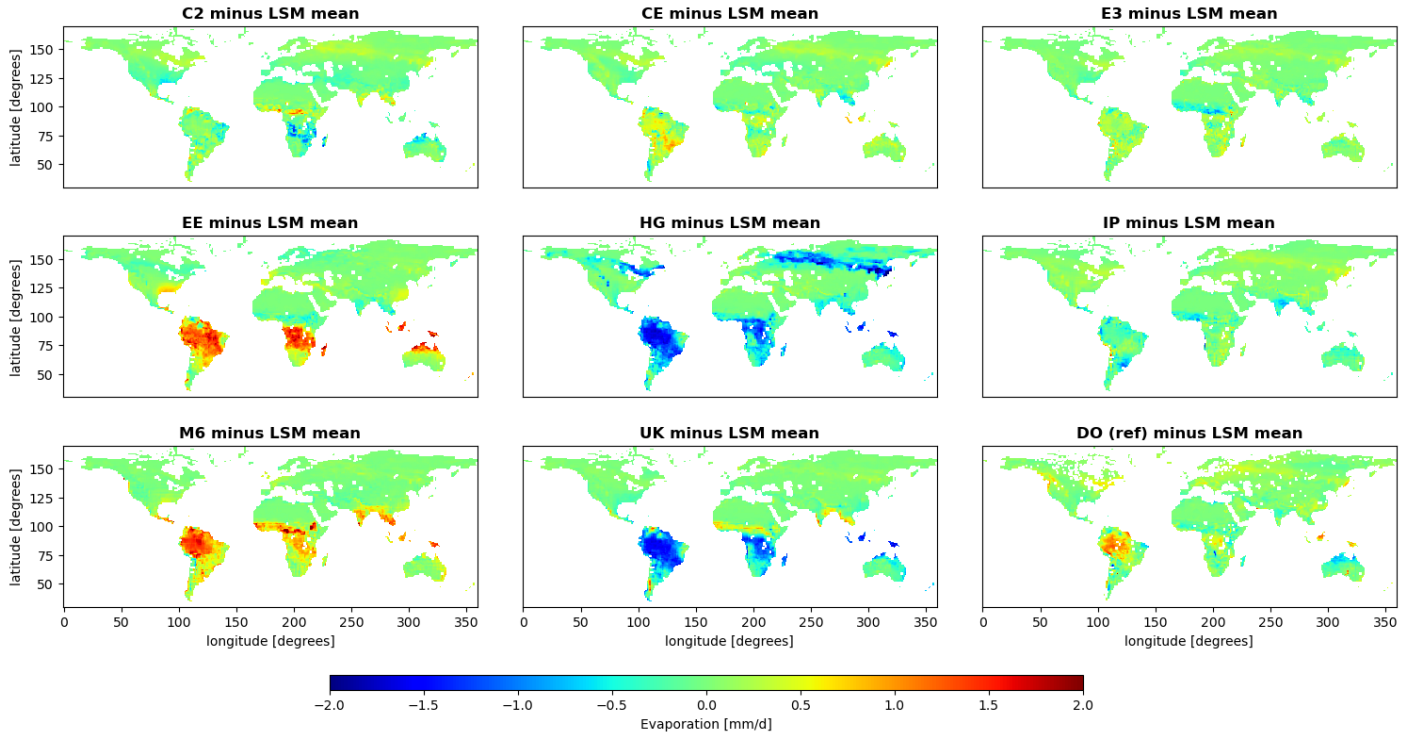


Figure 3.3: *Spatial characteristics of the individual mean LSM evaporation output as well as the DO reference product compared to the combined LSM mean for the* ***D****ecember,* ***J****anuary and* ***F****ebruary months between 2001 and 2010 in mm per day*

3.2 Analysis of study areas

The study areas were specified as discussed in section 2.3.3 and further clarified in section 2.4.

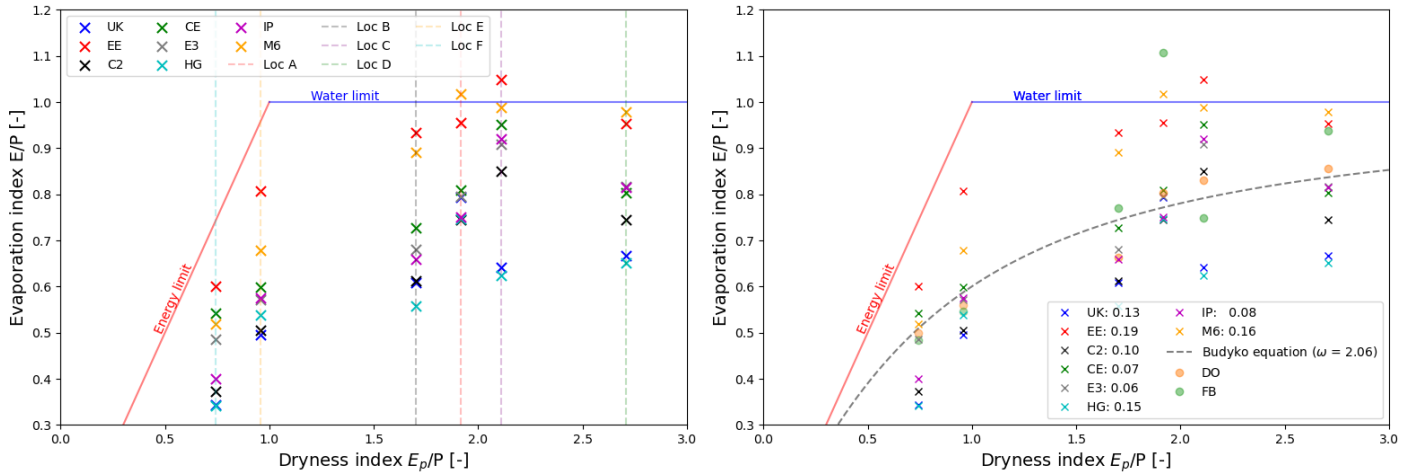
3.2.1 Budyko framework

The Budyko framework was constructed as discussed in section 2.2.3. All points were on the same vertical line per location due to the constant atmospheric forcing in the model. As a result, both P and E_p were constant for all LSMs per location.

Figure 3.4a showed that there was a significant spread in terms of the evaporation index among the LSMs. The average difference between the highest and lowest value for the evaporation index was equal to 0.33 [-]. Despite this significant spread, the distribution of LSMs remained fairly similar over all locations. For instance, HG and UK were always on the low side, while EE and M6 had the highest evaporation indices. From this, it could be concluded that the dryness index did not play a significant role in the distribution of LSMs in terms of the evaporation index and that the initial bias of the LSMs was preserved over varying Dryness indices. For location A, M6 exceeded the water limit which meant that the annual mean evaporation exceeded the annual mean precipitation value. The same was observed for EE at location C. The energy limit was not exceeded by any LSM at any location.

Figure 3.4b showed that the LSMs, in general, extended both above and below the calibrated Budyko equation per location. This with the exception of location A for which most of the LSMs were positioned above this line. FB showed an outlier well above the water limit in location A. Therefore, this point was not included in the calibration of the Budyko equation. Apart from location A, the reference data points were located relatively close to each other.

The root mean squared errors from the LSMs compared to the calibrated Budyko equation showed that E3 (0.06), CE (0.07), IP (0.08) and C2 (0.10) were the four best performing LSMs. EE (0.19), M6 (0.16), HG (0.15) and UK (0.13) were the four least performing LSMs in terms of the RMSE. These results were consistent with the findings of the seasonality in section 3.1.1, as these four LSMs were also the furthest from the reference data in that part of the analysis.



(a) *(b)*

Figure 3.4: ***(a)*** *The Budyko framework for all LSMs considered in this research. The vertical lines represent the various study areas. The red and blue lines represent the physical boundaries of the energy and water limit.* ***(b)*** *The Budyko framework for all LSM and the reference products DO and FB for all study areas. The ω parameter in the Budyko equation (Eq: 2.3) is calibrated based on the reference data. The RMSE is calculated for each individual LSM compared to this calibrated line.The red and blue lines represent the physical boundaries of the energy and water limit.*

3.2.2 Water balance

Figure 3.7 shows the mean annual water fluxes involved in the water balance (Eq: 2.2) for each LSM per study area. This figure shows that the majority of LSMs had a water balance close to, or equal to zero over the 10 year time period. However, UK, HG and CE were the exception to this. UK and HG showed a persistent increase in water storage as the incoming precipitation was significantly higher compared to the evaporation plus run-off. This was the case for locations B, C, D, E and F. In general, HG and UK had a relatively low evaporation which was not compensated in the water balance by a relatively high run-off. In contrast, for locations C and D, UK and HG showed a relatively low runoff.

CE showed a persistent decrease in water storage for locations B, E and F, since the incoming precipitation was lower compared to the evaporation plus run-off. This was mainly due to the relatively high run-off, especially for locations E and F. The CE evaporation was around the LSM average for these locations.

The remaining five LSMs had a water balance close to or equal to zero for all locations. However, what stood out was the magnitude of the run-off for EE and M6. EE had a run-off of almost zero for location C while the evaporation had already exceeded the precipitation. The same pattern was visible for M6 for location A. In general, EE and M6 showed a relatively high evaporation flux and a relatively low run-off. The high evaporation flux for these two LSMs was in line with the results found in sections 3.1.1, 3.1.2 and 3.2.1. Figure E.2 in appendix E displays the changes in storage in the investigated time series due to the fluxes involved in the water balance. Figure E.4 demonstrates that the contrasts observed in terms of the water balance do not match the results for the simulated total soil moisture, as all LSMs remain roughly stable over time.

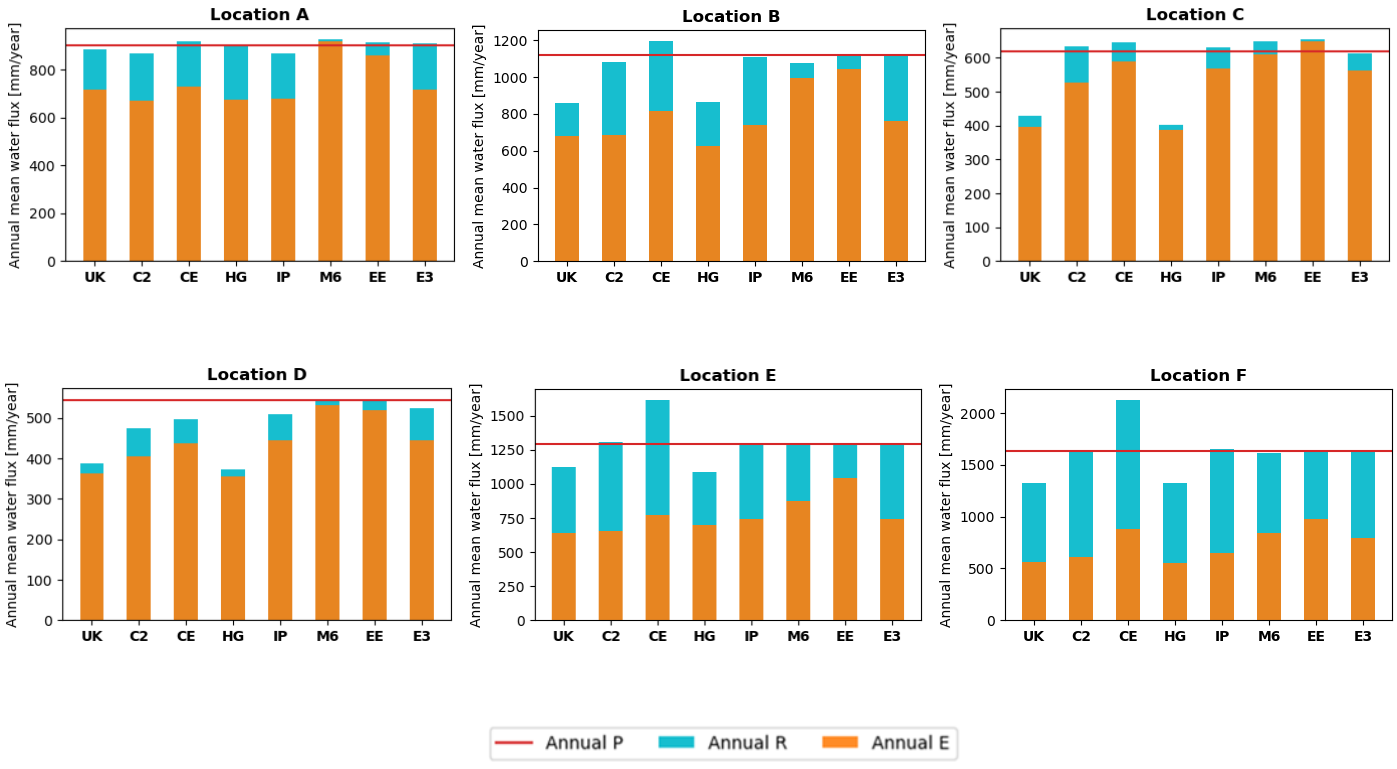


Figure 3.7: *The annual mean water fluxes between 2001 and 2010 involved in the water balance (Eq: 2.2) from all LSMs for all locations. The red line represents the annual precipitation. The evaporation (E) and run-off (R) are plotted on top of each other to show the sum of the two fluxes. The water balance is zero if the sum of the mean annual evaporation (E) and run-off (R) equals the mean annual precipitation (P).*

3.2.3 Water fluxes for the driest period in time

This section will zoom in on the LSM performance during droughts. For all study areas, the driest period within the investigated time series was determined as discussed in section 2.5.2.

Soil moisture deficits

Figure 3.8 shows a large spread of the SMDs in the LSMs. At all study areas there were significant differences between the accumulated soil moisture deficits during the driest period in time. The SMD of M6 was relatively large for locations A, B and C while the ones of UK and HG were relatively small for locations C, D and E. The mean relative error showed that C2 (0.17), CE (0.18), E3 (0.24) and EE (0.30) were the four best performing models compared to the reference data sets. IP (0.37), UK (0.41), HG (0.46) and M6 (0.62) performed the least compared to the reference data. The SMD value of FB for location A was excluded from the mean relative error calculation as Figure 3.4b showed that the reference data was well over the water limit for this location. The SMD value of FB at location A was in line with that finding, as it was relatively high. Apart from this location, the reference data sets did not deviate much from each other.

The results of the analysis did not entirely match with what was found in sections 3.1.1, 3.1.2 and 3.2.1, which were not focused on droughts. It became clear that M6 and EE both had relatively high rates of evaporation compared to other LSMs in Figures 3.1 and 3.4b. However, Figure 3.8 showed less agreement between these two LSMs. The high evaporation rate of M6 could be linked to the large SMD values and the overall worse performance compared to the reference data. On the other hand, EE had a lower relative error and was the fourth best performing model. This showed that EE evaporates less during dry periods and more during wetter periods in time compared to M6. This was especially true for locations A and C, which was confirmed by looking at the monthly mean LSM evaporation per study area (Figure B.3 in Appendix B). IP worsened slightly compared to the results found in the Budyko framework (3.4b) which underestimated the SMD for all locations. At the same time, HG and UK were grouped as LSMs with relatively low evaporation. However, for locations A and B, the SMDs were relatively large while they were relatively small for the remaining locations. For locations A and B, this showed that the UK and HG had a relatively high evaporation during the dry periods in time for these two locations while the Budyko framework of Figure 3.4b showed that these LSMs were at the low end in terms of the annual evaporation index for these two locations compared to the other LSMs. This was also confirmed in Figure B.3 in Appendix B. Therefore, these results showed that the annual pattern in terms of evaporation does not always match the LSM performance during a dry period in time.

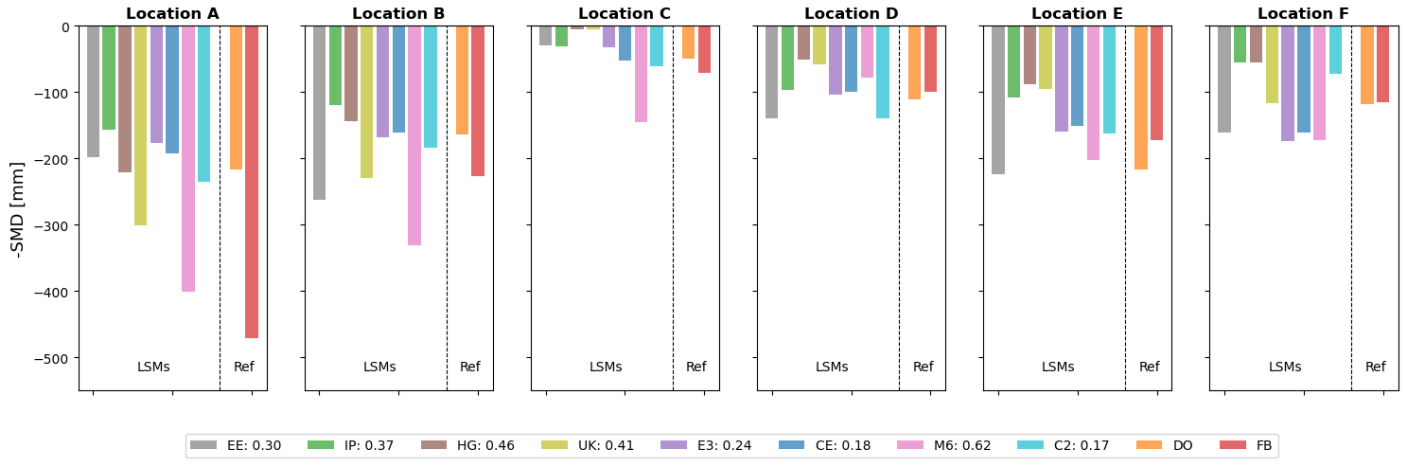


Figure 3.8: *The determined negative soil moisture deficit (Eq: 2.4) for the driest period within the investigated time series for all LSMs and reference products DO and FB. The numerical values in the legend represent the mean relative error [-] between the LSM compared to the average of the two reference products.*

Change in soil moisture

To compare the LSM soil moisture output with the SO reference data, only the first half a meter of soil was considered. Figure 3.9 showed that different patterns were visible for the change in soil moisture for the first half a meter of soil compared to the computed SMDs. For instance, M6 had the smallest relative error (0.29) compared to SO while it had the largest relative error for the difference in SMDs. Furthermore, C2 showed the largest relative error (0.84), while it had the lowest relative error for the difference in SMDs. All in all, according to this analysis, the M6, CE and UK LSMs performed the best for modeling the decrease in soil moisture for the first half a meter of soil, while C2, HG and IP performed the worst.

However, it was important to note that these results did not show the complete picture of the decrease in soil moisture. The total decrease in soil moisture over the entire depth and the comparison of this decrease to the SMDs during this dry period are displayed in figure 3.10 and Table D.2 (Appendix D). These results showed that the overall decrease in soil moisture was not always in line with the decrease for the first half a meter of soil. This showed that each LSM had a different decline in soil moisture over the depth. For example, it showed that M6 had the smallest decrease for the first half a meter for location C, while it had the largest decrease over the entire soil depth. Furthermore, it was important to note that the models differed in the total depth over which the soil moisture was calculated. For instance, IP, HG and UK used a soil layer of 2 meters, while CE used a soil layer of more than 35 metres. The relevant soil depths of all LSMs are displayed in Table D.1.

Overall, from figures 3.9 and 3.10 in combination with Table D.2 it could be concluded that the differences between the LSMs were less significant than for soil moisture deficits. This applies both for the first half a meter and over the entire soil depth. Furthermore, the results did not clearly follow the patterns found in the previous results. However, it became clear from these results that the distribution of soil moisture decrease over the depth can vary greatly from one LSM to another.

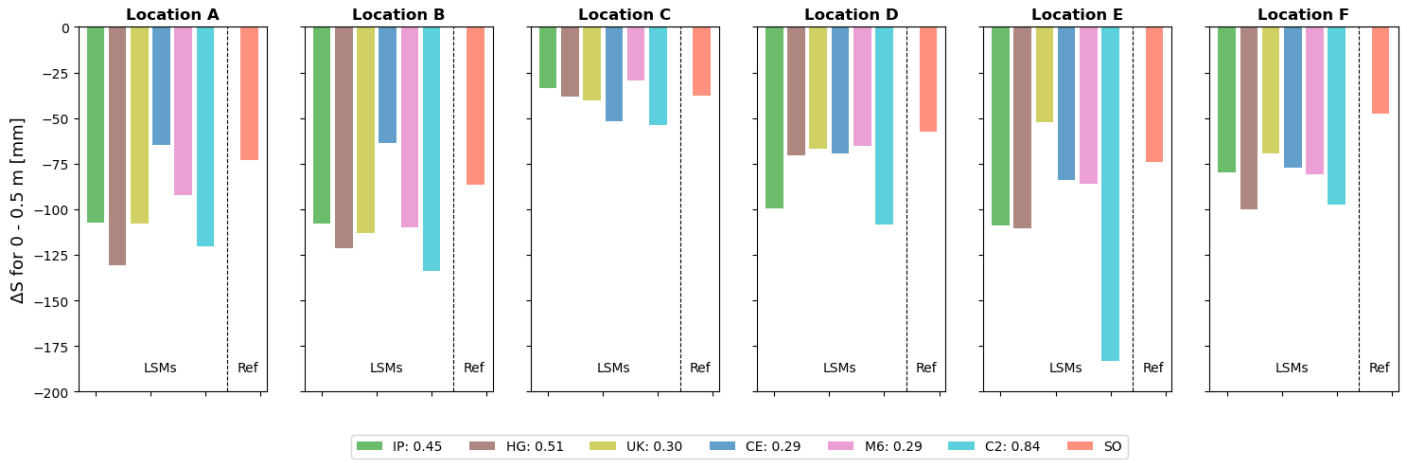


Figure 3.9: *The LSM soil moisture output for the first 0.5 meter of soil for the driest period in time, compared to the SO reference data set. Please note that no soil moisture output was available for this research from the EE and E3 LSMs. The numerical value in the legend represents the mean relative error [-] between the LSM compared to the reference product.*

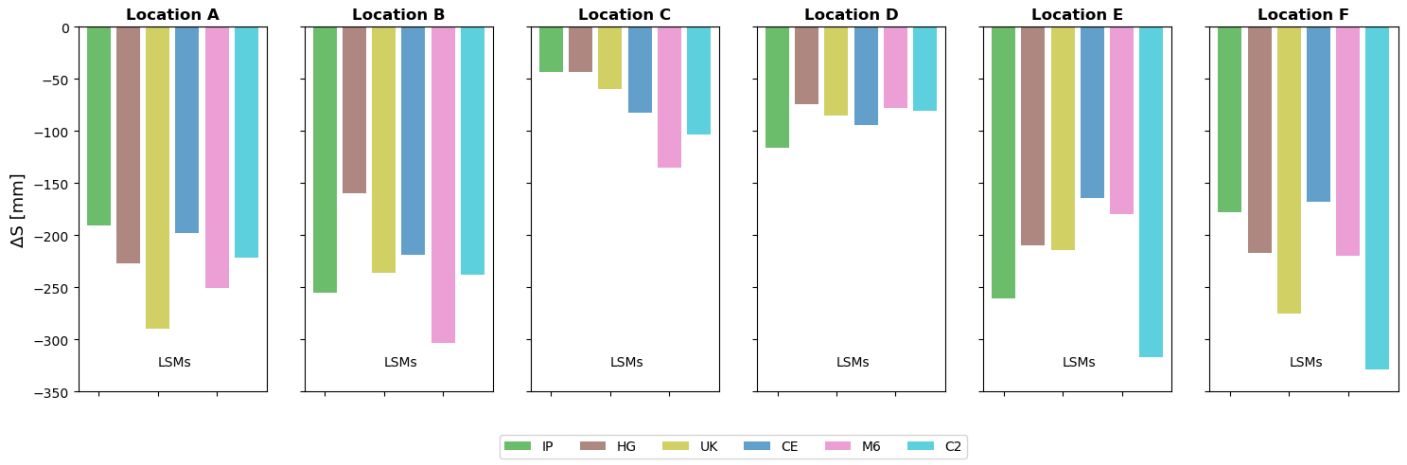


Figure 3.10: *The LSM soil moisture decrease over the entire soil depth for the driest period in time. Please note that no soil moisture output was available for this research from the EE and E3 LSMs.*

Total run-off

The total run-off for the driest period in time per location, shown in figure 3.11, in general, followed the results found for the run-off within the water balance in section 3.2.2. This showed that there were no major differences in the patterns of annual run-off and run-off during droughts. Both results had some significant variation between the LSMs for all locations. Furthermore, both results showed that the CE and C2 run-off were relatively high for all locations while it was relatively low for EE and M6. The remaining LSMs did not show major differences either.

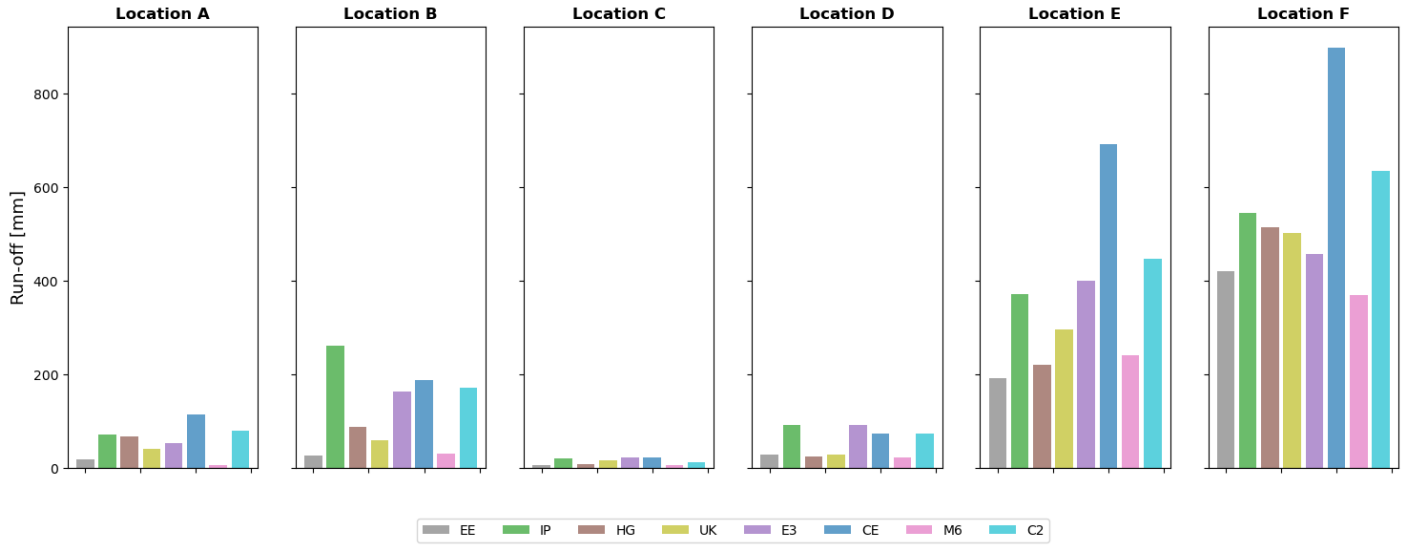


Figure 3.11: *The individual LSM run-off output over the driest period in time per location for all study areas. Please note that no reference data set is used to evaluate this output.*

3.2.4 Correlation with reference soil moisture

The results for this section were obtained based on the methods discussed in section 2.5.3. The results showed the correlation between the reference soil moisture and the calculated soil moisture deficits of the LSMs to evaluate whether it matched the timing and progression of the dry periods. Please note that the absolute values of the change in soil moisture and the SMDs were not comparable. For locations A (Figure 3.12) and B (Figure 3.13) there was an annual returning wet and dry period. For locations C (Figure 3.14), D (Figure 3.15), E (Figure 3.16) and F (Figure 3.17), a less obvious annual returning drought was visible. For this reason, only the three driest periods in time were considered for these locations.

Figures 3.12 and 3.13 display the water limited locations in the tropics (A and B), which showed that the SO reference data decreased strongly at the start of the dry period and weakened as the end of the dry period approached. The LSMs with the highest correlation were able to match this pattern over time to some extent. On the other hand, the LSMs with the lowest correlation showed a rather linear progression over time with almost the same gradient at the beginning as at the end of the dry period. As a result, these LSMs were unable to match the pattern of the reference SO data set. This pattern was similar for both locations, and in general a higher soil moisture deficit value was reached due to this persistent linear progression over time. The performance in correlation could be related to the results found in Figure 3.8, where the four LSMs with the lowest correlation had the highest SMD values for location A. In terms of correlation, IP, E3, CE, and EE were the best performing LSMs, while HG, C2, M6, and UK were the least performing.

Figures 3.14 and 3.15 show the water limited locations in the extratropics (C and D), where the soil moisture decline was slower compared to locations A and B. The LSMs that had the highest correlation were able to accurately capture the progression of the dry period over time. Lower correlation values indicated that the LSM was weak or unresponsive to changes in soil moisture. Therefore, a lower correlation value equaled a lower value for the SMD. In HG and UK, the SMD remained around zero during most of the dry periods. The exception to this was M6, which had a significant SMD at the beginning of the dry periods due to its relatively high evaporation rate. The results were similar for both locations, with CE, E3, and EE performing the best in terms of correlation, while M6, HG, and UK were the least performing LSMs.

The energy limited locations (E and F) depicted in Figures 3.16 and 3.17 showed similar results to those for locations C and D. The progression of the reference data was similar, and the LSMs showed a similar response. As in C and D, a lower correlation indicated a weak LSM response and a lower SMD buildup value. However, there was a notable difference in the time when the LSMs started accumulating SMD. While most LSMs remained at zero as the reference data values declined, high correlation LSMs began to follow this decline from an earlier stage. In terms of correlation, CE, E3, and EE were the best performing LSMs, while C2, IP, HG, and UK were the least performing.

Overall, when all locations were considered, E3, EE, and CE showed the highest correlation performance. IP, on the other hand, only performed well in locations A and B, while HG and UK had low performance overall. As for C2 and M6, they were found to be around average in terms of performance. For locations A and B low correlation was equal to a relatively high SMD while it was equal to a relatively low SMD for locations C,D,E and F. Partly because of this, no clear difference was found between water limited locations C and D and energy limited locations E and F. However, a clear difference was visible between the tropical (A and B) compared to the extratropical (C, D, E and F) locations.

Location A

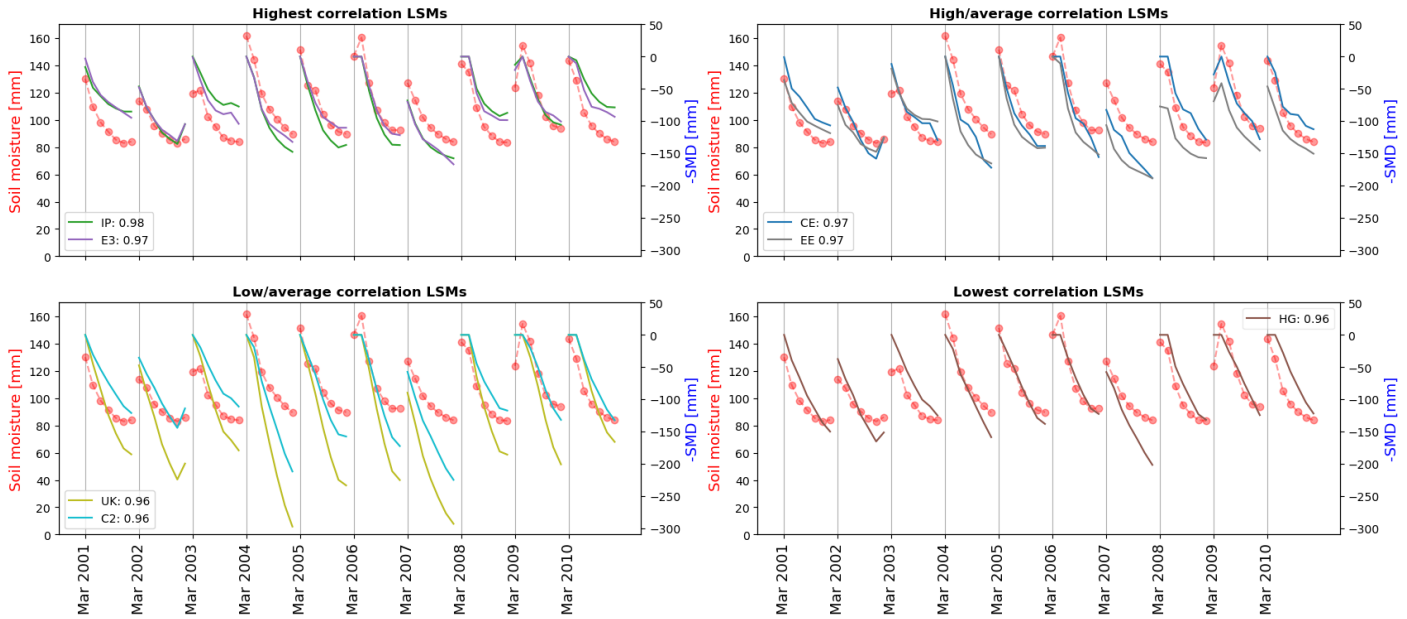


Figure 3.12: *Location A: LSM negative SMDs compared to the SO soil moisture over all selected dry periods between 2001 and 2010. The numerical value represents the Pearson correlation coefficient between the two. Please note the two different y-axis that are used. A dot in the reference data equals one monthly data point.*

Location B

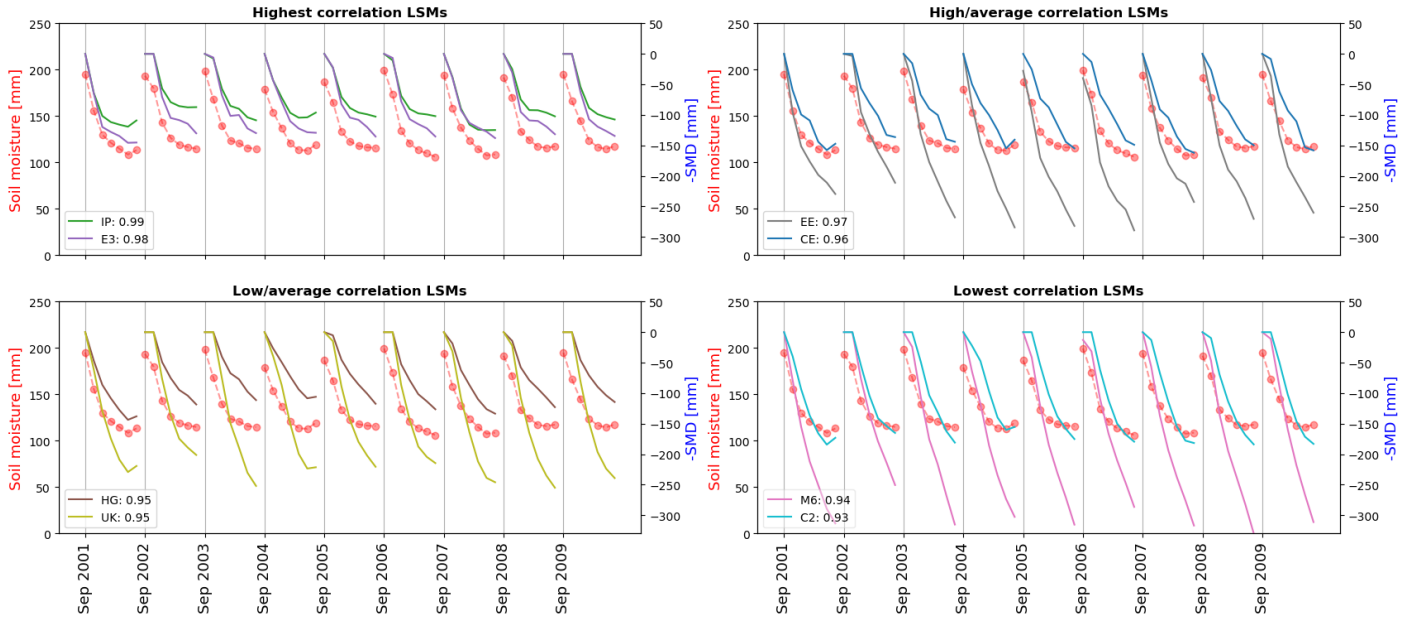


Figure 3.13: *Location B: LSM negative SMDs compared to the SO soil moisture over all selected dry periods between 2001 and 2010. The numerical value represents the Pearson correlation coefficient between the two. Please note the two different y-axis that are used. A dot in the reference data equals one monthly data point.*

Location C

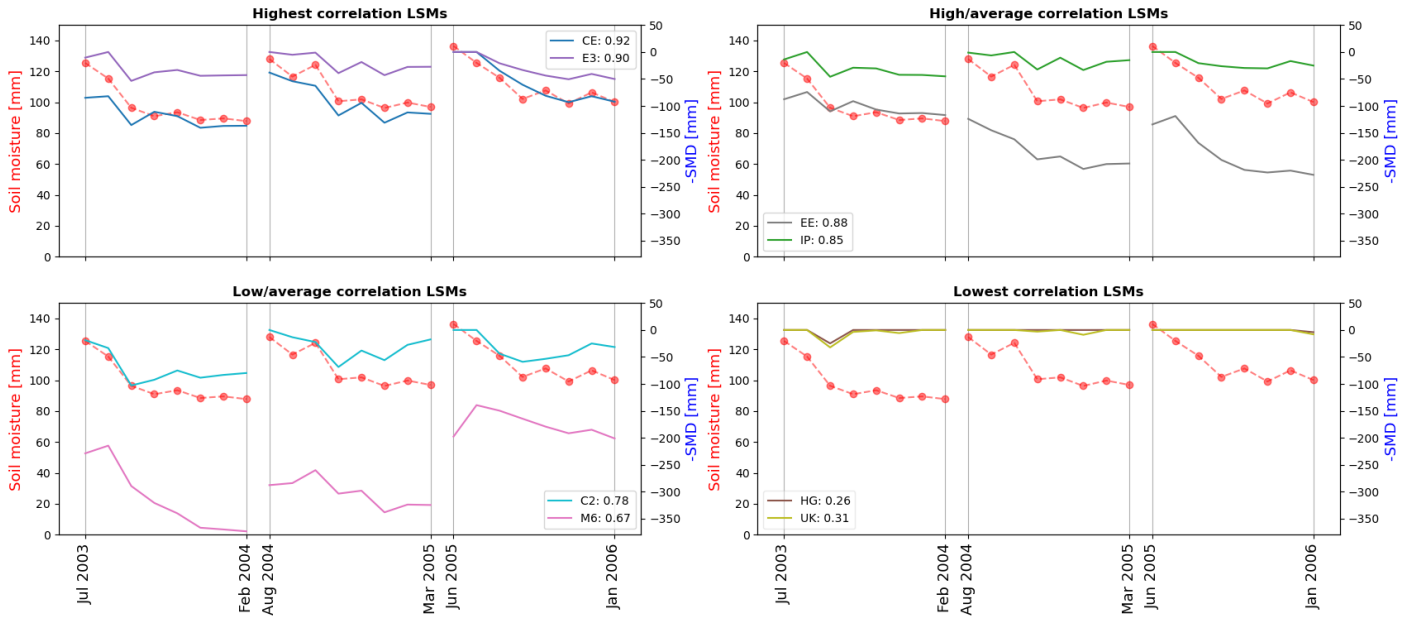


Figure 3.14: *Location C: LSM negative SMDs compared to the SO soil moisture over all selected dry periods between 2001 and 2010. The numerical value represents the Pearson correlation coefficient between the two. Please note the two different y-axis that are used. A dot in the reference data equals one monthly data point.*

Location D

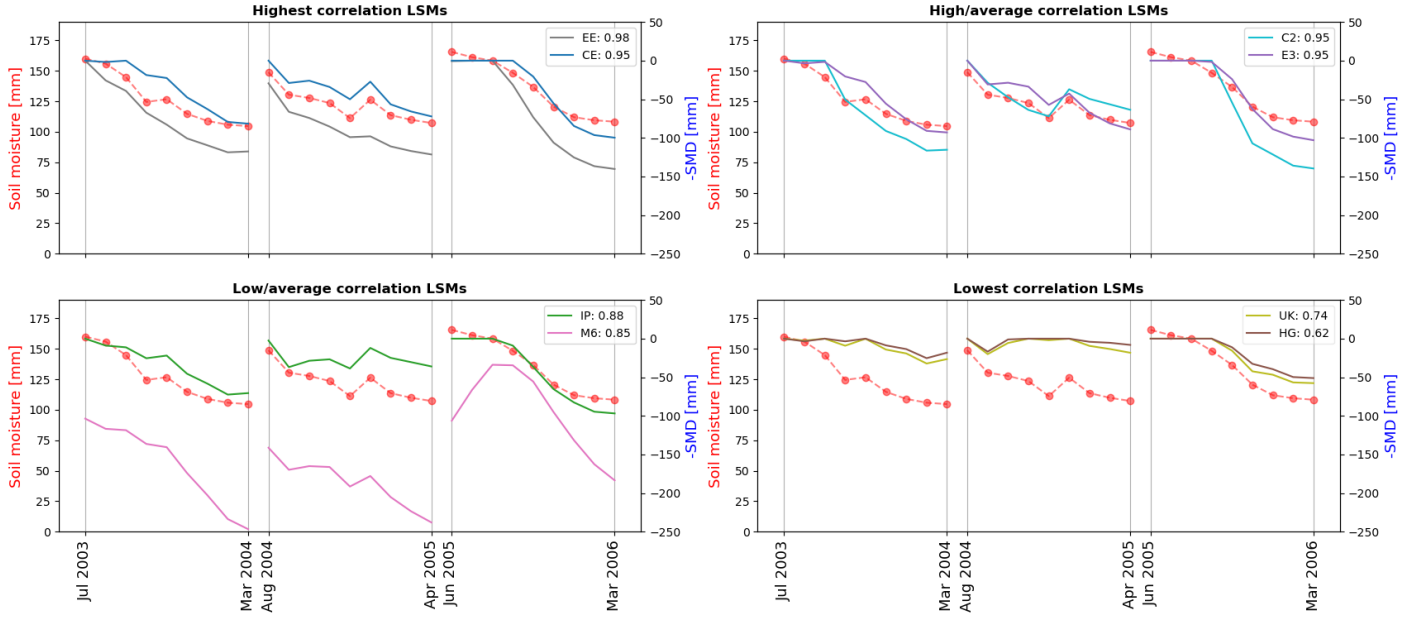


Figure 3.15: *Location D: LSM negative SMDs compared to the SO soil moisture over all selected dry periods between 2001 and 2010. The numerical value represents the Pearson correlation coefficient between the two. Please note the two different y-axis that are used. A dot in the reference data equals one monthly data point.*

Location E

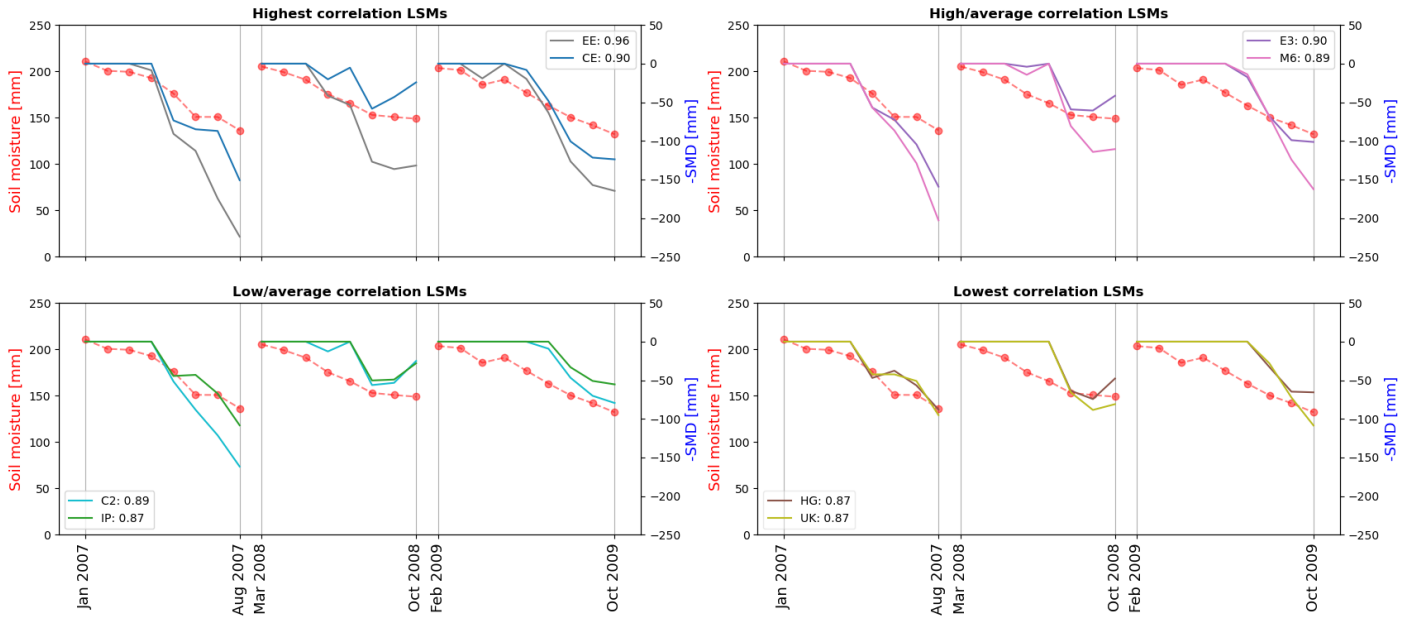


Figure 3.16: *Location E: LSM negative SMDs compared to the SO soil moisture over all selected dry periods between 2001 and 2010. The numerical value represents the Pearson correlation coefficient between the two. Please note the two different y-axis that are used. A dot in the reference data equals one monthly data point.*

Location F

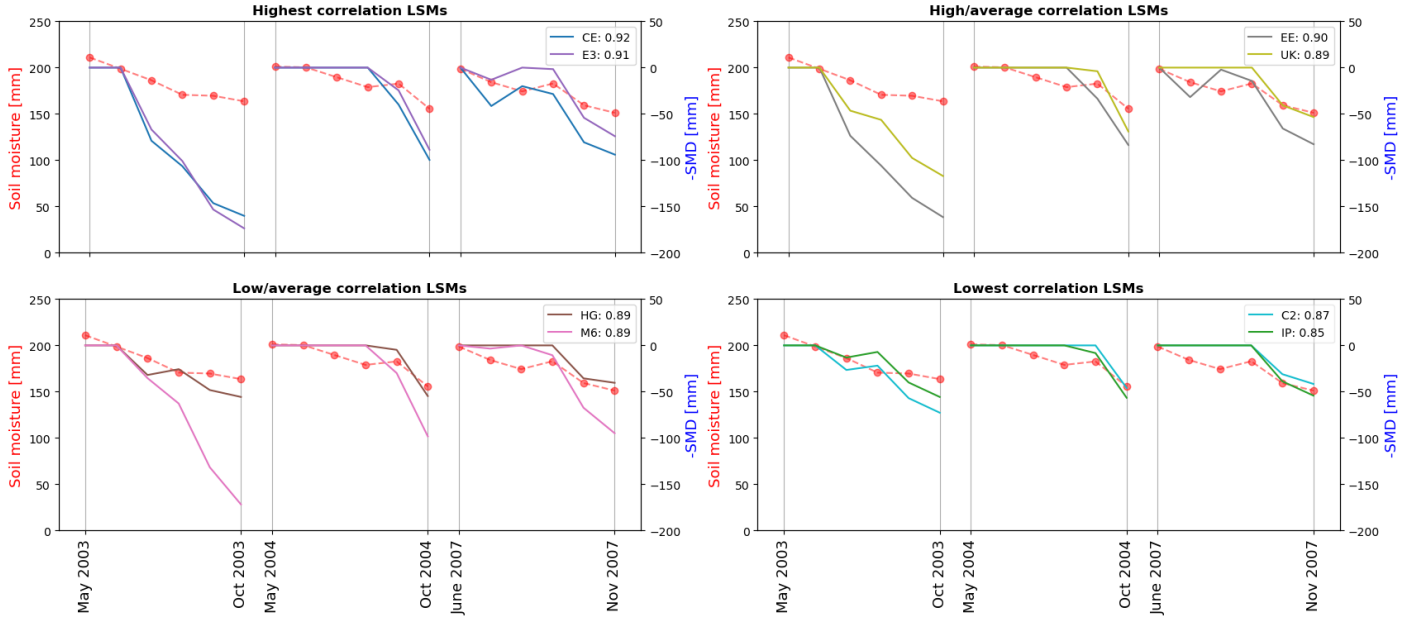


Figure 3.17: *Location F: LSM negative SMDs compared to the SO soil moisture over all selected dry periods between 2001 and 2010. The numerical value represents the Pearson correlation coefficient between the two. Please note the two different y-axis that are used. A dot in the reference data equals one monthly data point.*

3.2.5 Model sensitivity

The model sensitivity to precipitation anomalies for each LSM and the reference evaporation data was determined as discussed in section 2.5.4.

Evaporation

Figure 3.18 showed that there is a clear link between the dryness index and the $P - E$ anomaly response. The driest study areas showed a relatively high slope, while the wettest study areas (E and F) showed an average slope close to zero. Overall, EE was the LSM with highest sensitivity for all study areas. UK was the model with the lowest sensitivity for locations A, B, E and F. The best performing LSM, in terms of average relative error compared to the reference data sets, was C2, followed by HG and IP. Furthermore, this analysis showed that the average sensitivity of the LSM was relatively high compared to that of the reference data sets for all study areas.

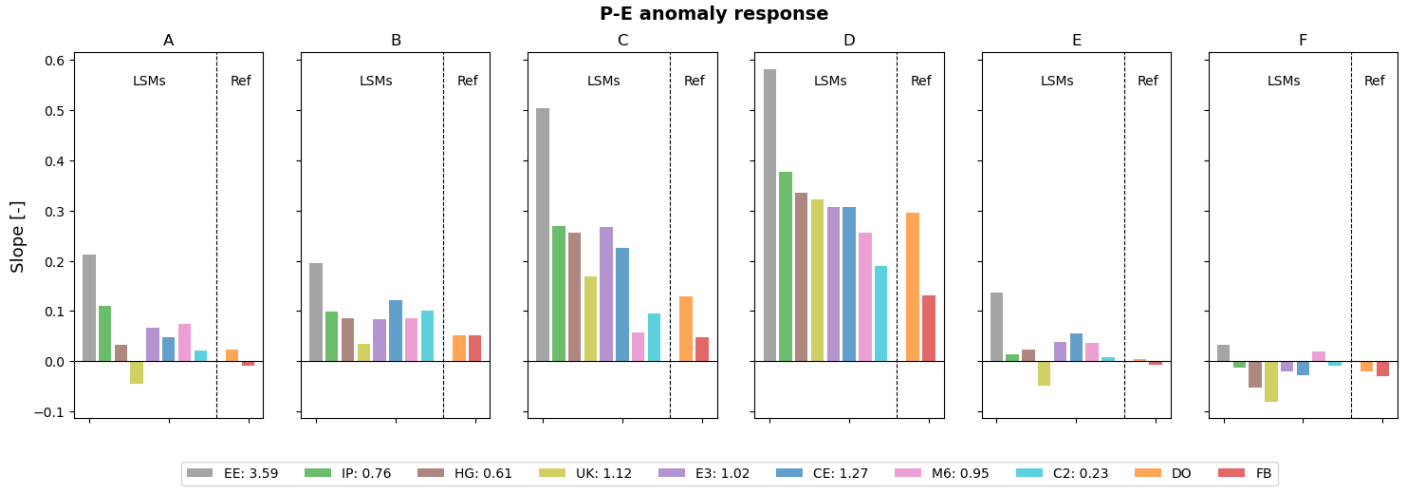


Figure 3.18: *Evaporation anomaly to precipitation anomaly response for all individual LSMs and reference products DO and FB for all study areas. The response is visualized as the slope of the fitted line through the data points as described in section 2.5.4. The numerical value in the legend represents the mean relative error compared to the average of the two reference products.*

Run-off

Figure 3.19 showed, in general, the opposite response for the run-off anomalies to precipitation anomalies compared to Figure 3.18. The driest locations showed the weakest response and the wettest locations showed the strongest run-off anomaly response to precipitation anomalies. EE, which showed the strongest $P - E$ anomaly response in Figure 3.18, displayed a relative weak response for the $P - R$ anomalies. However, this was evident for EE and was not an observed pattern for all LSMs. Furthermore, Figure 3.19 clearly showed similarities in distribution among the LSMs compared to the run-off during the driest period in time (Figure 3.11) and the annual average run-off (Figure 3.7).

Combining evaporation, run-off and soil moisture

The model sensitivity of soil moisture anomalies to precipitation anomalies, shown in Figure 3.20 showed no clear link to Figures 3.18 and 3.19. In addition, less spread was observed in the slope of the $P - SM$ response among the LSMs compared to the $P - E$ and $P - R$ response. IP, CE and C2 showed a relatively high $P - SM$ response for the driest locations C and D, implying that these LSMs make the most use of the soil moisture storage under these conditions. The cumulative slope of the E, R and SM response was also in general higher for these three LSMs. Furthermore, location A deviated from the other locations since it had a relatively low cumulative slope for the combined responses.

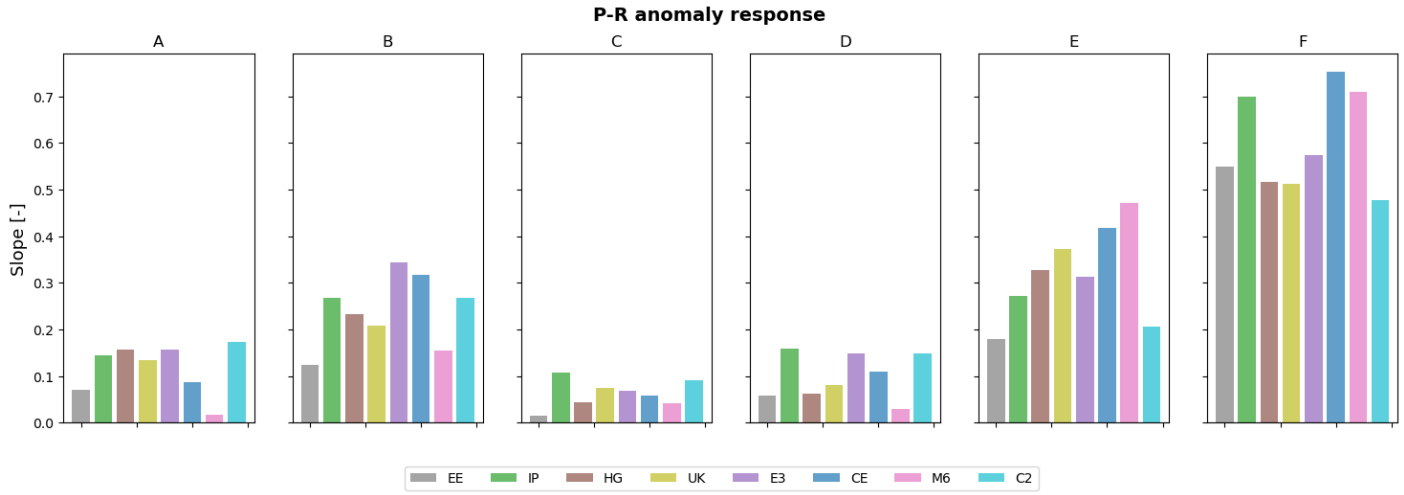


Figure 3.19: *Run-off anomaly to precipitation anomaly response for all individual LSMs for all study areas. The response is visualized as the slope of the fitted line through the data points as described in section 2.5.4.*

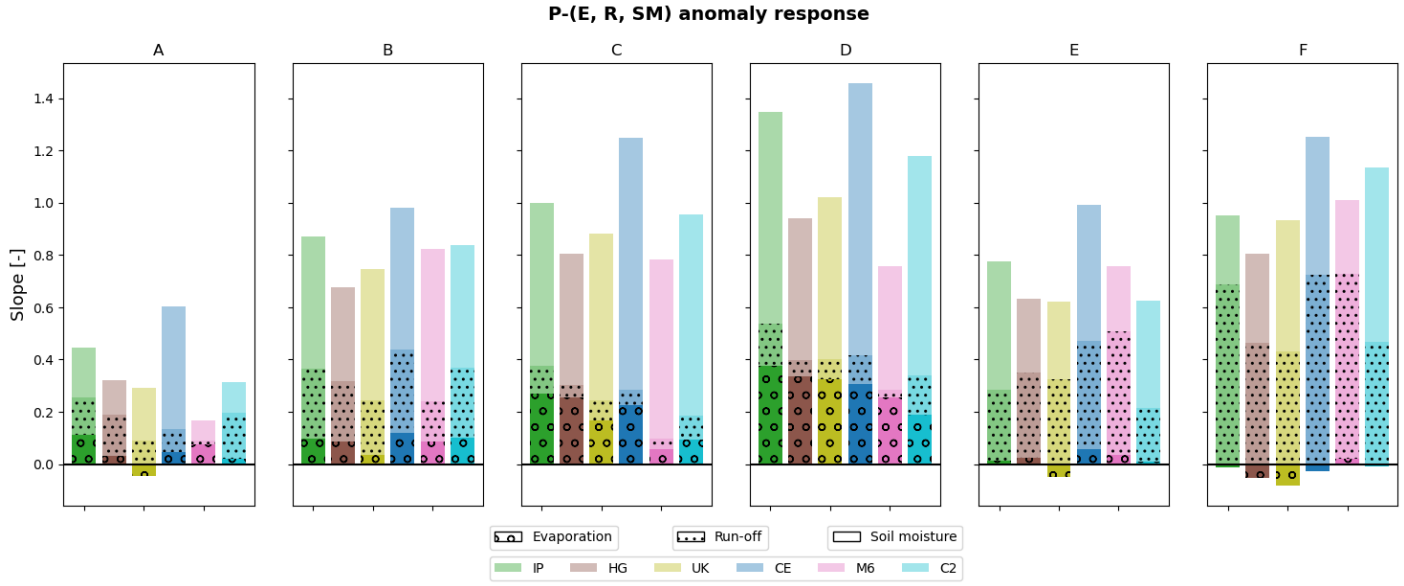


Figure 3.20: *Evaporation, run-off and soil moisture anomaly to precipitation anomaly response for all individual LSMs except EE and E3 for all study areas. The response is visualized as the slope of the fitted line through the data points as described in section 2.5.4*

3.2.6 Overall performances

Table 3.1 displays the relative performances of all LSMs considered, for all evaluation steps. For the annual mean water fluxes, E3 and IP performed best by closing the water balance and by being relatively close to the reference data in the Budyko framework. HG was the worst performing model in this regard. During droughts, E3 and C2 were the best performing LSMs, followed by CE, IP, and EE, respectively, while HG, M6, and UK were the least performing LSMs. Overall, as discussed in section 2.5.5, the LSMs could be ranked on their performance during droughts. This led to the following order of LSMs: E3, C2, CE, IP, EE, HG, M6, and UK.

For the overall LSM performance, E3 was the best performing LSM over all analyses, looking at the long term analyses as well as specifically during droughts. C2 performed slightly worse for the correlation with the reference soil moisture and the Budyko Framework, while CE had issues with closing the water balance. Besides this, these two LSMs performed relatively well. IP performed well on an annual basis but worsened for the performance during the drier months. EE performed poorly on an annual basis but did relatively well during the drier periods in time. M6 performed relatively poorly for both the annual averages as well as during the dry periods in time. UK and HG, together with M6, performed the worst overall while UK and HG also had issues considering the water balance.

	Long term analyses		Drought analyses		
LSM	**Water balance**	**Budyko**	**Accumulated SMD**	**Correlation SMD**	**Model sensitivity**
C2	+	+/−	+	+/−	+
CE	−	+	+	+	−
E3	+	+	+	+	+/−
EE	+	−	+/−	+	−
HG	−	−	−	−	+
IP	+	+	+/−	+/−	+
M6	+	−	−	−	+/−
UK	−	+/−	−	−	−

Table 3.1: *All considered LSMs and their performances for the various evaluation steps as discussed in section 2.5.5. The performances are relative as the three best performing LSMs got a '+', numbers 4 and 5 a '+/−' and the worst three LSMs a '−'. For the water balance, this does not apply as this is a binary classification. A '+' means no issues while a '−' equals problems with closing the water balance.*

4 — Discussion

4.1 Interpretation of results

The results obtained in this research will be discussed in three parts after which an overall conclusion will be formed.

1. The results from sections 3.1.1 and 3.1.2 showed the global characteristics of the LSM evaporation output.
2. The results from sections 3.2.1 and 3.2.2 showed annual mean water fluxes for the study areas and how they relate to some key physical principles. These results are not specifically focused on dry periods in time.
3. The results from sections 3.2.3, 3.2.4 and 3.2.5 showed the LSM performance during dry periods in time for the study areas.

Part 1

Global characteristics

The first part of the analysis provided some relevant information on the general bias of the LSMs. In addition, it was shown that the medians of all LSMs were significantly different through a statistical test. After the t-test, 3 groups of LSMs with a similar mean could be made. However, it was expected that each LSM had some initial bias in it and it was therefore more relevant to look at how the LSMs with a similar mean differed from each other in other aspects, and for this research, specifically during dry periods in time. For the LSMs with a similar bias, EE and M6 had the largest differences just below the equator for the JJA months and just above the equator for the DJF months. UK and HG showed a similar pattern. The same areas also led to differences when comparing the C2, CE, IP and E3 LSMs. The differences in the JJA months occurred in the hemisphere where it is winter at that time while it is the other way around for the DJF months. This may seem contradictory, as the hemisphere that experiences summer is expected to have the greatest evaporation and is therefore more likely to have the greatest absolute error among the LSMs.

Intertropical Convergence Zone

In these contrasting areas around the equator, precipitation falls mainly under the influence of the Intertropical Convergence Zone (ITCZ). This global belt is characterized by the presence of low-pressure systems, convection and heavy precipitation. Due to the Hadley cell circulation, descending dry air is present at around 30 degrees latitude distance from the ITCZ. Because of the absence of convection, these high-pressure systems experience almost no precipitation. The location of the ITCZ varies throughout the year as can be seen in Figure 4.1, being the most North in July and the most South in January. Hence, the regions near the equator undergo distinct annual wet and dry seasons. For the results found in the global analysis this means that the dry high-pressure systems are located at the areas in which the greatest contrasts among the LSMs were observed. For the JJA months this dry belt lies just below the equator while it is located just above the equator for the DJF months. In the annual rainy season(s) these regions receive significant precipitation. Therefore, the available soil moisture for evaporation, as well as the parameterization of water stress during these dry months seem to play a major role in these differences between LSMs. For these regions in the summer months, the energy is high, the humidity is low and the precipitation is almost negligible. This makes the potential evaporation very high,

pushing hard on the soil and leaves to release water. This evaporation will mainly depend on the amount of water still available in the soil (E_S), open water evaporation (E_O) and the amount of water from the rootzone that can be absorbed by the plant roots and evaporates by transpiration (E_T). Interception (E_I) is expected to be negligible during these dry periods in time.

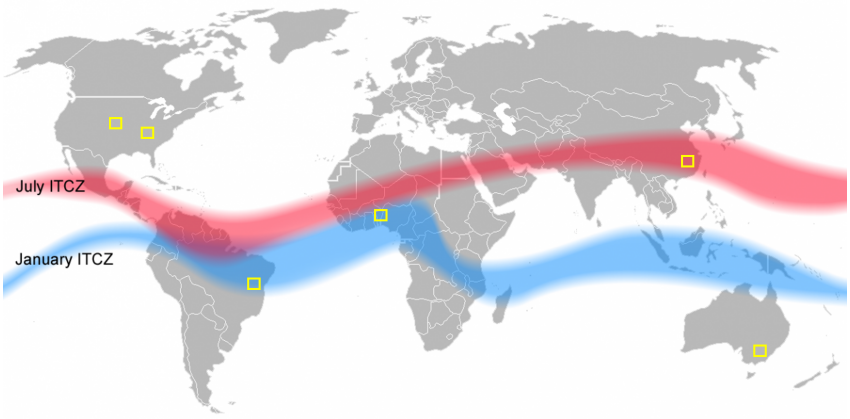


Figure 4.1: *Intertropical Convergence Zone for the most Northern (July) and most Southern (January) positions [Halldin, 2006]. The remaining months, the ITCZ is located someplace in between these two belts. The yellow squares represent the investigated study areas.*

Part 2

Budyko Framework

The Budyko Framework showed that the distribution among the LSMs, in terms of the evaporation index, remained roughly the same for both wetter and drier areas. This suggests that the LSM performance is not sensitive for the dryness of a region. Furthermore, the relative errors were in line with the conclusions drawn in part 1 of this section as the four LSMs with the largest error also deviated the most from the reference products for the global seasonality.

The exceeding of the water limit, by two of the LSMs, suggests that water from another source needs to be added to close the water balance. From the water balance analysis it can be concluded that the total run-off is nearly equal to zero for these LSMs when the water limit is exceeded. This way, the LSM ensures that the water limit is not exceeded even further. Since it is known that the LSMs incorporated in the climate models do not use lateral connectivity or have any interaction with groundwater, it is assumed that exceeding the water limit is accompanied by a (small) decrease in soil moisture storage over the investigated time series [Kim and Mohanty, 2016]. From the water balance it can be concluded there are several LSMs that exceed the water limit when adding the run-off to the evaporation. This leads to an almost equal negative water balance at locations A and C compared to M6 and EE. However, exceeding the water limit by the evaporation flux alone combined with a negligible run-off makes the LSM output of M6 at location A and EE at location C rather questionable. The Budyko Framework also showed that reference data FB significantly exceeds the water limit at location A. However, the cause of this is less clear as the precipitation data representing the water limit was not used as input for FB, which was the case for the LSMs. As a result, certain biases in the precipitation data may be carried over into this product. In addition, FB, which had observational data as input, may take the role of other water sources into account, such as irrigation water and groundwater inflow. Wang-Erlandsson et al. (2016), in their paper, discuss the exceedance of evaporation over precipitation for various satellite based E and P products. For location A, the comparison of E_{CSM} to P_{CHIRPS} showed a similar exceedance of the water limit as for FB compared to P_{GSWP3}. They found that this exceedance occurs primarily during the drier years (Figure 4.2), suggesting that the water demand in this case is satisfied by soil with higher potential or by groundwater uptake from vegetation [Wang-Erlandsson et al., 2016].

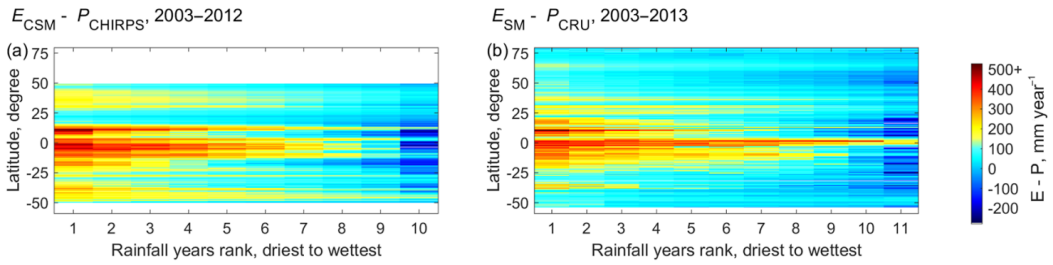


Figure 4.2: *Mean latitudinal difference between various satellite based E and P products, sorted from the driest to the wettest years. The figure from Wang-Erlandsson et al. (2016) only includes regions where accumulated E minus P over the entire available time series (2003–2012 and 2003–2013, respectively) are positive [Wang-Erlandsson et al., 2016]*

Water balance

Considering the issues with closing the water balance of HG and UK, the problem seems to lie in a numerical error within the LSM's calculations. The simulated evaporation is relatively low and the calculated run-off does not compensate for this phenomenon. However, there seems to be no problem with the conservation of mass principle, as storage increases in this case. As can be seen in Figure E.4, the total soil moisture remains roughly stable for HG and UK for all study areas. Therefore, it is not apparent where the remaining water is directed. However, it could also be a deliberate choice by the modelling groups not to close the water balance and ensure that the model might perform better at some other point as a result. CE shows an inverse pattern compared to UK and HG which therefore seems to have a different cause. Figure E.2 in appendix E clearly shows that CE breaches the conservation of mass principle by decreasing by thousands of millimetres over 10 years for locations E and F. As Figure E.4 demonstrates, the total soil moisture output for CE shows no decline or increase over the investigated decade for all LSMs. Therefore, making it unclear where the additional water is obtained from.

Part 3

Water fluxes for the driest period in time

The SMDs, change in soil moisture and total run-off from the LSMs, for the driest period in time, can be considered as the extreme values with a return period of 10 years. For the SMDs, it was concluded that there was a significant spread in magnitude among the LSMs. Although this was in line with the results that had been found in the Budyko framework, some differences were observed. EE increased in performance compared to the results discussed in part 1 and 2 while it seemed to be the opposite for IP. EE showed that it was extremely sensitive to precipitation anomalies compared to the other LSMs. Although this was found for the anomalies, this is entirely in line with the fact that EE displayed a high evaporation rate during the wetter months and a more average evaporation rate during the drier months (Figure B.3 in Appendix B). M6, which also showed a relatively high annual evaporation rate, had a much lower sensitivity in terms of precipitation anomalies and was the least performing LSM in terms of the SMDs. This high sensitivity of EE, and therefore being rather water limited, seemed to compensate for the high evaporation rate and makes the LSM perform relatively well in simulating SMDs. The overestimation of soil moisture for the first half a meter of soil could be caused by a poor representation of vertical fluxes and redistribution of water in the soil. The results found for the soil moisture did not seem to show the same patterns as found for the computed SMDs. Table D.2 in Appendix D shows the differences between the SMDs and the total decrease in soil moisture for every LSM per location. If the SMD exceeds the value of the total decrease in soil moisture, some kind of lateral inflow is expected. The opposite holds for the situation when the total decrease in soil moisture exceeds the SMD value. In this case lateral outflow through run-off is expected. This way of thinking is illustrated in Figure 4.3 and indicates that some LSMs do not follow this concept as the run-off is either close to or above zero and there is no lateral inflow in the models. For several study areas, M6 shows errors in following this concept as lateral inflow is required to close the water balance as the soil moisture decrease is less than the SMD.

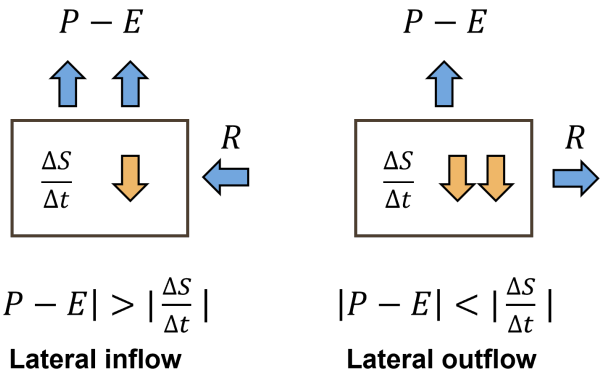


Figure 4.3: *Schematization of what certain mean $P-E$ and $\frac{\Delta S}{\Delta t}$ fluxes would mean for the lateral flow. This schematizion is linked to Table D.2 in appendix D.*

A strong correlation was observed between the total run-off during the driest period in time and the determined SMDs. Generally, LSMs with higher SMD values had a lower run-off value and vice versa. However, the LSMs that displayed water balance issues appeared to be exceptions to this trend. For example, UK and HG had the lowest SMD and run-off values for location D, and the relatively high run-off values of CE for locations E and F did not match with a relatively low SMD value for the same locations. Although not unexpected, it is relevant for this research that the overall water balance issues of certain LSMs were also reflected for the water fluxes during the driest period in time.

Correlation between SMDs and reference soil moisture

For locations A and B, which are the water limited locations in the tropics, the least performing LSMs appeared to overestimate the SMD. This by showing a fairly linear response over time. It seemed that these least performing LSMs do not experience much water stress as the dry period progresses. The reason for this could be that the LSM has too much water available or that too much energy goes into evaporation. It is expected that besides the decrease in soil moisture, and thus increased water stress, vegetation is able to adapt to the drier conditions in water controlled regions by limiting the transpiration [Laio et al., 2001]. This would in this case mean that the slope of soil moisture decline decreases as the dry period progresses. The reference soil moisture meets this line of thought for locations A and B. In contrast to locations C, D, E and F, none of the LSMs seem to underestimate the timing and progression of the SMD over time for these locations. Therefore it appears, that under these conditions of high seasonality on precipitation and consistent high temperatures (B.1a and B.1b), there are no LSMs that overestimate the water stress over the dy period in time for these locations.

For locations C and D, which are water limited locations in the extratropics, the least performing LSMs appeared to underestimate the SMD over time. Notably, HG and UK almost showed no increase in SMD, while the reference SM data persistently decreased over time. Locations E and F, which are energy limited locations, showed a similar response compared to locations C and D. The low correlation LSMs appeared to underestimate the magnitude of the SMDs. They responded with a certain delay to the decrease in soil moisture, keeping the SMD at zero for the first few months of the dry period. Although this was observed for all LSMs, the low correlation LSMs showed the slowest response. Since this research distinguished between energy- and water-limited locations, it was expected that there might be differences in how the LSMs would respond over a dry period. However, the results of C and D were fairly similar to those of E and F. There was a similar performance among the LSMs and the lowest correlation equaled the relatively low SMD values. In the Budyko Framework, a distinction between water and energy limited locations was made. However, this was determined for an annual average. For locations C, D, E and F, the clear distinction between water and energy limited locations no longer applied, as it seems likely that an energy limited location could behave like a water limited location during droughts. This due to the fact that during droughts the amount of available water is more likely to become a limiting factor instead of the available energy for evaporation. This may explain the similarities between the results of locations C, D, E and F.

Difference in correlation between the tropics and extratropics

The main difference between the tropical (A and B) and extratropical locations (C, D, E and F) is expected to be caused by the differences in temperature and precipitation seasonality as can be seen in Appendix B. A combination of high energy levels with practically no precipitation within the dry periods (locations A and B) leads to high potential evaporation values and a relatively low humidity. This creates ideal conditions for high evaporation rates with water as limiting factor. Therefore, to simulate the actual evaporation it appears to be mainly important to accurately parameterize the soil moisture dynamics and water stress functions of the soil and vegetation. Furthermore, it is expected that the initial soil moisture storage in the rootzone, available for soil evaporation and transpiration, at the start of a dry period, plays a dominant role in this. For locations outside the tropics (C, D, E and F), lower fluctuating energy levels apply and precipitation is more present during the drier months. This also leads to higher humidity levels compared to locations A and B. As a result, the evaporation rate depends more on the fallen precipitation and how the LSM transforms this precipitation into evaporation, change in soil moisture and run-off. Therefore, soil moisture dynamics, initial soil moisture storage as well as the water stress functions are expected to be of less importance. Seneviratne et. al (2010) described this phenomenon and showed how the dry to wet transitional zones with a certain soil wetness are most affected by the $SM - E$ coupling (Figure 4.4). A low value for the soil wetness leads to an insignificant evaporation rate while a high value for the soil wetness would lead to a low $SM - E$ coupling. The highest impact of this coupling is therefore reached for the range in soil wetness, displayed with the grey band. In the same range, the standard deviation of E is at its peak. For our research, it is assumed that locations A and B are in this range of soil wetness and are therefore the most impacted by the $SM - E$ coupling.

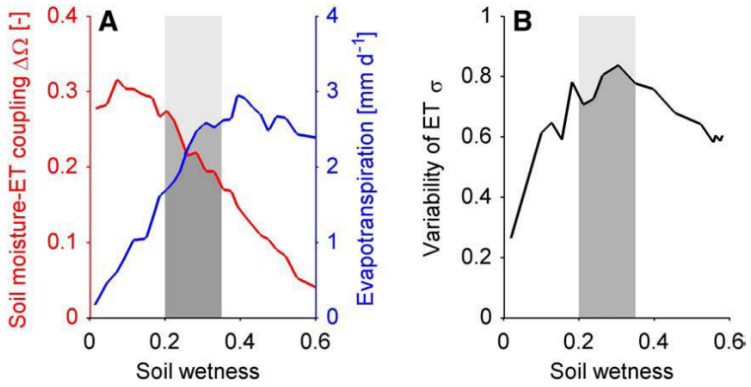


Figure 4.4: *The coupling between soil moisture and ET, which is equal to E in this research, over varying soil wetness in GLACE simulations (A) and the standard deviation of ET over varying soil wetness for the same simulations (B) [Seneviratne et al., 2010]*

Model sensitivty

EE stood out in the model sensitivity analysis for being the most sensitive for all study areas, while UK was the most insensitive for four out of the six locations. Furthermore, it was determined that in general the LSMs were overly sensitive compared to the reference data sets. The $P - E$ anomaly response can be seen as a measure of how water limited the LSMs are in terms of evaporation. For our research, the negative precipitation anomalies were the most relevant as they represent periods in time with less water available compared to normal conditions (i. e. droughts). The fact that almost all LSMs seem to overestimate the $P - E$ anomaly sensitivity means that the evaporation rate will be exaggerated in periods with below average precipitation. The results, at the same time, showed that the LSMs are overly sensitive for above average precipitation. A cause for the LSMs being overly sensitive could be the inaccurate LSM representation of vegetation changes under changing precipitation, like the representation of the stomatal resistance (r_s) and Leaf Area Index (LAI) [Zeng et al., 2018]. Another potential cause is an inaccurate representation of the soil moisture-evaporation coupling under changing precipitation [Seneviratne et al., 2010].

The $P - R$ anomaly response generally showed the opposite pattern to the $P - E$ anomaly response. This was expected

since the change in precipitation is partly split among evaporation and total run-off. From the $P-SM$ anomaly response, it was concluded that CE, C2 and IP make the most use of the soil moisture under these conditions, especially for the driest locations C and D. These LSMs also showed in general the highest cumulative response. Among the locations, location A stood out as it had a relatively low cumulative response. A combined slope of approximately 0.3 means that only 30 percent of the change in precipitation is reflected in the changes of E, R and SM. It is not entirely clear why this was the case. For the remaining locations, a more unanimous overall slope was observed.

Overall interpretation

For all findings discussed in Parts 1,2 and 3, the LSMs showed a significant spread in the results. In addition to the initial evaporation bias of each LSM, the main contrasts were observed in the dry to wet transition zones within the tropics. This is expected to be due to the strong soil moisture-climate coupling caused by the specific climate conditions in these regions. Differences in modeling soil moisture dynamics and water stress functions of soil and vegetation are expected to play a major role in this. The challenges in representing these processes could underlie the fact that the models were generally too sensitive to precipitation anomalies. Furthermore, the main differences for the tropical and extratropical locations are expected to be caused by differences between the strong soil moisture-climate coupling in the tropics and the more precipitation dominant and radiation limited evaporation regimes in the extratropics. This was especially evident for the correlation analysis with reference soil moisture. The low correlation LSMs seem to overestimate the SMDs in the tropics and underestimate them in the extratropics.

Considering the types of Land Surface Models that were implemented in the Earth System Models, the considered models in our research can be grouped. Although not the same version, CE (version 4.5) and C2 (version 5.0) used the Community Land Model (CLM) as the basis for their land component ([Lovato et al., 2022]; [Danabasoglu et al., 2020]). E3 used the CLM (version 4.5) as a basis, made some adjustments, and gave it the name E3SM Land Model (ELM) [Golaz et al., 2019]. EE used HTESSEL [Döscher et al., 2021], HG and UK, both developed by the Met Office Hadley Centre, used JULES as the basis for their Land Surface Model [Best et al., 2011], IP used ORCHIDEE (version 2.0) [Boucher et al., 2020] and M6 used MATSIRO [Qiang et al., 2021]. It is not the case that each Earth System Model gives the same output with the same LSM basis, as each model makes its own small adjustments for the implementation. However, in our research, the similarities within these groups of models were evident. As can be seen in Figure 4.5, ranking them on their performance during droughts, it became clear in our research that the CLM outperforms all other LSM bases. Specifically, the modified version implemented in E3. Overall ORCHIDEE was second best and HTESSEL third. The worst performing LSM bases were MATSIRO (M6) and JULES (HG and UK).

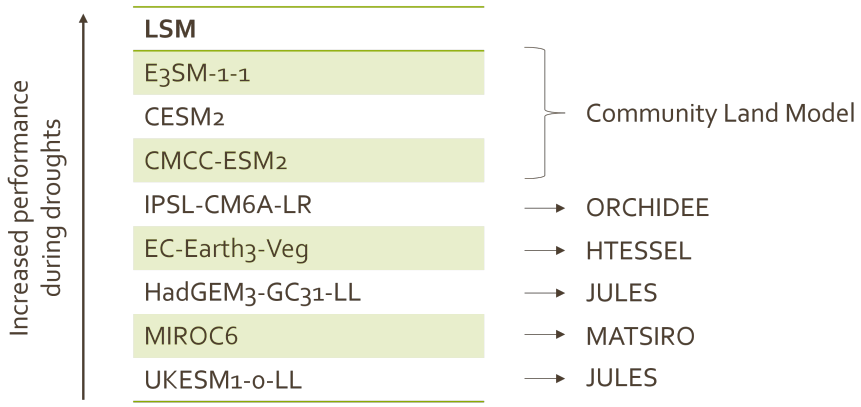


Figure 4.5: *The LSM ranking in terms of simulating droughts as well as the model bases implemented in them. The ranking is based on the results displayed in Table 3.1.*

Hydrological perspective

From a hydrological perspective, certain limitations of LSMs can be linked to the findings of our research, as LSMs are known to lack some essential aspects to accurately represent the water cycle ([Clark et al., 2015]; [Fisher and Koven, 2020]).

One of the major limitations is the neglect of groundwater in the models, ignoring discharge and recharge processes which are a major aspect of the hydrological cycle. Furthermore, LSMs neglect the lateral flows between grid cells, which results in an inaccurate representation of surface and subsurface water flows [Kim and Mohanty, 2016]. Research has also shown that the influence of lateral connectivity on soil evaporation and transpiration is more significant during dry periods in time due to stronger soil moisture-evaporation coupling [Ji et al., 2017]. For our research this could be one of the reasons that some LSMs had issues with closing the water balance. However, with the limitations mentioned above, you would expect some minor deviations. This was not the case as the storage increased or decreased by thousands of millimeters over the years for some LSMs. Because of this, you could argue that this was a deliberate choice by some of the LSMs to serve another purpose. Perhaps to improve the model performance in some other aspect.

Within the LSMs, the water balance is simulated for each grid cell separately, and the residual water in the vertical column is attributed to run-off, ignoring the lateral movement of water between neighboring cells (see Figure 4.6). In reality this is a very simplistic approach and the run-off is influenced by many other factors within the specific grid cell, such as irrigation practices and land-use. This method shows that, within the LSMs, the least focus is on the run-off and that it mainly functions as a water flux to close the water balance. This is in contrast to the operation of many hydrological models in which the run-off is often used as a starting point for the calibration. A link can be made between these limitations and our research, since the results showed a significant spread in the run-off output between the LSMs for the water balance and during the driest periods in time. This suggests that certain uncertainties and biases seem to work through in the final run-off flux.

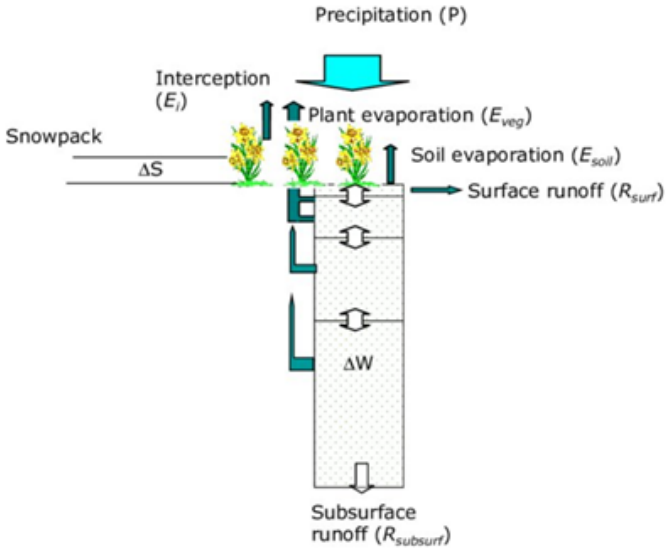


Figure 4.6: *Water balance of the HTESSEL land surface scheme implemented in EC-Earth3-Veg.*

Another limitation of the LSMs is the inability to capture the heterogeneity of the, often, very coarse resolution. One grid cell can vary greatly in, for example, elevation, land-use, soil properties and rootzone storage capacitiy. Appendix C clearly shows the spatial variability of soil moisture for all study areas, implying that one average value would be a major oversimplification. Even though this is a well known limitation, its consequences on the results of our research are difficult to identify.

The same applies to the vertical heterogeneity of the soil, as it is a known limitation but hard to link to the results of this study. Within the LSMs, as can be seen in Figure 4.6, the entire soil layer is simplified into a few layers that interact with

each other. This neglects the vertical heterogeneity of the soil and the presence of, for example, preferential flow paths. Each LSM has its own schematisation and parameterisations of the soil moisture layers and water fluxes. Furthermore, as can be seen in Table D.1, there are major variations in soil layer depth used by the LSMs for the simulations. Figure E.4 shows that is also the case for the magnitude of total soil moisture. M6 has a far greater soil moisture storage, which can be linked to the fact that M6 has, in general, a relatively high evaporation rate throughout the year. Even during drier periods, this evaporation rate hardly decreases, which is somewhat contradictory from a hydrological perspective in water limited regions. Overall, it is expected that these limitations and variations among the LSMs attribute to the large spread in performances. Another challenge in terms of the representation of the soil, is accurately representing the soil evaporation stress, which is a key contributor to the SM-E coupling and is often oversimplified in LSMs [Dong et al., 2020]. As mentioned before, in our findings this can be linked to the differences between the results found in the tropical and extratropical study areas. The correlation results for study areas A and B suggest that the worst performing LSMs experience too little water stress as the dry period progresses, in these areas with strong $SM-E$ coupling. The mentioned limitations and oversimplifications of the soil could be one of the causes for this, as it could become more challenging to adjust water stress under changing drought conditions. For the extratropical study areas, water stress functions appear to be less important, as evaporation is more interception dominated.

Similar as for the representation of the soil, it is known that vegetation phenology is often oversimplified or misunderstood within the LSMs ([Lian et al., 2020]; [L. Harris et al., 2023]). During droughts, vegetation is able to adapt to the below average amount of water and the interannual variations, which subsequently affects the magnitude of evaporation [Zeng et al., 2018]. For our results, it is expected that this could be related to the fact that the LSMs were found to be overly sensitive to precipitation anomalies in terms of their evaporation output. Together with the aforementioned limitations and oversimplifications of the soil, it is likely that the LSMs have difficulties dealing with precipitation anomalies. Furthermore, the challenges in the representation of vegetation phenology can also be linked to differences in the correlation results between tropical and extratropical study areas. Namely, that the limitations and oversimplifications of the vegetation create the challenge of accurately adjusting the vegetation water stress under changing drought conditions. As a result, the worst performing LSMs in the tropics show a fairly constant evaporation rate over the drought duration. Just as for the soil evaporation stress, for the extratropical study areas, vegetation water stress seems less important as the evaporation is more interception dominated.

Lastly, a major challenge of the LSMs is to correctly implement the role of human impacts [Al-Yaari et al., 2021]. It is well known that humans have a significant impact on the water cycle through climate change, land use changes, the building of dams, irrigation and the extraction of groundwater, creating besides spatial also temporal variability [Clark et al., 2015]. Appendix A shows that a large part of the world as well as some locations considered in our research can be classified as cropland, which implies that biases due to irrigation could be present in the simulation of evaporation. In this case, this would equal predominantly negative biases relative to the reference data. The global mean analysis showed that the LSM global mean is slightly below the reference data. However, it would be questionable to link this only to the role of irrigation. For the study areas, it can be argued that the locations where the water limit was exceeded in the Budyko Framework may have been influenced by the role of irrigation. This led to negligible runoff for some LSMs to close the water balance. Irrigation could be the missing link in this to improve LSM performance. Apart from this, the influence of irrigation or other human impacts is not evident in the results of this research.

4.2 Comparison to previous research

Several researches have been conducted focusing on the terrestrial evaporation in CMIP6 Earth System Models, focusing for example on global annual ET averages [Wang et al., 2020], regional temperature biases due to evaporation [Dong et al., 2020] and the simulation of land surface energy and water fluxes [Li et al., 2021]. Furthermore, several researches have been conducted evaluating the performances of Land Surface Models in particular, which function as the basis for the land part within the Earth System Models used in CMIP6. For instance, studies have been done on the LSM performance during evaporative droughts [Ukkola et al., 2016] and the performance of JULES (used in HG and UK) in simulating baseflow [Zulkafli et al., 2013].

CMIP6 models assessment

Wang et al. (2020) evaluated the state-of-the-art climate models included in CMIP6 comparing the evaporation output to the GLEAM data set. This reference data was also part of DOLCE V3 which was one of the two reference data sets used in our research. One of the main results from this research was that almost all CMIP6 models overestimated the global annual evaporation compared to GLEAM between 1984 and 2010. This is in conflict with the results found in our research. Figure 3.1 showed that on average the models do not over or underestimate the evaporation. However, our research examined a different time scale, reference data and a different selection of models. Looking at the spatial characteristics of the GLEAM data set minus the CMIP6 ensemble mean (Figure 4.7) similarities can be seen with Figures 3.2 and 3.3. Just like the difference between reference data DO and the LSM mean, there was mainly a positive bias observed over the regions with a tropical rainforest climate. Similarities for the negative bias were less clear.

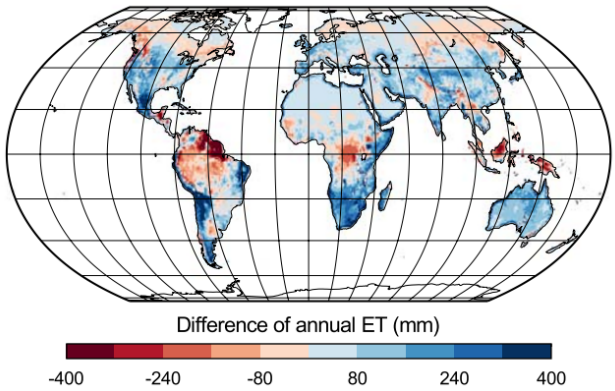


Figure 4.7: *Biases in the multi- model mean with respect to the GLEAM data [Wang et al., 2020]*

According to Dong et al. (2020), the CMIP6 climate models' evaporation in the Central United States is too sensitive to seasonal changes in precipitation. They mention that this may be because the models cannot accurately account for the complex relationships between vegetation, root zone storage, and climate in transitional areas from dry to wet. During the dry summer months, evaporation relies more on deep soil moisture that accumulated during spring. However, CMIP6 models struggle to capture this process and instead focus more on precipitation-related processes like interception and surface soil moisture. The results of this research were completely in line with our findings and conclusions. Namely, the models appeared to be overly sensitive to precipitation anomalies (Figure 3.18), the areas where dry conditions transition to wet conditions exhibited the most significant deviations, and the overall conclusions regarding the underlying causes were similar. Furthermore, Dong et al. concluded that the ensemble of LSMs outperforms all seperate LSMs. This is an interesting finding what was not investigated in depth in our research. However, based on Figures 3.1 and 3.4b it can be concluded that the LSM ensemble was close to the reference data in terms of the annual evaporation rate.

LSMs assessment

Ukkola et al. (2016) conducted a research on the performance of eight different LSMs in simulating evaporative droughts. The research concluded that most LSMs tended to overestimate the intensity, duration, and magnitude of evaporative droughts, failing to accurately capture realistic drought responses. Although our research concluded the same about failing to accurately capture drought responses, it was not evident that it was mainly due to overestimating. From Section 3.2.4 it was concluded that this appeared to be the case for the worst performing LSMs at locations A and B while it was the other way around for locations C, D, E and F. Therefore, this was not entirely in line with what Ukkola et al. found.

Ukkola et al. identified The Joint UK Land Environment Simulator (JULES) as an LSM that enters hydrological droughts too slowly while overestimating ensuing drought duration and severity. This was found to be the case for the results of our research in Section 3.2.4, especially for locations A and B, where HG and UK showed a relatively linear response over

time, overestimating ensuing drought duration and magnitude. For locations C, D, E, F, HG and UK were among the worst performing LSMs as they appeared to be unresponsive for the first months of the dry period in time. Furthermore, research has shown that JULES and ORCHIDEE, among several other LSMs, were oversensitive to short-term precipitation variability, which is consistent with the results found in Section 3.18 [Tallaksen and Stahl, 2014].

Zulkafli et al. (2013) concluded that the JULES LSM can be unreliable for hydrological predictions due to the poor simulation of the base flow and by errors in the water balance. They also found that the LSM was unable to represent the high evaporation rates for the study area. Both results are in line with the results found in our research. HG and UK were not able close the water balance for all locations (Figure 3.7) and showed a relatively low evaporation rate (Figures 3.2 and 3.3).

Overall, our research is in general in line with the research already done on both the CMIP6 models and the LSMs that are implemented in them. However, our research has shown that there can be differences in terms of how a model performs on an annual basis and how it performs during droughts. This implies that not every study investigating the evaporation of LSMs can be easily compared.

4.3 Limitations

One of the main limitations is that this research focused only on the performance of eight different climate models from CMIP6, while there are many more models involved in the comparison project. Therefore, the findings of this research may not be generalisable to all climate models of CMIP6, but are more relevant for concluding which LSM basis performs relatively well. However, it is a possibility to conclude on certain patterns and characteristics and use these to make predictions for the remaining CMIP6 models.

Second, a big assumption made in this research is that the shallow soil moisture and the computed soil moisture deficits should be perfectly correlated (Section 3.2.4). Although it is evident that they should be correlated during dry periods in time, the response is not necessarily equal. The reference soil moisture still interacts with the deeper soil in terms of infiltration and recharge. The soil moisture deficit was calculated by considering the soil as a bucket that only decreases or increases in volume through evaporation and precipitation. Therefore, the soil moisture deficit is a simplification of what happens in reality which means that some differences compared to the reference soil moisture data are expected.

Third, like the LSMs, the reference data also contains some uncertainties and biases. DO and FB do not entirely match in their evaporation values and spatial characteristics. Figure 3.8 clearly shows these differences between the two in terms of the soil moisture deficits. The selection of a specific reference data set, therefore, may lead to different final results and conclusions. Additionally, the reference soil moisture data set (SO) is also expected to have certain biases. The measurement points are poorly distributed globally, after which a machine learning algorithm attempts to interpolate between them. Furthermore, point data does not capture the heterogeneity of the soil in a certain area. For instance, these data is potentially obtained from easily passable places with a less than average number of tree roots in the soil. To minimize the biases in this research due to measurement uncertainties and due to the interpolation algorithm, it was decided to investigate only areas with at least two measurement points in them.

Another limitation for this research was the limited knowledge of the operation of the considered LSMs. Since each LSM has its own structure and parameterisations in terms of land-atmosphere processes, it was decided not to address this for all specific LSMs. Therefore, the focus was on the meteorological inputs and subsequent model outputs, followed by conclusions on the operation, biases and uncertainties of the LSMs, based on previous research and from a hydrological perspective. However, a more in-depth knowledge of the parameterisation and components of different LSMs might have strengthened the final conclusions of this research.

Additionally, this research has some limitations in terms of how it investigated the performance during droughts. The periods examined were limited to a particular range of months, based on the most significant decreases in soil moisture

determined within the selected time series. However, the LSMs exhibited variations in the duration of droughts beyond this specific range of months. Unfortunately, these variations were not considered in this study. Figure 4.8 illustrates the variability in duration among the LSMs of annual droughts, using location C as an example. It's important to note that choosing a fixed period may lead to an LSM performing unexpectedly well, while in fact it extends the dry period for a few months. Moreover, this research did not investigate multi-year droughts, as well as the recovery process after droughts, while both are important aspects of how the LSM deals with dry periods in time. Figure 4.8 also highlights the differences among the LSMs in this regard, emphasizing the importance of investigating these aspects of droughts as well.

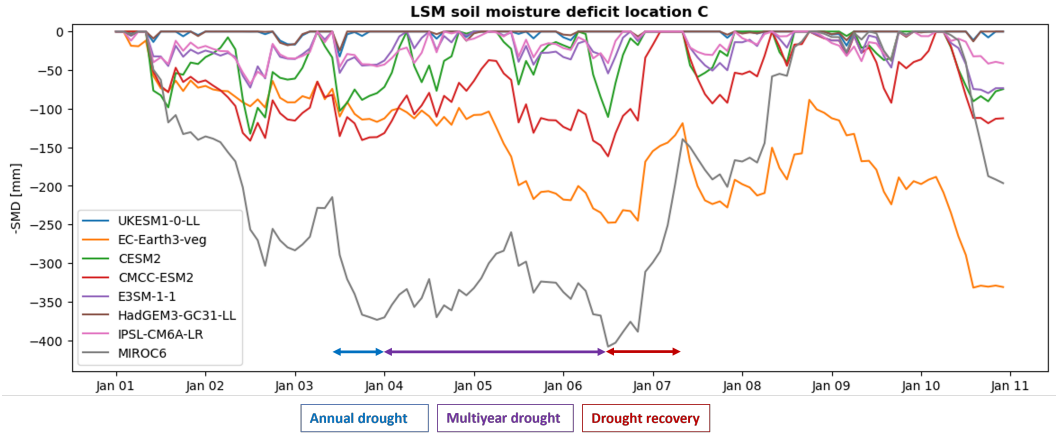


Figure 4.8: *The soil moisture deficits for all LSMs, for location C: Central US, over the considered time series.*

Another limitation of this research is the way the overall performance ranking was determined (see Table 3.1). By using pluses and minuses, all performances are relative to each other and do not say much about how good they really are. In addition, it gives the impression that these scores add up when in fact the columns cannot necessarily be called independent. For instance, the correlation and absolute value of soil moisture deficits are linked. This could create a skewed picture between the performances of the different LSMs. This table mainly shows what the best and worst LSMs are per component and it is somewhat uncertain when viewed as a combination of performances.

Lastly, during this research, it became clear that the EC-Earth3-Veg simulation had an error in its representation of total soil moisture. Every 20 to 40 years, the initial conditions, in terms of total soil moisture, reset. This meant, in this case, a sudden rise in total soil moisture with a gradual decrease afterwards. Such a rise occured within the investigated time series of this research at the start of 2005. This peak in soil moisture also affected the simulation of run-off and evaporation as they are related. Figure 4.9 shows this reset in total soil moisture and the effect this had on the global mean evaporation magnitude of EE. It can be observed that in 2005 the highest value of global mean evaporation was reached and it is expected that it also affected the years up to 2010 in terms of the magnitude of the evaporation. This finding makes it difficult to draw a conclusion on the performance of EC-Earth3-Veg in simulating droughts. It is unclear how the results would have been different if this error had not occurred.

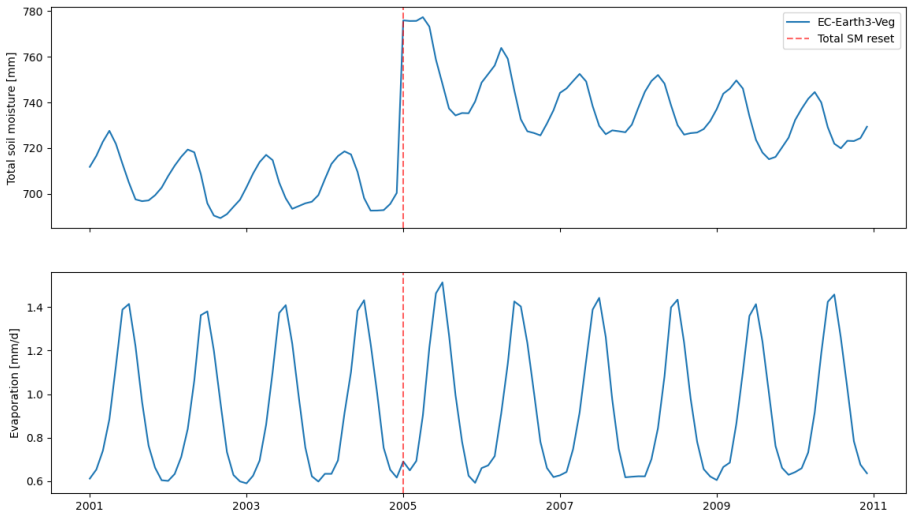


Figure 4.9: *Global mean daily evaporation and global mean total soil moisture by EC-Earth3-Veg between 2001 and 2010. The red dashed line represents the moment in time in which the total soil moisture resets.*

In general, most of the limitations of this research offer opportunities for follow-up research with the aim of ultimately strengthening the determined conclusions drawn in this research. More knowledge of individual models and the focus on other aspects of droughts will help explain patterns and performances and will create a clearer picture overall. Furthermore, adding more CMIP6 models will also strengthen this research.

4.4 Implications

This research evaluated the performance of eight different LSMs in the simulation of soil moisture droughts. In addition to the focus on droughts, this research investigated the annual water fluxes involved in the water balance. Overall, this has provided a conclusion on the characteristics and performances of the LSMs.

This research can contribute to the understanding of LSMs used in ESMs in terms of simulating droughts, as it identifies the uncertainties and biases in them. One strength of this research is that it considers various water fluxes and is not focused solely on the evaporation output. As a result, conclusions can be drawn about the water balance within the LSMs and why this leads to particular deviations. For each LSM, the characteristics for long-term analyses and the distribution between evaporation and runoff were determined in various climates. By conducting the same analysis during droughts, it becomes clear whether the model maintains the same level of performance. The fact that this research showed that the LSMs are, in general, overly sensitive to precipitation anomalies is a relevant finding, looking ahead in time. Climate change is expected to lead to more extreme droughts and intensified precipitation. Thus, the results imply that the current LSMs would not be able to accurately simulate these extremes that deviate from the average seasonal pattern.

In addition, the results show the strengths and weaknesses of the LSMs for different parts of the analysis. This creates an opportunity to delve into the model components and figure out why a certain model performs better in a specific aspect. This aligns well with the objective of LS3MIP, which is to improve LSMs through experiments and comparisons. The results of this study also demonstrate the need for intercomparison studies such as LS3MIP, as the variation between LSMs is significant for all parts of the analysis. This suggests that the focus should be more on improving the existing LSM components and perhaps less on introducing new aspects.

The ranking of LSMs provided in this research show which models perform better than others for certain parts of the analysis. This creates the possibility of selecting a specific LSM for a particular task. The LSM ranking also leads to a ranking of the model bases implemented in them. This allows us to predict what the performance of other LSMs, not considered in this research and with the same model base, will be.

Furthermore, this research can contribute to the overall understanding of droughts in the climate modeling community as it highlights certain hydrological aspects that are important and not yet well incorporated within the LSMs. These aspects were linked to the findings and served as hypotheses as to why certain deviations and differences were found. LSMs are known to be quite oversimplified compared to hydrological models used by hydrologists. This research can contribute to bridging these two different fields of research.

In general, the identification of strengths and weaknesses as well as the rankings determined in this research could aid in the development and improvement of LSMs. Improved simulations can provide better predictions of the future effects of climate change, allowing vulnerable drought-prone areas to better prepare and increase overall climate adaptation and mitigation. Water scarcity and food insecurity are phenomena strongly linked to droughts and will hopefully be limited by improved simulations and predictions, contributing to Sustainable Development Goals 2 and 6 (zero hunger and clean water and sanitation). Therefore, the implications of this research extend beyond assisting in improving drought simulations and into the larger context of climate change and its effects on vulnerable populations. By improving our understanding and modeling capabilities of the mechanisms that lead to droughts, we can better prepare ourselves for the future and hopefully mitigate the impacts of climate change.

5 — Conclusion

This research evaluated eight Land Surface Models (LSMs) included in the Land Surface, Snow and Soil Moisture Intercomparison project (LS3MIP) on their performance in simulating soil moisture droughts. From the intercomparison project, the Land-Hist offline simulations were considered. After a global analysis on the LSM evaporation characteristics, six study areas with diverse climate characteristics were selected. For these locations, the LSM performances in terms of the water balance, Budyko Framework, simulation of soil moisture deficits, and the model sensitivity to precipitation anomalies were evaluated, with a particular focus on dry periods in time.

All parts of the evaluation displayed a wide spread in LSM characteristics and performances. Some LSMs revealed issues in closing the water balance or exceeded the water limit in the Budyko Framework. On a global scale, the greatest contrasts among the LSMs were observed at dry to wet transitional zones within the tropical regions, especially during the driest months. In this latitudinal range, the worst performing models generally overestimated the accumulation of soil moisture deficits and exaggerated the drought severity. For the extratropical locations, the worst performing LSMs displayed an opposite pattern and underestimated the severity of droughts. Additionally, all areas studied revealed excessive sensitivity of evaporation to precipitation anomalies. From a hydrological perspective, the findings of this research could be linked to some known limitations of the LSMs. Neglecting groundwater and lateral connectivity may be related to issues with the water balance, while a poor representation of the soil moisture - evaporation coupling could be linked to the differences in findings between the tropical and extratropical regions. Additionally, oversimplified soil and vegetation dynamics could contribute to the LSMs being overly sensitive to precipitation anomalies.

Despite the significant variations, several LSMs exhibited relatively strong performances across all conducted analyses. Ranking the five different model bases, implemented in the eight evaluated LSMs, on their performance during droughts, it became clear that the Community Land Model (implemented in E3SM-1-1, CESM2 and CMCC-ESM2) was predominantly the best performing. This was followed by ORCHIDEE (IPSL-CM6A-LR) and HTESSEL (EC-Earth3-Veg). MATSIRO (MIROC6) and JULES (HadGEM3-GC31-LL and UKESM1-0-LL) were the least performing LSM bases.

The findings of this research could have significant implications for LS3MIP and the Land Surface Modeling community, as they highlight the strengths and weaknesses of LSMs in simulating soil moisture droughts. The wide variability in the performance of the models underlines the importance of intercomparison studies and the improvement of existing LSM components, rather than solely focusing on the addition of new components. Ultimately, this research could contribute to the development and enhancement of LSMs, improving our understanding and modeling capabilities of drought mechanisms. From there, the implications extend beyond model improvement and into the larger context of climate change predictions, mitigation, and adaptation, with a specific focus on drought-prone regions.

6 — Code availability

All analyses were done in the python programming language and processed in jupyter notebooks. All the notebooks are available on GitHub at the following url:
`https://github.com/PdenBlaauwen/The-performance-of-Earth-System-Models-in-simulating-droughts`
The notebooks are also accessible through the following DOI: 10.5281/zenodo.7950674.

Bibliography

[Al-Yaari et al., 2021] Al-Yaari, A., Ducharne, A., Tafasca, S., Mizuochi, H., and Cheruy, F. (2021). Infllunce of irrigation on the bias between orchidee and fluxcom evapotranspiration products. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*.

[Andrews et al., 2020] Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., Burke, E., Dittus, A. J., Florek, P., Gray, L. J., and et al. (2020). Historical simulations with hadgem3-gc3.1 for cmip6. *Journal of Advances in Modeling Earth Systems*, 12(6).

[Best et al., 2011] Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., and et al. (2011). The joint uk land environment simulator (jules), model description – part 1: Energy and water fluxes. *Geoscientific Model Development*, 4(3):677–699.

[Bogenschutz et al., 2020] Bogenschutz, P. A., Tang, S., Caldwell, P. M., Xie, S., Lin, W., and Chen, Y.-S. (2020). The e3sm version 1 single-column model. *Geoscientific Model Development*, 13(9):4443–4458.

[Boucher et al., 2020] Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., and et al. (2020). Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, 12(7).

[Chen and Sivapalan, 2020] Chen, X. and Sivapalan, M. (2020). Hydrological basis of the budyko curve: Data-guided exploration of the mediating role of soil moisture. *Water Resources Research*, 56(10).

[Chen et al., 2021] Chen, Y., Feng, X., and Fu, B. (2021). An improved global remote-sensing-based surface soil moisture (rsssm) dataset covering 2003–2018. *Earth System Science Data*, 13(1):1–31.

[Chiew et al., 1995] Chiew, F., Kamaladasa, N., Malano, H., and McMahon, T. (1995). Penman-monteith, fao-24 reference crop evapotranspiration and class-a pan data in australia. *Agricultural Water Management*, 28(1):9–21.

[Clark et al., 2015] Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., and et al. (2015). Improving the representation of hydrologic processes in earth system models. *Water Resources Research*, 51(8):5929–5956.

[Danabasoglu et al., 2020] Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., and et al. (2020). The community earth system model version 2 (cesm2). *Journal of Advances in Modeling Earth Systems*, 12(2).

[de Boer-Euser et al., 2016] de Boer-Euser, T., McMillan, H. K., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H. (2016). Influence of soil and climate on root zone storage capacity. *Water Resources Research*, 52(3):2009–2024.

[Delft High Performance Computing Centre (DHPC), 2022] Delft High Performance Computing Centre (DHPC) (2022). DelftBlue Supercomputer (Phase 1). `https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase1`.

[Dirmeyer, 2011] Dirmeyer, P. A. (2011). A history and review of the global soil wetness project (gswp). *Journal of Hydrometeorology*, 12(5):729–749.

[Dong et al., 2020] Dong, J., Dirmeyer, P. A., Lei, F., Anderson, M. C., Holmes, T. R., Hain, C., and Crow, W. T. (2020). Soil evaporation stress determines soil moisture-evapotranspiration coupling strength in land surface modeling. *Geophysical Research Letters*, 47(21).

[Döscher et al., 2021] Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., and et al. (2021). The ec-earth3 earth system model for the climate modelintercomparison project 6.

[Essery et al., 2020] Essery, R., Kim, H., Wang, L., Bartlett, P., Boone, A., Brutel-Vuilmet, C., Burke, E., Cuntz, M., Decharme, B., Dutra, E., and et al. (2020). Snow cover duration trends observed at sites and predicted by multiple models. *The Cryosphere*, 14(12):4687–4698.

[Eyring et al., 2016] Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958.

[Fisher and Koven, 2020] Fisher, R. A. and Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, 12(4).

[Gamo et al., 2013] Gamo, M., Shinoda, M., and Maeda, T. (2013). Classification of arid lands, including soil degradation and irrigated areas, based on vegetation and aridity indices. *International Journal of Remote Sensing*, 34(19):6701–6722.

[Golaz et al., 2019] Golaz, J., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-Davis, X. S., Bader, D. C., and et al. (2019). The doe e3sm coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, 11(7):2089–2129.

[Greve et al., 2020] Greve, P., Burek, P., and Wada, Y. (2020). Using the budyko framework for calibrating a global hydrological model. *Water Resources Research*, 56(6).

[Guhathakurta and Saji, 2013] Guhathakurta, P. and Saji, E. (2013). Detecting changes in rainfall pattern and seasonality index vis-à-vis increasing water scarcity in maharashtra. *Journal of Earth System Science*, 122(3):639–649.

[Halldin, 2006] Halldin (2006). Inter tropical convergence zone (itcz).

[Hargreaves and Samani, 1985] Hargreaves, G. H. and Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, 1(2):96–99.

[Hobeichi et al., 2018] Hobeichi, S., Abramowitz, G., Evans, J., and Ukkola, A. (2018). Derived optimal linear combination evapotranspiration (dolce): A;global gridded synthesis et estimate. *Hydrology and Earth System Sciences*, 22(2):1317–1336.

[Hobeichi et al., 2021] Hobeichi, S., Abramowitz, G., and Evans, J. P. (2021). Robust historical evapotranspiration trends across climate regimes. *Hydrology and Earth System Sciences*, 25(7):3855–3874.

[Ji et al., 2017] Ji, P., Yuan, X., and Liang, X.-Z. (2017). Do lateral flows matter for the hyperresolution land surface modeling? *Journal of Geophysical Research: Atmospheres*, 122(22).

[Jiang and Ryu, 2016] Jiang, C. and Ryu, Y. (2016). Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from breathing earth system simulator (bess). *Remote Sensing of Environment*, 186:528–547.

[Jung et al., 2019] Jung, Martin & Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., Reichstein, M., and et al. (2019). The fluxcom ensemble of global land-atmosphere energy fluxes. *Scientific Data*, 6(1).

[Kim and Mohanty, 2016] Kim, J. and Mohanty, B. P. (2016). Influence of lateral subsurface flow and connectivity on soil water storage in land surface modeling. *Journal of Geophysical Research: Atmospheres*, 121(2):704–721.

[L. Harris et al., 2023] L. Harris, B., M. Taylor, C., Quaife, T., and P. Harris, P. (2023). Contrasting responses of vegetation to intraseasonal rainfall in earth system models.

[Laio et al., 2001] Laio, F., Porporato, A., Ridolfi, L., and Rodriguez-Iturbe, I. (2001). Plants in water-controlled ecosystems: Active role in hydrologic processes and response to water stress. *Advances in Water Resources*, 24(7):707–723.

[Li et al., 2021] Li, J., Miao, C., Wei, W., Zhang, G., Hua, L., Chen, Y., and Wang, X. (2021). Evaluation of cmip6 global climate models for simulating land surface energy and water fluxes during 1979–2014. *Journal of Advances in Modeling Earth Systems*, 13(6).

[Li and Qu, 2018] Li, X. and Qu, Y. (2018). Evaluation of vegetation responses to climatic factors and global vegetation trends using glass lai from 1982 to 2010. *Canadian Journal of Remote Sensing*, 44(4):357–372.

[Lian et al., 2020] Lian, X., Piao, S., Li, L. Z., Li, Y., Huntingford, C., Ciais, P., Cescatti, A., Janssens, I. A., Peñuelas, J., Buermann, W., and et al. (2020). Summer soil drying exacerbated by earlier spring greening of northern vegetation. *Science Advances*, 6(1).

[Lovato et al., 2022] Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., Fogli, P. G., Cherchi, A., Bellucci, A., Gualdi, S., and et al. (2022). Cmip6 simulations with the cmcc earth system model (cmcc-esm2). *Journal of Advances in Modeling Earth Systems*, 14(3).

[Martens et al., 2017] Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. (2017). Gleam;v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, 10(5):1903–1925.

[Mueller et al., 2013] Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., and et al. (2013). Benchmark products for land evapotranspiration: landflux-eval multi-data set synthesis. *Hydrology and Earth System Sciences*, 17(10):3707–3720.

[Mueller and Seneviratne, 2014] Mueller, B. and Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in cmip5 simulations. *Geophysical Research Letters*, 41(1):128–134.

[O. and Orth, 2021] O., S. and Orth, R. (2021). Global soil moisture data derived through machine learning trained with in-situ measurements. *Scientific Data*, 8(1).

[Oliveras and Malhi, 2016] Oliveras, I. and Malhi, Y. (2016). Many shades of green: The dynamic tropical forest–savannah transition zones. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1703):20150308.

[Peel et al., 2007] Peel, M. C., Finlayson, B. L., and McMahon, T. A. (2007). Updated world map of the köppen-geiger climate classification. *Hydrology and Earth System Sciences*, 11(5):1633–1644.

[Pielke et al., 1998] Pielke, R. A., Avissar, R., Raupach, M., Dolman, J. A., Zeng, X., and Denning, S. A. (1998). Interactions between the atmosphere and terrestrial ecosystems: Influence on weather and climate. *Global Change Biology*, 4(5):461–475.

[Qiang et al., 2021] Qiang, G., Kino, K., Li, S., Nitta, T., Takeshima, A., and et al. (2021). Description of the land surface model matsiro6. 70.

[Qiao et al., 2022] Qiao, L., Zuo, Z., and Xiao, D. (2022). Evaluation of soil moisture in cmip6 simulations. *Journal of Climate*, 35(2):779–800.

[Schulzweida, 2022] Schulzweida, U. (2022). Cdo user guide. *Max Planck Institute for Meteorology*.

[Sellar et al., 2019] Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., and et al. (2019). Ukesm1: Description and evaluation of the u.k. earth system model. *Journal of Advances in Modeling Earth Systems*, 11(12):4513–4558.

[Seneviratne et al., 2010] Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4):125–161.

[Singh et al., 2020] Singh, C., Wang-Erlandsson, L., Fetzer, I., Rockström, J., and van der Ent, R. (2020). Rootzone storage capacity reveals drought coping strategies along rainforest-savanna transitions. *Environmental Research Letters*, 15(12):124021.

[Tallaksen and Stahl, 2014] Tallaksen, L. M. and Stahl, K. (2014). Spatial and temporal patterns of large-scale droughts in europe: Model dispersion and performance. *Geophysical Research Letters*, 41(2):429–434.

[Ukkola et al., 2016] Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., Decker, M., and Haughton, N. (2016). Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environmental Research Letters*, 11(10):104012.

[van den Berg, 2023] van den Berg, R. G. (2023). How to run levene's test in spss?

[van den Hurk et al., 2016] van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., and et al. (2016). Ls3mip (v1.0) contribution to cmip6: The land surface, snow and soilmoisture model intercomparison project – aims, setup and expected outcome. *Geoscientific Model Development*, 9(8):2809–2832.

[Van Loon, 2015] Van Loon, A. F. (2015). Hydrological drought explained. *WIREs Water*, 2(4):359–392.

[van Oorschot et al., 2021] van Oorschot, F., van der Ent, R. J., Hrachowitz, M., and Alessandri, A. (2021). Climate-controlled root zone parameters show potential to improve water flux simulations by land surface models. *Earth System Dynamics*, 12(2):725–743.

[Vieira et al., 2020] Vieira, R. M., Tomasella, J., Barbosa, A. A., Martins, M. A., Rodriguez, D. A., Rezende, F. S., Carriello, F., and Santana, M. D. (2020). Desertification risk assessment in northeast brazil: Current trends and future scenarios. *Land Degradation; Development*, 32(1):224–240.

[Wang et al., 2020] Wang, Z., Zhan, C., Ning, L., and Guo, H. (2020). Evaluation of global terrestrial evapotranspiration in cmip6 models. *Theoretical and Applied Climatology*, 143(1-2):521–531.

[Wang-Erlandsson et al., 2016] Wang-Erlandsson, L., Bastiaanssen, W. G., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I., Guerschman, J. P., Keys, P. W., Gordon, L. J., Savenije, H. H., and et al. (2016). Global root zone storage capacity from satellite-based evaporation. *Hydrology and Earth System Sciences*, 20(4):1459–1481.

[Watanabe et al., 2011] Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., and et al. (2011). Miroc-esm: Model description and basic results of cmip5-20c3m experiments.

[Wouters et al., 2019] Wouters, H., Petrova, I. Y., van Heerwaarden, C. C., Vilà-Guerau de Arellano, J., Teuling, A. J., Meulenberg, V., Santanello, J. A., and Miralles, D. G. (2019). Atmospheric boundary layer dynamics from balloon soundings worldwide: Class4gl v1.0. *Geoscientific Model Development*, 12(5):2139–2153.

[Zach, 2022] Zach (2022). Kruskal-wallis test: Definition, formula, and example.

[Zampieri et al., 2012] Zampieri, M., Serpetzoglou, E., Anagnostou, E., Nikolopoulos, E., and Papadopoulos, A. (2012). Improving the representation of river–groundwater interactions in land surface modeling at the regional scale: Observational evidence and parameterization applied in the community land model. *Journal of Hydrology*, 420-421:72–86.

[Zeng et al., 2018] Zeng, Z., Peng, L., and Piao, S. (2018). Response of terrestrial evapotranspiration to earth's greening. *Current Opinion in Environmental Sustainability*, 33:9–25.

[Zhang et al., 2016] Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y. Y., and et al. (2016). Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific Reports*, 6(1).

[Zulkafli et al., 2013] Zulkafli, Z., Buytaert, W., Onof, C., Lavado, W., and Guyot, J. L. (2013). A critical assessment of the jules land surface model hydrology for humid tropical environments. *Hydrology and Earth System Sciences*, 17(3):1113–1132.

A — Study areas: Land cover

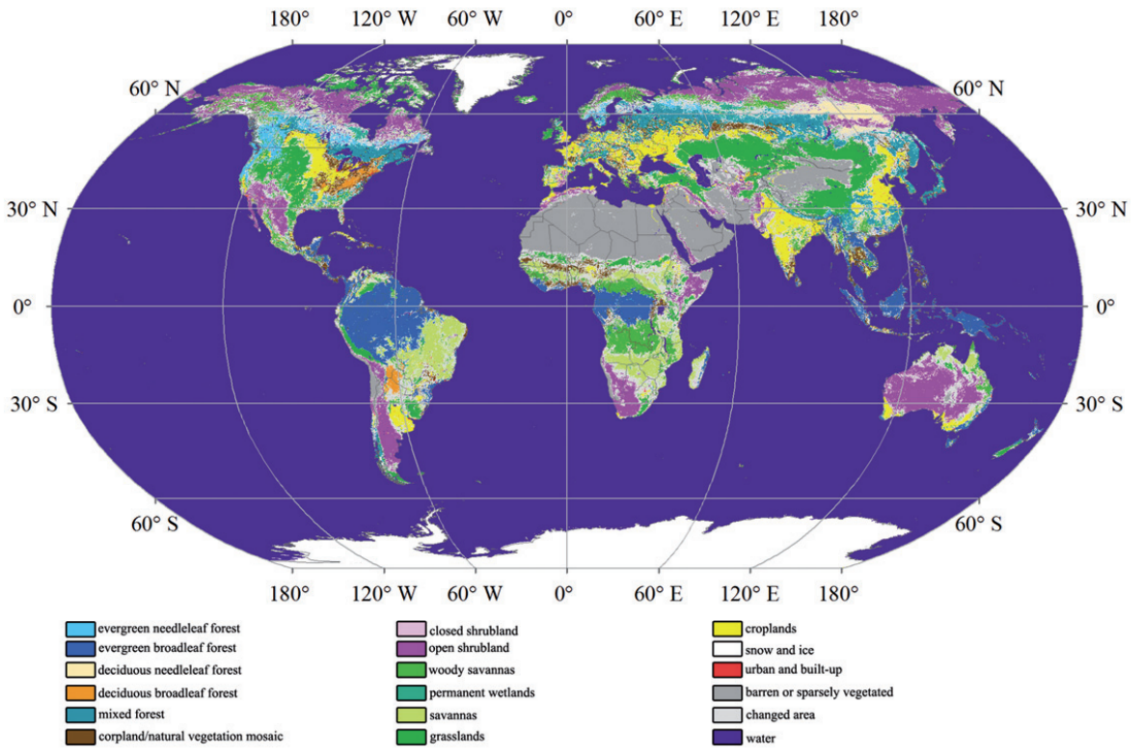


Figure A.1: *Global land cover type from 2001 until 2010 [Li and Qu, 2018]*

Study area	Dominant land cover type(s)
A: East Brazil	Savannas
B: West Africa	Savannas and cropland/natural vegetation mosaic
C: Central US	Woody savannas and croplands
D: Southeast Australia	Evergreen broadleaf forest and croplands
E: East US	cropland/natural vegetation mosaic and deciduous broadleaf forest
F: East China	Croplands and evergreen needleleaf forest

Table A.1: *Dominant land cover type(s) for all study areas*

B — Study areas: P, T and E seasonality

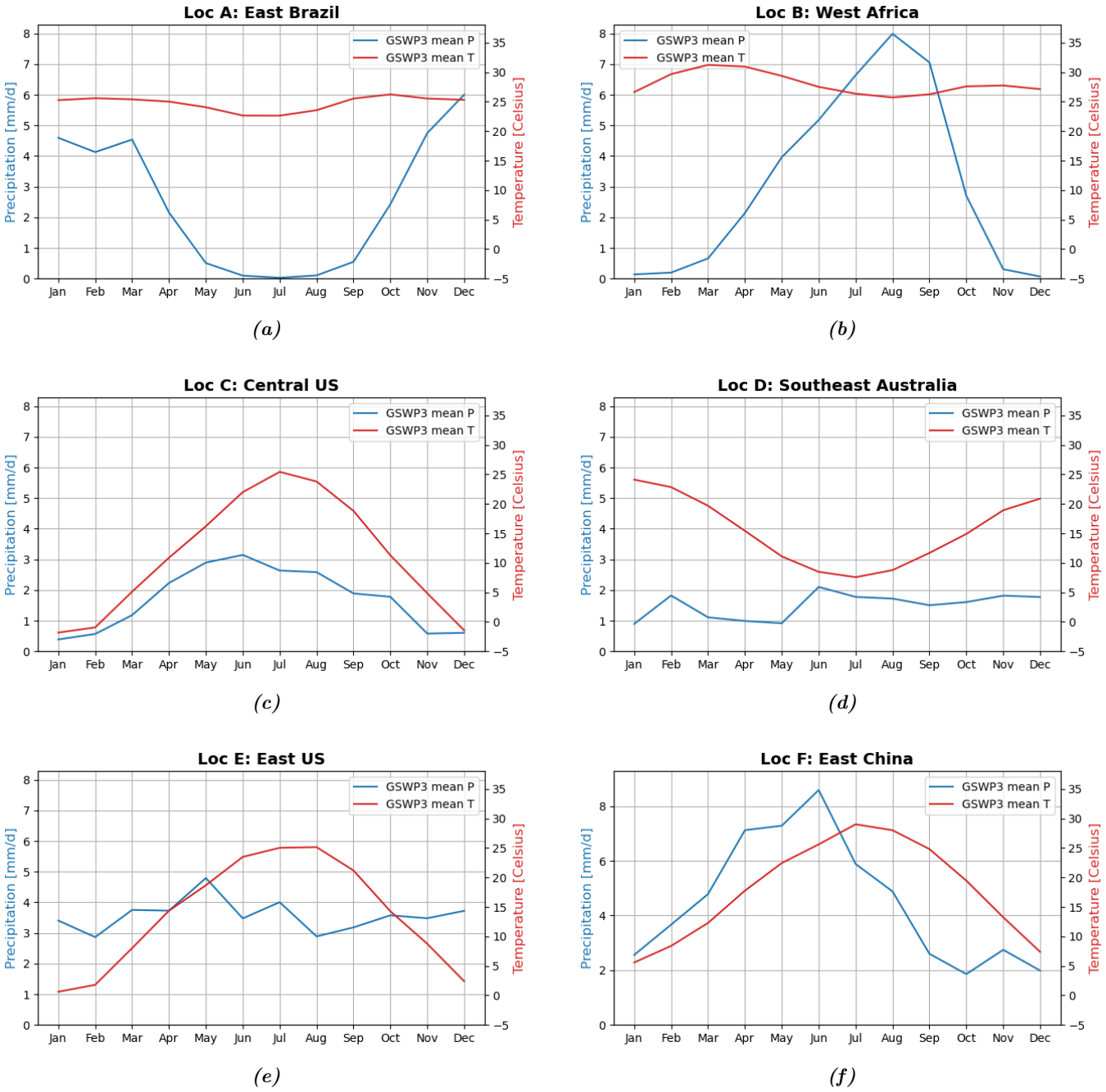


Figure B.1: *Monthly mean precipitation in mm per day and monthly mean temperature in Celsius within the time series 2001-2010 for all study areas*

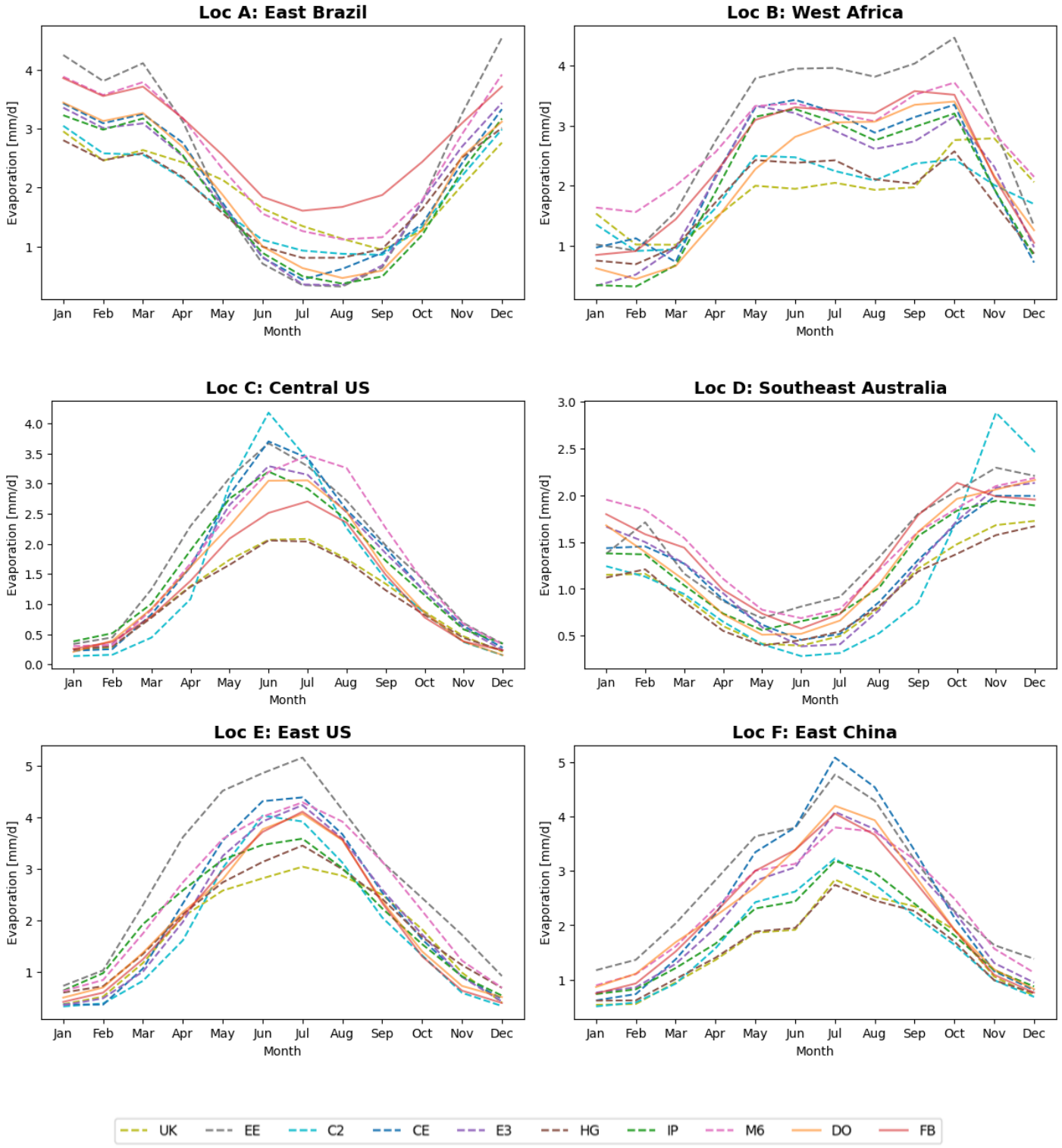


Figure B.3: *Monthly mean evaporation in mm per day for all LSMs and reference data DO and FB over the investigated time series for all study areas*

C — Study areas: Mean reference soil moisture

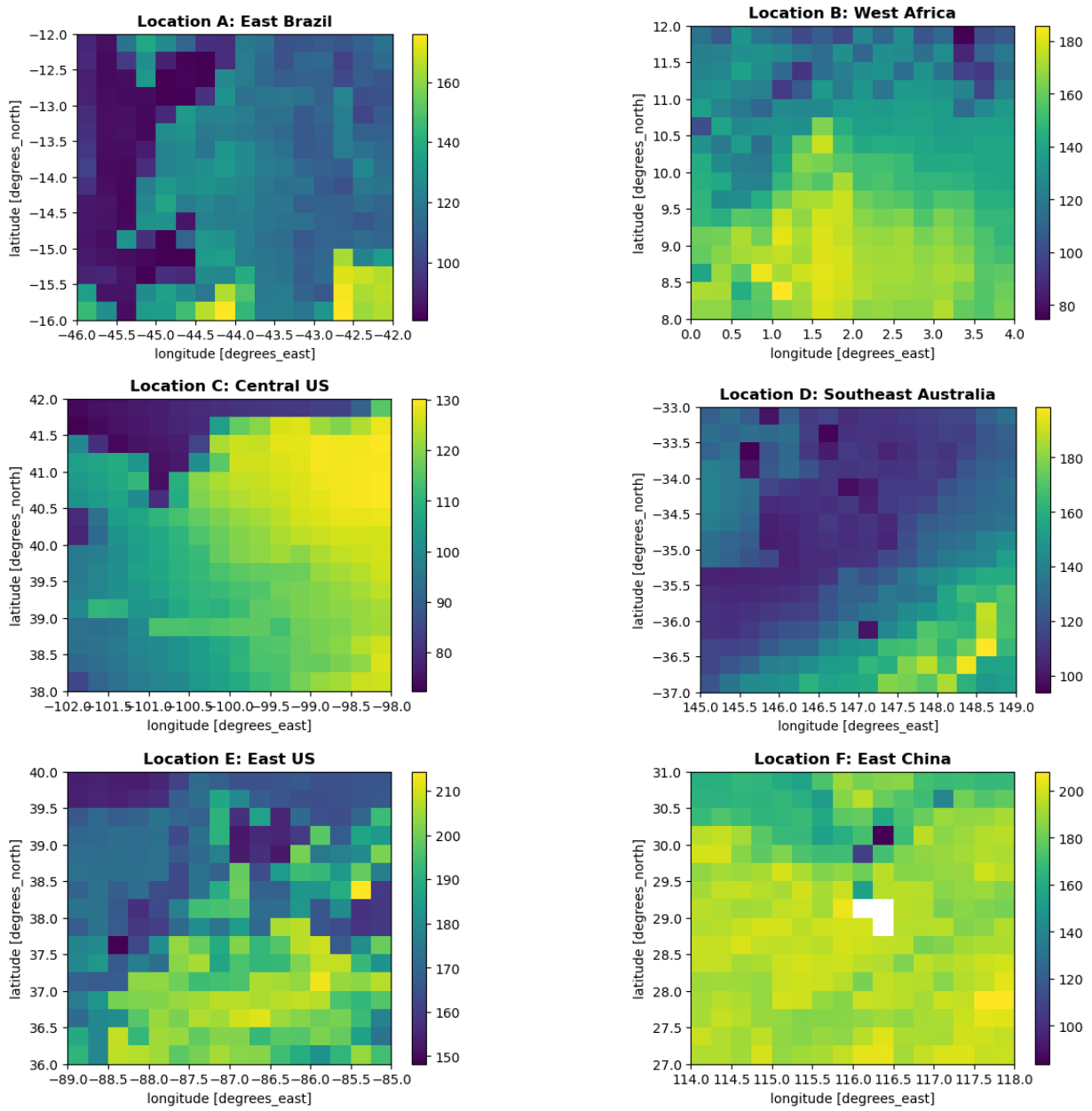


Figure C.1: *Mean spatial distribution in mm between 2001 and 2010 of the SoMo.ml soil moisture data for all study areas*

D — Study areas: LSM soil moisture and run-off driest period

LSM	Soil layer depth [m]
IP	2.0
HG	2.0
UK	2.0
CE	35.18
M6	9.0
C2	8.03

Table D.1: *The soil layer depths incorporated in the LSM models involved in the computation of the total soil moisture*

	Location A: dry period of 2007/2007			Location B: dry period of 2001/2002		
Dataset	**ΔS [mm]**	**-SMD - ΔS [mm]**	**R [mm]**	**ΔS [mm]**	**-SMD - ΔS [mm]**	**R [mm]**
MIROC6	-250.86	-149.14	2.41	-303.02	-27.71	29.49
EC-Earth3-veg	-	-	17.11	-	-	25.17
UKESM1-0-LL	-289.43	-10.97	39.91	-235.72	6.48	58.66
CESM2	-221.30	-14.40	79.62	-237.65	53.38	171.11
E3SM-1-1	-	-	52.18	-	-	162.63
CMCC-ESM2	-197.47	5.52	114.28	-218.56	57.75	186.67
HadGEM3-GC31-LL	-226.61	5.77	66.06	-159.91	16.25	87.55
IPSL-CM6A-LR	-190.45	34.63	71.81	-254.99	135.65	261.39

	Location C: dry period of 2003/2003			Location D: dry period of 2005/2006		
Dataset	**ΔS [mm]**	**-SMD - ΔS [mm]**	**R [mm]**	**ΔS [mm]**	**-SMD - ΔS [mm]**	**R [mm]**
MIROC6	-134.64	-10.14	1.32	-77.36	0.1	22.10
EC-Earth3-veg	-	-	2.11	-	-	28.32
UKESM1-0-LL	-59.34	59.34	15.65	-84.74	27.14	28.90
CESM2	-103.48	42.48	11.90	-80.76	-58.88	73.90
E3SM-1-1	-	-	22.71	-	-	91.04
CMCC-ESM2	-82.33	30.45	21.92	-94.44	-5.31	73.64
HadGEM3-GC31-LL	-43.48	43.48	7.06	-74.02	22.96	23.89
IPSL-CM6A-LR	-42.95	11.62	19.75	-115.69	18.86	91.95

	Location E: dry period of 2007/2007			Location F: dry period of 2003/2003		
Dataset	**ΔS [mm]**	**-SMD - ΔS [mm]**	**R [mm]**	**ΔS [mm]**	**-SMD - ΔS [mm]**	**R [mm]**
MIROC6	-179.44	-23.44	240.35	-219.80	48.03	368.74
EC-Earth3-veg	-	-	192.15	-	-	418.90
UKESM1-0-LL	-214.11	118.97	295.70	-275.50	158.45	500.70
CESM2	-316.95	155.19	446.85	-328.57	255.85	634.34
E3SM-1-1	-	-	399.71	-	-	455.48
CMCC-ESM2	-164.15	13.25	630.93	-168.08	7.99	897.62
HadGEM3-GC31-LL	-209.49	121.53	219.66	-217.27	161.57	514.12
IPSL-CM6A-LR	-260.80	152.44	371.73	-177.59	121.86	544.92

Table D.2: *The total decrease in soil moisture (ΔS), the comparison with the soil moisture deficit and the total run-off for the investigated dry period in time for all study areas*

E — Study areas: Water balance and SM storage over time

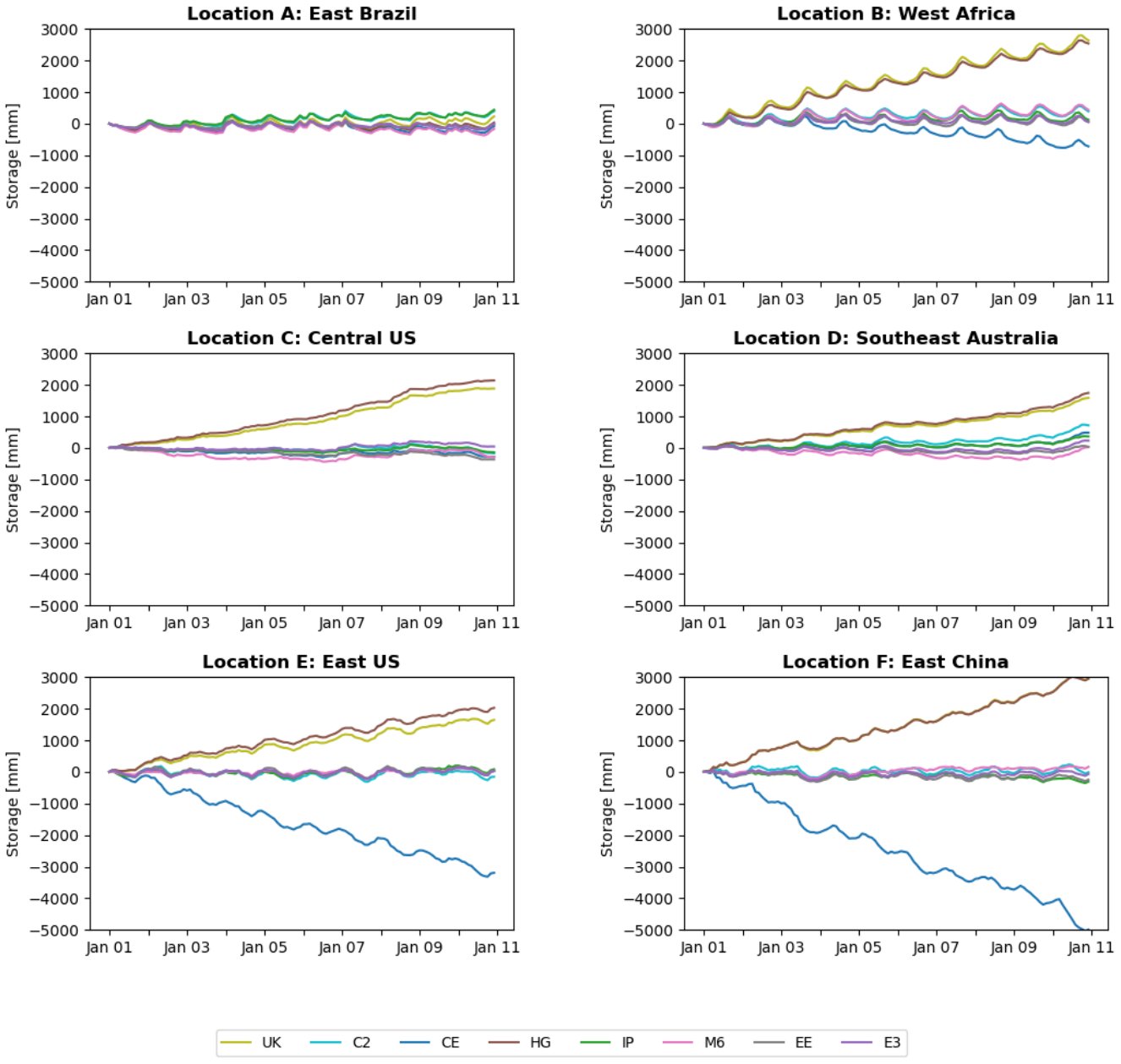


Figure E.2: *The change in storage over time based on the water balance (Eq: 2.2), for all LSMs and study areas*

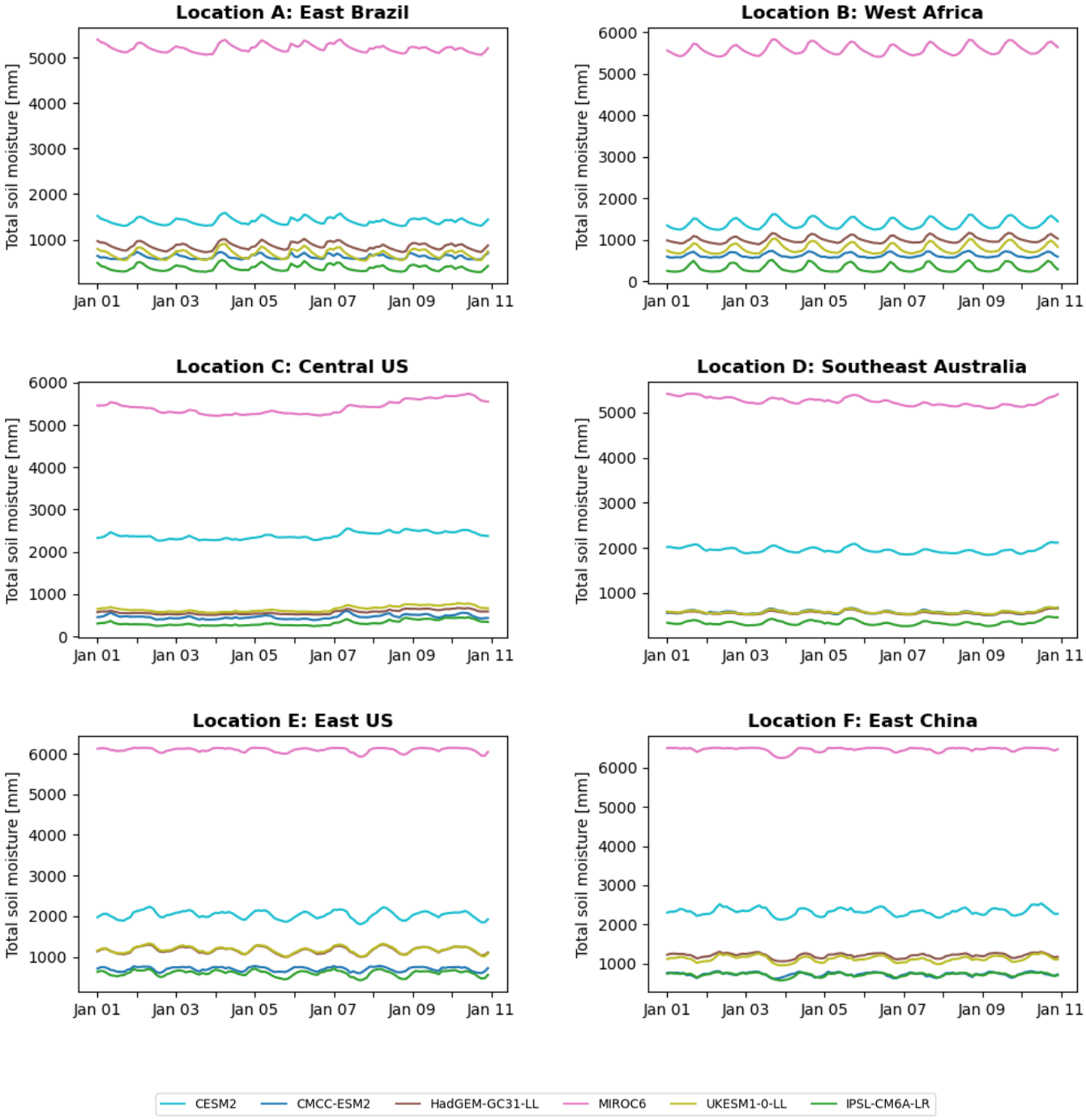


Figure E.4: *The total soil moisture over time for all study areas. The total soil moisture data from E3 and EE were not available for this research.*

F — Global analysis: JJA and DJF

JJA

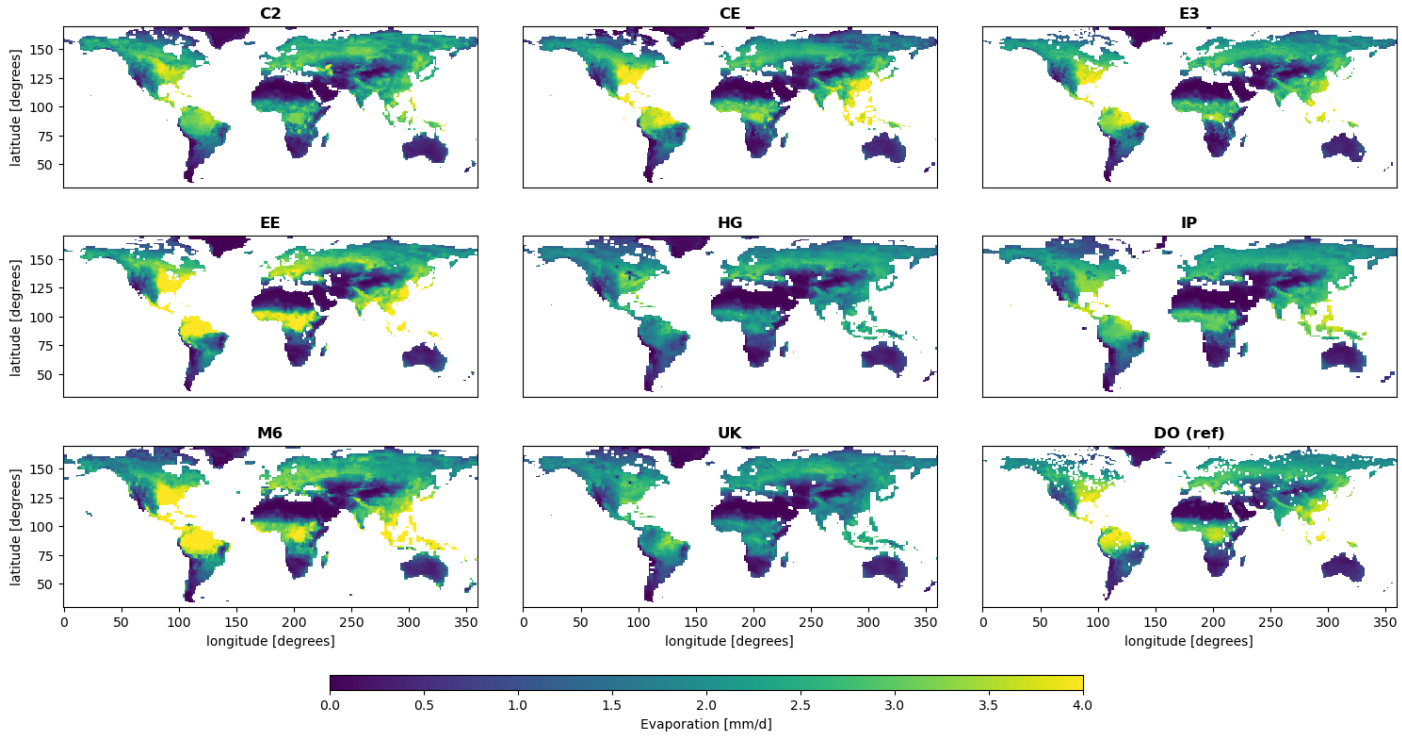


Figure F.1: *Spatial characteristics of the individual mean LSM evaporation output as well as the DO reference product for the **J**une, **J**uly and **A**ugust months between 2001 and 2010*

DJF

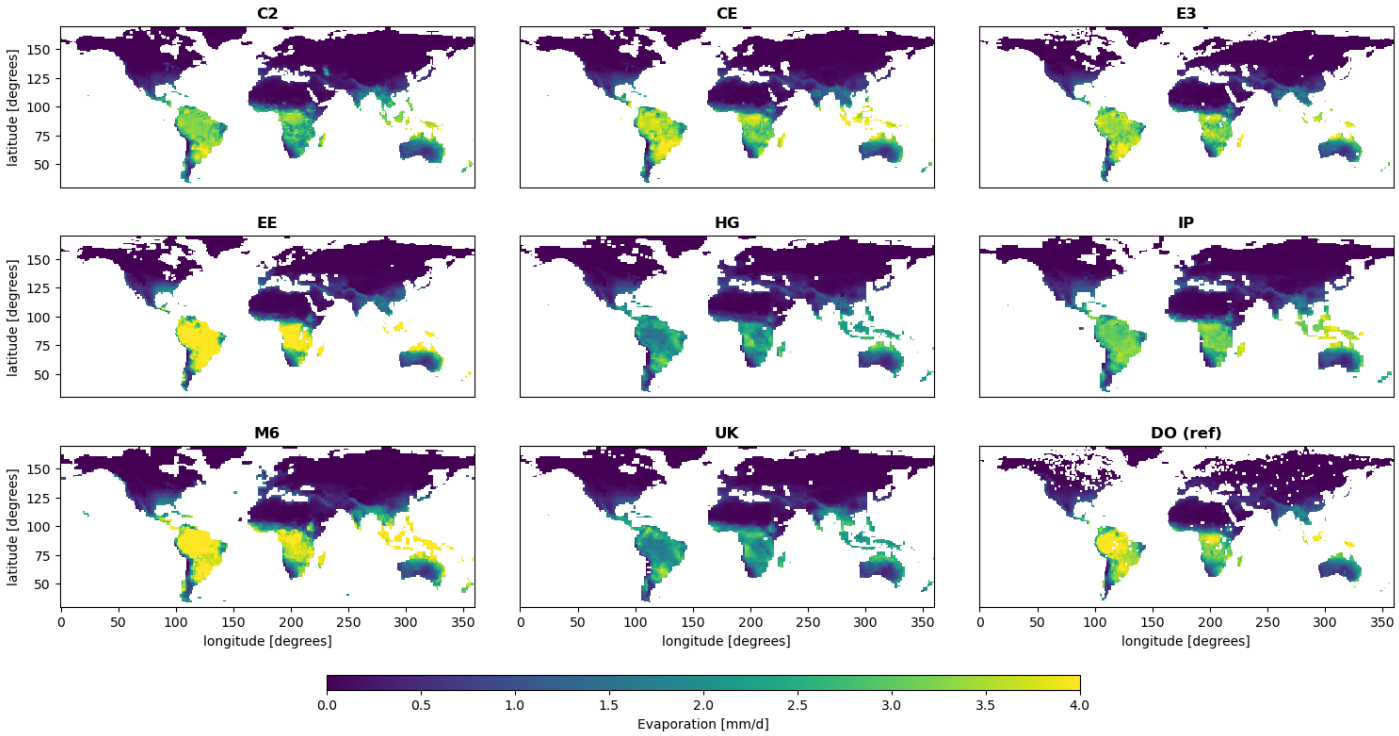


Figure F.2: *Spatial characteristics of the individual mean LSM evaporation output as well as the DO reference product for the* ***D****ecember,* ***J****anuari and* ***F****ebruari months between 2001 and 2010*

G — Köppen-Geiger climate classification

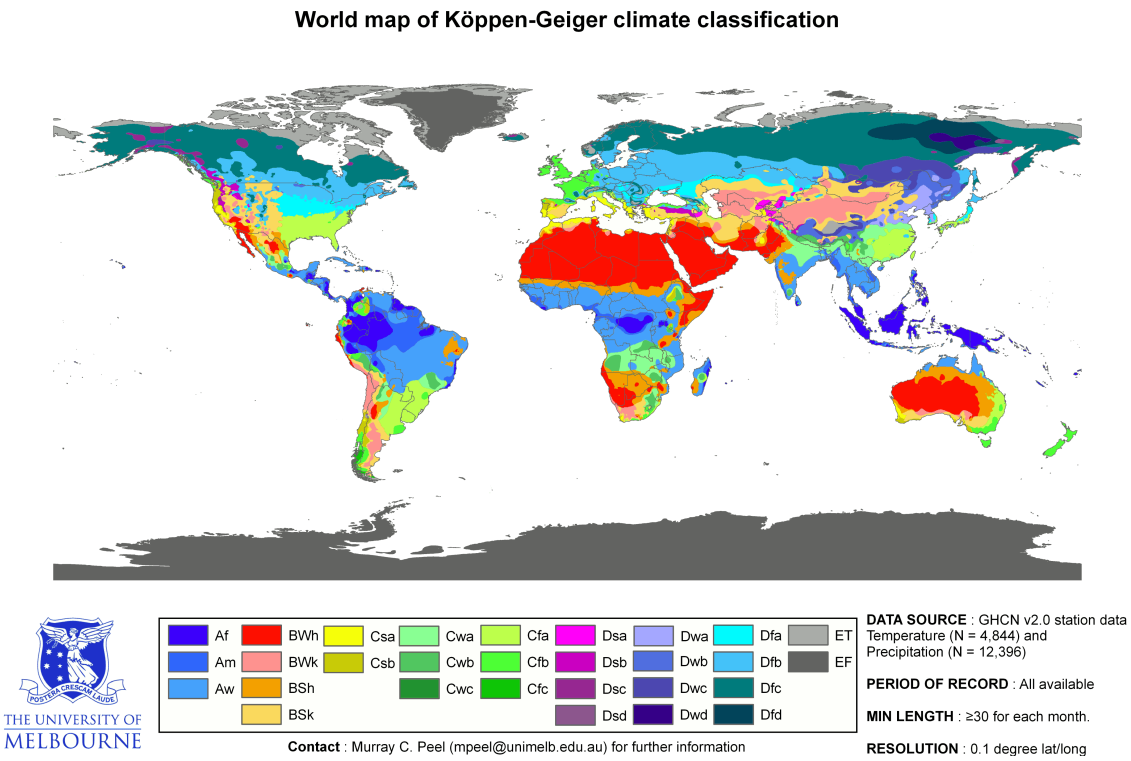


Figure G.1: *World map of the Köppen - Geiger climate classification [Peel et al., 2007]*

Af	Tropical rainforest	**Dfa**	Hot-summer humid continental
Am	Tropical monsoon	**Dfb**	Warm-summer humid continental
Aw	Tropical savanna	**Dfc**	Subarctic
BSh	hot semi-arid (steppe)	**Dfd**	Extremely cold subarctic
BSk	Cold semi-arid (steppe)	**Dsa**	Hot, dry-summer continental
BWh	Hot desert	**Dsb**	Warm, dry-summer continental
BWk	Cold desert	**Dsc**	Dry-summer subarctic cold summer
Cfa	Humid subtropical	**Dsd**	Dry-summer subarctic very cold winter
Cfb	Temperate oceanic	**Dwa**	Monsoon-influenced hot-summer humid continental
Cfc	Subpolar oceanic	**Dwb**	Monsoon-influenced warm-summer humid continental
Csa	Hot-summer mediterranean	**Dwc**	Monsoon-influenced subarctic
Csb	Warm-summer Mediterranean	**Dwd**	Monsoon-influenced extremely cold subarctic
Cwa	Monsoon-influenced humid subtropical	**EF**	Ice cap
Cwb	Subtropical highland or temperate oceanic	**ET**	Tundra
Cwc	Cold subtropical highland or subpolar oceanic		

Table G.1: *Clarification of the types of climates used in the Köppen-Geiger classification*

H — Standard deviation reference data

As discussed in section 2.5.4, the standard deviation of the DOLCE V3 anomaly needed to be compared with that of ERA5, FLUXCOM RS and GLEAM. This was done for locations A and C. The results in Table H.1 showed that the standard deviation of DOLCE V3 remained close to GLEAM and ERA5 while the standard deviation of FLUXCOM-RS was lower. This table suggested that the standard deviation of DOLCE V3 was not significantly influenced by the fact that it was a combination of other reference data sets. Therefore, suggesting that it was valid to compare the model sensitivity of individual LSMs with reanalysis data sets such as DO.

	Loc A	**Loc C**
DOLCE V3	0.36	0.29
ERA5	0.32	0.30
FLUXCOM RS	0.16	0.18
GLEAM	0.43	0.31

Table H.1: *The standard deviation in mm per day of the anomalies for locations A and C for various benchmark evaporation products*

I — T-test LSM evaporation means

Multiple Comparisons

Dependent Variable:

Bonferroni

(I) Groep		Mean Difference (I-J)	Std. Error	Sig.
C2	CE	-0.058699952466808	0.032415316912639	1.000
	E3	-0.039064809872198	0.032415316912639	1.000
	EE	-,263388897162521*	0.032415316912639	0.000
	HG	,278454309600493*	0.032415316912639	0.000
	IP	0.021589848964885	0.032415316912639	1.000
	M6	-,263100111523831*	0.032415316912639	0.000
	UK	,202571985217423*	0.032415316912639	0.000
CE	C2	0.058699952466808	0.032415316912639	1.000
	E3	0.019635142594610	0.032415316912639	1.000
	EE	-,204688944695713*	0.032415316912639	0.000
	HG	,337154262067301*	0.032415316912639	0.000
	IP	0.080289801431693	0.032415316912639	0.376
	M6	-,204400159057023*	0.032415316912639	0.000
	UK	,261271937684231*	0.032415316912639	0.000
E3	C2	0.039064809872198	0.032415316912639	1.000
	CE	-0.019635142594610	0.032415316912639	1.000
	EE	-,224324087290323*	0.032415316912639	0.000
	HG	,317519119472691*	0.032415316912639	0.000
	IP	0.060654658837083	0.032415316912639	1.000
	M6	-,224035301651632*	0.032415316912639	0.000
	UK	,241636795089622*	0.032415316912639	0.000

EE	C2	,263388897162521*	0.032415316912639	0.000
	CE	,204688944695713*	0.032415316912639	0.000
	E3	,224324087290323*	0.032415316912639	0.000
	HG	,541843206763014*	0.032415316912639	0.000
	IP	,284978746127406*	0.032415316912639	0.000
	M6	0.000288785638691	0.032415316912639	1.000
	UK	,465960882379945*	0.032415316912639	0.000
HG	C2	-,278454309600493*	0.032415316912639	0.000
	CE	-,337154262067301*	0.032415316912639	0.000
	E3	-,317519119472691*	0.032415316912639	0.000
	EE	-,541843206763014*	0.032415316912639	0.000
	IP	-,256864460635608*	0.032415316912639	0.000
	M6	-,541554421124324*	0.032415316912639	0.000
	UK	-0.075882324383070	0.032415316912639	0.544
IP	C2	-0.021589848964885	0.032415316912639	1.000
	CE	-0.080289801431693	0.032415316912639	0.376
	E3	-0.060654658837083	0.032415316912639	1.000
	EE	-,284978746127406*	0.032415316912639	0.000
	HG	,256864460635608*	0.032415316912639	0.000
	M6	-,284689960488716*	0.032415316912639	0.000
	UK	,180982136252538*	0.032415316912639	0.000
M6	C2	,263100111523831*	0.032415316912639	0.000
	CE	,204400159057023*	0.032415316912639	0.000
	E3	,224035301651632*	0.032415316912639	0.000
	EE	-0.000288785638691	0.032415316912639	1.000
	HG	,541554421124324*	0.032415316912639	0.000
	IP	,284689960488716*	0.032415316912639	0.000
	UK	,465672096741254*	0.032415316912639	0.000
UK	C2	-,202571985217423*	0.032415316912639	0.000
	CE	-,261271937684231*	0.032415316912639	0.000
	E3	-,241636795089622*	0.032415316912639	0.000
	EE	-,465960882379945*	0.032415316912639	0.000
	HG	0.075882324383070	0.032415316912639	0.544
	IP	-,180982136252538*	0.032415316912639	0.000
	M6	-,465672096741254*	0.032415316912639	0.000

Based on observed means.

*. The mean difference is significant at the ,05 level.

Figure I.1: *The mean difference between the LSMs with Bonferroni correction among all global LSM evaporation means. The Sig. is the p-value after the correction.*