Evaluating the effects of precipitation and evapotranspiration on soil moisture variability

Xuan Xi[1], Pierre Gentine[2], Qianlai Zhuang[1,3], Seungbum Kim[4]

[1]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, 47907.

[2]Department of Earth of Environmental Engineering, Columbia University, New York, NY, 10027.

[3]Department of Agronomy, Purdue University, West Lafayette, IN 47907.

[4]NASA Jet Propulsion Laboratory, Pasadena, CA 91109.

Corresponding author: Qianlai Zhuang (qzhuang@purdue.edu)

Key Points:

- Models underestimate weekly to seasonal variability and overestimate seasonal to annual variability of external effects on soil moisture.
- Simulated variability of precipitation and evapotranspiration is underestimated as being dominant factors controlling soil moisture.
- Earth system models shall be improved to correctly characterize the effects of precipitation and evapotranspiration on soil moisture.

Abstract

The effects of precipitation (Pr) and evapotranspiration (ET) on soil moisture play an essential role in the land-atmosphere system. Here we evaluate multimodel differences of these effects within the Coupled Model Intercomparison Project Phase 5 (CMIP5) compared to Soil Moisture Active Passive (SMAP) products in the frequency domain. The variability of surface soil moisture (SSM), Pr, and ET within three frequency bands (7 ~ 30 days, 30 ~ 90 days, and 90 ~ 365 days) after normalization is quantified using Fourier transform. We then analyze the impact of ET and Pr on SSM variability based on a transfer function assuming these variables with a linear time-invariant (LTI) system. For the simulated effects of ET and Pr on SSM variability, models underestimate them in the two higher frequency bands and overestimate them in the lowest frequency band but show better estimates in transitional zones between dry and wet climates. Besides, the effects on SSM by Pr and ET are found to be different across the three frequency bands, and models underestimate the one of Pr and ET as the dominant factor controlling SSM variability in each frequency band. This study identifies the spatiotemporal distribution of the CMIP5 model deficiencies in simulating ET and Pr effects on SSM. Overcoming these deficiencies could improve the interpretability and predictability of Earth system models in simulating interactions among the three variables.

Plain Language Summary

Surface climate is influenced by the interactions between the land surface and atmosphere boundary, and soil moisture is a key component of these physical processes. Precipitation and evapotranspiration, as two major variables involved in these interactions, have been largely regarded as essential processes affecting soil moisture dynamics. However, Earth system models have large uncertainties in simulating these effects. This study identifies that (1) models

underestimate the total effect of precipitation and evapotranspiration on soil moisture variability at weekly to seasonal time scales and overestimate it at seasonal to annual time scales; (2) soil moisture is mainly affected by precipitation at shorter scales and by evapotranspiration at longer time scales, and models underestimate the degree of this control over the whole weekly to annual frequency band; (3) model generally have better performance in the transitional climate regions on capturing the effects of precipitation and evapotranspiration on soil moisture. This study reveals the deficiencies of Earth system models in simulating the relationships between soil moisture, precipitation, and evapotranspiration compared to satellite observations, which will help improve the quantification of soil moisture dynamics in these models.

1 Introduction

As one of the essential components in the Earth system, soil moisture plays an important role in land-atmosphere interactions (Green et al., 2019; Koster et al., 2004; Seneviratne et al., 2006; Seneviratne et al., 2010). The exploration and quantification of land-atmosphere interactions are significant for Earth system study and climate-change projections (Santanello et al., 2018; Seneviratne et al., 2010; Suni et al., 2015).

The dynamics of soil moisture (SM) depend on the interplay between variability in multiple hydrological processes, such as precipitation, interception, evapotranspiration, runoff, and drainage (Bonan, 1996). Since these processes are complex and show large heterogeneity spatiotemporally, it is hard to quantify the effects of their resulting impact on soil moisture. We here focus on the two largest fluxes: precipitation (Pr), which is the water source of soil moisture and also one of the atmospheric forcing variables for land surface processes; and evapotranspiration (ET), which is a primary water loss relative to soil moisture.

Both soil moisture-precipitation (SM-Pr) and soil moisture-evapotranspiration (SM-ET) interactions are some of the central issues in the climate research community and have been studied for a while (Berg and Sheffield, 2018; Dong et al., 2020; Koster et al., 2004; Seneviratne et al., 2010; Wang et al., 2007; Wei and Dirmeyer, 2012). Basically, soil moisture-atmosphere coupling can be separated into two parts: the coupling between SM and ET and the coupling between ET and Pr (Guo et al., 2016; Seneviratne et al., 2010; Wei and Dirmeyer, 2010). The SM-ET coupling is linked to the impact of SM on ET variability as a regulator of energy partitioning (Seneviratne et al., 2010) and is mostly a local process (Wei and Dirmeyer, 2012). On the other hand, the SM-Pr coupling, which includes the effect of SM on ET and the effect of ET on Pr, is more elusive due to the series of atmospheric processes, especially the interactions between ET and Pr (see Seneviratne et al., 2010).

Studies on the effects of Pr and ET on the temporal variability of SM focused on analyzing autocorrelations of SM time series. Considering the SM dynamics as being forced by a random precipitation time series (i.e., white noise) and damped by an exponential damping term related to evapotranspiration losses, the temporal variability of SM can be reasonably governed by a first-order Markov process, which results in the SM time series to exhibiting a red noise spectrum (Delworth and Manabe, 1988). Based on this, many studies worked on characterizing these effects from a time-frequency domain. The response of SM to Pr at long time scales was investigated and revealed the amplitude decrease and the phase shift of soil hydrology with soil depth (Wu et al., 2002). This phase shift as to how SM spectra related to Pr was further explored using the integral time scale to show that SM spectra decay more rapidly than a red noise due to Pr departing from white noise at high frequency, and the damping term of ET losses was found to be bounded by the maximum of ET (Katul et al., 2007). Similarly, the integral time scale was used to reveal the dynamics of SM memory and its correlation with Pr and ET (Ghannam et al., 2016). Based on previous studies (Katul et al., 2007), the SM spectrum could not be explained only by precipitation effects on longer time scales (Nakai et al., 2014). This concept has also been used to investigate the effects of Pr on SM variability on a regional scale (Zhou et al., 2020).

Although the SM-Pr and SM-ET couplings have been studied for a long time, the effects of Pr and ET on soil moisture variability are still not completely understood (Seneviratne et al., 2010). Even less understood is how Earth system models perform in capturing these effects globally and at different time scales. There are two major challenges. One is the lack of enough in-situ soil moisture measurements at the global scale. Nowadays, remote sensing technology, such as NASA's Soil Moisture Active Passive (SMAP) mission (Entekhabi et al., 2010), provides global observation of soil moisture at a high spatiotemporal resolution that can be used to constrain land-atmosphere interaction observations over different spatiotemporal scales (Guillod et al., 2015; Tuttle and Salvucci, 2016). Additionally, although it only provides surface soil moisture (top ~5cm of the soil column), several studies have shown that surface soil moisture (SSM) and root-zone soil moisture (RZSM) have strong correlations in quantifying surface flux (Akbar et al., 2018; Ford et al., 2014; Qiu et al., 2016), indicating that SSM can be regarded as a proxy for RZSM under most conditions (McColl et al., 2019). Another challenge is that, due to the complexity and the large number of processes involved in land-atmosphere interactions, the representation of couplings between SM, Pr, and ET highly relies on parameterizations within Earth system models, which leads to large uncertainties in identifying the effects of Pr and ET on SM variability. The transfer function (Haykin and Van Veen, 2007), as a mathematical representation of the differential equation of system dynamics, can be used to describe the relationship between the signal input and response assuming a linear time-invariant (LTI) system (Phillips et al., 2003) using a time-frequency analysis, without considering its specific structure and parameters. Therefore, it can be used to investigate the effects of Pr and ET on SM in the frequency domain, assuming they are nearly an LTI system. The spectral analysis based on the LTI system has been applied to other hydrological research like the runoff-storage relationship (Riegger and Tourian 2014) and the surface flow in the river during floods (Bailly-Comte et al., 2008).

The fifth phase of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012), which integrated a set of model experiments to improve our knowledge of climate change and climate variability, provides an opportunity for the multimodel assessment of land-atmospheric processes and variability. Evaluation of CMIP5 has been the ongoing interest of the research community (Yuan et al., 2021). Although evaluations of land-atmosphere interactions related to soil moisture within CMIP5 have been performed earlier (e.g., Berg and Sheffield, 2018; Dirmeyer et al., 2013; Levine et al., 2016), few studies have characterized the temporal behavior of SM globally in order to illustrate the model performance across frequency regimes. Therefore, this study takes advantage of the CMIP5 intercomparison project to evaluate 14 Earth system models (ESMs) in simulating the effects of Pr and ET on SSM variability. We aim to address two main objectives in this study: 1) how the effects of Pr and ET on SSM variability are at different time scales, and 2) how the ESMs within CMIP5 perform in capturing these effects. Specifically, these effects are analyzed within three frequency bands: 1) weekly to monthly time scales (1/7 ~ 1/30 day^{-1}), 2) monthly to seasonal time scales (1/30 ~ 1/90 day^{-1}), and 3) seasonal to annual time scales (1/90 ~ 1/365 day^{-1}) at the global scale. Further, a Fourier analysis is conducted to determine the variability and power spectra over those various periods (Thomson and Emery, 2014; Wilks, 2011). Similar approaches to decomposing the time series into different frequency bands have been used to understand the precipitation and soil moisture variability (Ruane and Roads, 2007; Wei et al., 2010; Xi et al., 2022).

In section 2, we first describe the models and data used. Then, we detail our methodology for spectral analysis. In section 3, we show the results of observation-based data in the first part. In the second part, we perform comparative analyses to evaluate the multimodel differences within CMIP5. In the third part, we investigate uncertainties that may exist in this study. Finally, in section 4, we summarize our findings and discuss the impacts of the research.

2 Methods

2.1 Overview

We first describe the data collection within CMIP5, SMAP observation data, and ERA5 reanalysis data. Second, we detail the methodology from data preprocessing to the final multimodel comparison (Figure 1). Specifically, section 2.3 describes the preprocessing of SMAP products and CMIP5 simulations. Section 2.4 defines the normalized variability of SSM, Pr, and ET and how to get them within the three frequency bands. Next, section 2.5 introduces two ratios used to investigate the effects of Pr and ET on SSM based on the normalized variability defined in section 2.4. Section 2.6 describes the spectral slopes of SSM, Pr, and ET time series and how to depict them as the color of noise. Finally, section 2.7 describes how the models are compared with the observation-based data and illustrates the significance test.

2.2 Data Organizing

CMIP5 integrated a set of model experiments to improve our knowledge of climate variability from past to present to future (Taylor et al., 2012). Here we use the daily simulations of 14 ESMs from the historical experiment within CMIP5. The models are selected based on the availability of daily outputs required for the spectral analysis within the same temporal coverage from 01/01/1950 to 12/31/2005 (Table S1). To evaluate the effects of Pr and ET (i.e., atmospheric water supply and loss) on SSM variability, we analyze the simulated SSM (top 10 cm), Pr, and ET (variable *mrsos*, *pr*, and *hfls* in the CMIP5 archive, respectively). We use only one ensemble member – "r1i1p1" (where *r* for realization, *i* for initialization, and *p* for physics).

Observation data of SSM are taken from SMAP (Entekhabi et al., 2010). For Pr and ET, we use reanalysis data from ERA5 (Copernicus Climate Change Service (C3S), 2017), the fifth-generation reanalysis of ECMWF (European Centre for Medium-Range Weather Forecasts) as the next generation of representative satellite-observational data, as a reference to compare with CMIP5 simulations on the global scale. To ensure that the data are consistent, we use datasets from SMAP and ERA5 with the same temporal coverage, spanning 1 April 2015 to 31 December 2020.

The NASA SMAP satellite was launched in January 2015 and has been measuring SSM (moisture in the top ~5 cm of the soil column) globally every 2~3 days (Entekhabi et al., 2010). SMAP soil moisture matches well *in situ* SSM observations (Chan et al., 2016, 2018; Colliander et al., 2017, 2021) and shows higher accuracy measured by a global average anomaly correlation over the majority of available land pixels compared to two other satellite products (Chen et al., 2018). Additionally, SMAP has been shown to have high information content relative to four other retrieval products of soil moisture (Kumar et al., 2018). In this study, we use its Level 3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 7 (O'Neill et al., 2020) with the retrievals from both 6 am descending passes and 6 pm ascending passes. Although its 6 pm retrievals show more degradation than its 6 am retrievals due to the required vertical thermal equilibrium assumption in its algorithm, this degradation was shown to be small (Chan et al., 2018; O'Neill et al., 2018). Therefore, we use both retrievals to best use the observational information. The Level 3 product of SMAP was developed based on geophysical parameters derived from its Level 1 and Level 2 products. It was spatiotemporally re-sampled to the global cylindrical EASE-Grid 2.0 to make each grid cell has a nominal size of approximately 36×36 km^2 regardless of longitude and latitude (Brodzik et al., 2012).

The reference observation-based data of precipitation (Pr) and evapotranspiration (ET) are collected from ERA5. ERA5 reanalysis is achieved by data assimilation, which combines weather forecasts with observations in an optimal way every few hours to produce the best estimate of the state of the atmosphere. In this way, ERA5 combines model data and observations into a globally complete and consistent dataset. ERA5 reanalysis has been evaluated extensively on regional and global scales and shows great improvements over its popular predecessor ERA-Interim and is a potential reference as proxies for observations for the hydrological process modeling (Jiao et al., 2021; Martens et al., 2020; Rivoire et al., 2021; Tarek et al., 2020). In this study, we use "total precipitation" (units: m) and "evaporation" (unit: m of water equivalent) estimates on single levels as the observation-based Pr and ET, respectively (Copernicus Climate Change Service (C3S), 2017). This dataset has a spatial resolution of 0.25°×0.25° for the atmosphere, spanning 1979 to the present, with an hourly temporal resolution. We collect the ERA5 hourly data within the same period as SMAP. Then we convert them into daily total precipitation and evapotranspiration (units: m) based on (Copernicus Climate Change Service (C3S), 2017):

$$Pr_d = \sum_{hr=1}^{23} Pr_{hr} + Pr_{d+1\ 00UTC} \tag{1}$$

$$ET_d = \sum_{hr=1}^{23} ET_{hr} + ET_{d+1\ 00UTC} \quad (2)$$

where hr is hour and d is the day of interest ($d + 1$ is the next day). This means that we need two days of data to get total precipitation and evapotranspiration per day. For example, to calculate total precipitation for 1 April 2015, we need hourly data on 1 April 2015 with time = 01 – 23 to cover 00 – 23 UTC for 1 April 2015 and the hourly data on 2 April 2015 with time = 00 to cover 23 – 24 UTC for 1 April 2015. In this way, we get daily precipitation and evapotranspiration time series (i.e., Pr_d and ET_d) for further analysis. We also use the same subset of ERA5 datasets and the same method to collect daily potential evaporation (PE) data (units: m) with the same temporal coverage for comparison with ET.

2.3 Data Preprocessing

The data in this study are preprocessed before the spectral analysis, as described in our previous study (Xi et al., 2022). Basically, since the global SSM retrievals from SMAP are temporally discontinuous on a daily time scale, we first perform a gap-filling to make it a daily dataset. For the model estimations within CMIP5 that span decades, there might be long-memory fluctuations on such time scales (Mudelsee, 2013). To avoid such long-memory trends from introducing errors into the power spectrum when performing Fourier analysis, we detrend the CMIP5 data to obtain a stationary signal by subtracting an optimal (least squares) fitted linear regression from original data. In this way, the time series after detrending has a mean value of zero, and we focus on their intra-annual fluctuations.

2.4 Normalized variability of SSM, Pr, and ET

Normalized variability of SSM (SSM_n), Pr (Pr_n), and ET (ET_n) of CMIP5 models and observation-based data are both calculated for comparison. We aim to use SSM_n, Pr_n, and ET_n to indicate the proportion of the temporal variability over different frequency bands. These normalized variabilities are further used to evaluate the effects of Pr and ET on SSM variability in section 2.5. The procedures to get SSM_n, Pr_n, and ET_n from time series of SSM ($ssm(t)$), Pr ($pr(t)$), and ET ($et(t)$), are shown in Figure 1 (for a detailed version, see Figure S1). The details of the steps to process SSM_n are explained in a previous study (Xi et al., 2022). The processing of Pr_n and ET_n follows a similar procedure. Here we give a basic idea of the strategy used to process X_n.

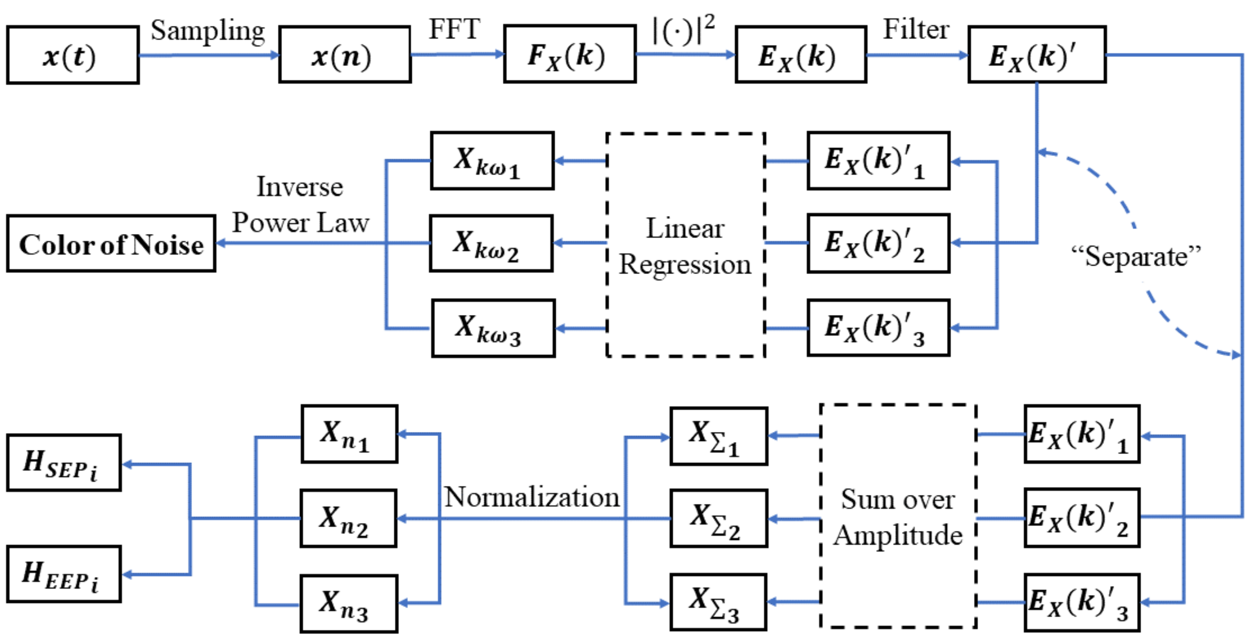

Figure 1. Steps to get the normalized variability (X_{n_1}, X_{n_2}, and X_{n_3}, hereafter collectively referred to as X_n) and the spectral slope (X_{kw_1}, X_{kw_2}, and X_{kw_3}, hereafter collectively referred to as X_{kw}) of the variable X from its time series ($X(t)$). X here means SSM, Pr, and ET, since the procedure to deal with $ssm(t)$, $pr(t)$, and $et(t)$ is the same. The number "1", "2", and "3" (hereafter being referred as i)

represent three frequency bands in the order of weekly to monthly (7 ~ 30 days), monthly to seasonal (30 ~ 90 days), and seasonal to annual (90 ~ 365 days) time scales. $x(n)$ is the discrete series sampled from $x(t)$. $F_X(k)$ is the amplitude spectrum of X from $x(n)$ using Fast Fourier Transform (FFT). $E_X(k)$ is the power spectrum of X as the square of the absolute value of its amplitude. $E_X(k)'$ is the filtered $E_X(k)$ to a frequency band within 7 to 365 days. ${E_X(k)'}_i$ is $E_X(k)'$ being "separated" into the three frequency bands: weekly to monthly ($i = 1$), monthly to seasonal ($i = 2$), and seasonal to annual ($i = 3$). The sum of spectral amplitudes of X ($X_{\sum_i}$) and X_{kw_i} is gotten from ${E_{SSM}(k)'}_i$ based on "sum over amplitude" and "linear regression" within the ith frequency band, respectively. X_{n_i} is gotten from $X_{\sum_i}$ based on normalization across the three frequency bands, and then ${H_{SEP}}_i$ and ${H_{EEP}}_i$ are two ratios used to analyze the effects of Pr and ET on SSM defined in section 2.5.

The computation of X_n for models and observations is the same. It is based on the Fast Fourier Transform (FFT), a faster algorithm for the Discrete Fourier Transform (DFT). They decompose the time series into orthogonal sinusoidal frequency components so that the variability within each component can be investigated separately. In this way, the oscillations of time series ($x(t)$) can be identified through the spectra in the frequency domain. All computations and statistical analyses in this study are programmed in MATLAB (http://www.mathworks.com/).

First, we use FFT to get the amplitude spectrum of X ($F_X(k)$) from $x(n)$, which is the discrete series sampled from $x(t)$ based on the sampling number (N) (i.e., the number of days). Then we get the power spectrum of X from its amplitude spectrum as $E_X(k) = |F_X(k)|^2$. We only keep $E_X(k)$ with the frequency ranges from 1/2 to $1/N$ day^{-1} since the spectrum is symmetrical about the Nyquist frequency ($f_s/2$, where f_s is sampling frequency). For all time-series data (i.e., CMIP5 simulations, SMAP, and ERA5 references), we use 1 day^{-1} as the sampling frequency from $x(t)$ to $F_X(k)$ since they are all with daily resolution.

Then, we restrict our investigation within a weekly to annual frequency band by using a low-pass filter and a high-pass filter with the cutoff frequency as 1/7 day^{-1} and 1/365 day^{-1}, respectively. Next, we separate the filtered $E_X(k)$ ($E_X(k)'$) into three frequency bands: weekly to monthly time scales (7 ~ 30 days), monthly to seasonal time scales (30 ~ 90 days), and seasonal to annual time scales (90 ~ 365 days). Finally, we define the normalized variability of *X* as the spectral power of each frequency band divided by the total spectral power of $E_X(k)'$:

$$X_{n_i} = \frac{\sum_j E_{X_i}(k_j)'}{\sum_{i=1}^{3} \sum_j E_{X_i}(k_j)'} \tag{3}$$

where $E_{X_i}(k_j)'$ represents the spectral power of *X* for the *j*th frequency in the *i*th frequency band, *i* is the ordinal number representing the three frequency bands from high to low, and *j* is the ordinal number of each frequency within each frequency band. Thus, we denote X_{n_i} as the normalized variability of *X* in the *i*th frequency band. In this way, X_{n_i}, as a value between 0 and 1, indicates the proportion of the temporal variability of $x(t)$ in the *i*th frequency band.

2.5 Analysis of the Effects of Pr and ET on SSM Variability

Both Pr and ET affect SSM variability. Pr is the water source of SSM, while ET is the water loss term affecting SSM. Thus, increasing Pr will increase SSM while increasing ET will decrease SSM (without considering the saturation condition), which can be expressed as (neglecting other processes):

$$\frac{\mathrm{d}ssm(t)}{\mathrm{d}t} = pr(t) - et(t) \tag{4}$$

Here we aim at examining the effects of ET and Pr on SSM variability within the three frequency bands based on the transfer function of a conceptual LTI system. The related theory of the LTI system and transfer function can be found in Appendix A.

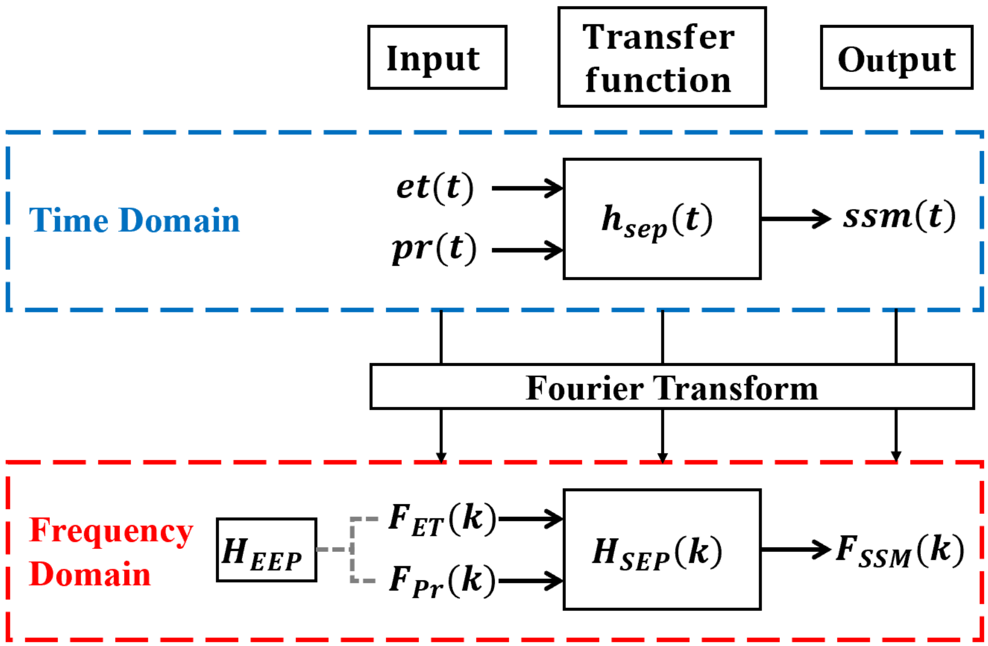

Figure 2. Diagram of the conceptual LTI system with the excitations as $et(t)$ and $p(t)$, the response as $ssm(t)$, and the transfer function as $h_{sep}(t)$ in the time domain. The form in the time domain is shown in the blue box. By performing Fourier transform, the corresponding form of the LTI system in the frequency domain is shown in the red box, where $F_{ET}(k)$, $F_{Pr}(k)$, and $F_{SSM}(k)$ is the Fourier transform (amplitude spectrum) of $et(t)$, $p(t)$, and $ssm(t)$, and $H_{SEP}(k)$ is the Fourier transform of the transfer function $h_{sep}(t)$. H_{EEP} is the fraction of ET variability to the sum of ET and Pr variability in the frequency domain.

To capture the total effects of ET and Pr on the SSM variability, we use a conceptual LTI system with the excitation as $et(t)$ and $p(t)$ together and the response as $ssm(t)$ (Figure 2). Since ET and Pr have different spectral characteristics in the frequency domain (Katul et al., 2007; Nakai et al., 2014; also from Figure 3 in section 3.1), here we separate their effects on SSM as two inputs and determine the total effects as an identical transfer function. Regarding this system as a “black-box” model, we can focus on the relationship between excitation (i.e., ET and Pr) and response (i.e., SSM) without caring about the internal variations of the system. In this way, the relationship between $ssm(t)$, $et(t)$, and $p(t)$ can be expressed :

$$ssm(t) = et(t) \otimes h_{sep}(t) + pr(t) \otimes h_{sep}(t) \tag{5}$$

where $h_{sep}(t)$ is the transfer function of the LTI system shown in Figure 2. Then, equation (2) can be expressed as spectrum analysis in the frequency domain:

$$F_{SSM}(k) = F_{ET}(k) \cdot H_{SEP}(k) + F_{Pr}(k) \cdot H_{SEP}(k) \tag{6}$$

where $H_{SEP}(k)$ is the Fourier transform of the transfer function $h_{sep}(t)$. Thus, the variations of the excitation and response spectra of the LTI system are determined by the transfer function $H_{SEP}(k)$ as:

$$H_{SEP}(k) = \frac{F_{SSM}(k)}{F_{ET}(k) + F_{Pr}(k)} \tag{7}$$

where $F_{ET}(k)$, $F_{Pr}(k)$, and $F_{SSM}(k)$ is the Fourier transform (amplitude spectrum) of $et(t)$, $p(t)$, and $ssm(t)$.

In order to characterize the total effects of ET and Pr on SSM variability within the three frequency bands, we process equation (5) based on the normalized variability (SSM_{n_i}, ET_{n_i}, and Pr_{n_i}) defined in section 2.4:

$$H_{SEP_{n_i}} = \frac{SSM_{n_i}}{ET_{n_i} + Pr_{n_i}} \tag{8}$$

where $H_{SEP_{n_i}}$ is the fraction of SSM variability to the sum of ET and Pr variability (i.e., demand and supply) in the *i*th frequency band. The higher this ratio, the stronger influences on the temporal variability of SSM by ET and Pr. We also aim to define the dominant factor on SSM variability (i.e., whether ET or Pr) within the three frequency bands. Therefore, we define another ratio:

$$H_{EEP_{n_i}} = \frac{ET_{n_i}}{ET_{n_i} + Pr_{n_i}} \tag{9}$$

where $H_{EEP_{n_i}}$ is the fraction of ET variability to the sum of ET and Pr variability in the *i*th frequency band. This ratio is greater than one-half means that ET has larger variability than Pr and thus a greater impact on the temporal variability of SSM and vice versa. In this way, we use $H_{SEP_{n_i}}$ and $H_{EEP_{n_i}}$ as two indicators to characterize the effects of ET and Pr on SSM variability in the three frequency bands – $H_{SEP_{n_i}}$ measures the total effect of ET and Pr on SSM variability and $H_{EEP_{n_i}}$ determines which process is dominant. A detailed procedure to get $H_{SEP_{n_i}}$ and $H_{EEP_{n_i}}$ can be found in Figure S1.

2.6 Analysis of Spectral Slope of SSM, Pr, and ET

The spectral slope exhibits characteristics of the soil moisture's physical behavior. This factor can explain how ET and Pr variability contribute to the spectrum of soil moisture (Katul et al., 2007). Being considered power-law noise signals, the spectral densities of time series vary as proportional to $1/f^{\beta}$ (i.e., inverse frequency), where β is the inverse number of the spectral slope (Bourke, 1998). In this way, the color of the noise, which is related to the power spectrum of noise signals, can be used to indicate the spectral slopes of SSM, Pr, and ET. The basic theory of the color of noise can be found in Text S2.

The noise colors can be divided into several types according to the slope of their power spectral density. We use white noise and five main colored noises (violet, blue, pink, red, and black noise) to characterize the spectral slopes for SSM ($SSM_{k\omega_i}$), Pr ($Pr_{k\omega_i}$), and ET ($ET_{k\omega_i}$) in the *i*th frequency band. The corresponding spectral slope (equal to β in inverse power law $1/f^{\beta}$) of violet, blue, white, pink, and red noise (or Brownian noise) is 2, 1, 0 (i.e., the spectral density of white noise is flat), -1, and -2, respectively, and the spectral slope of black noise is smaller than -2. The smaller the spectral slope in the frequency domain, the longer the memory of the signals represented as different colors of noise (excluding violet and blue noise). For example, a signal with its spectrum shown as white noise means the contribution to its variance is equal across all frequencies, while a signal with its spectrum shown as red noise means low-frequency periodic components dominate the contribution to its variance. Therefore, we use $SSM_{k\omega}$, $Pr_{k\omega}$, and $ET_{k\omega}$ to characterize the memory of SSM, Pr, and ET. The steps to get these variables can also be found in Figure 1.

2.7 Analysis of Differences between Models and Observational references

We evaluate two multimodel differences within CMIP5 compared to SMAP and ERA5 data: 1) differences in H_{SEP_n} and H_{EEP_n}; and 2) differences in $SSM_{k\omega}$, $ET_{k\omega}$, and $P_{k\omega}$, by subtracting observation-based data from model averages. In addition, we calculate the coefficient of variation across 14 models to show the degree of the statistical dispersion of the quantities.

The spatial resolution and the land cover between CMIP5 models and observational references (i.e., SMAP and ERA5), as well as among models themselves, are different. Here we re-grid all products with the same spatial resolution (36 km×36 km) and land cover as SMAP based on the nearest neighbor binning so they can be compared with each other (details on the spatial resolution projection see previous work (Xi et al., 2022)). In addition, we perform a significance test on these differences to avoid that multimodel differences in some regions may be caused by only a few models or even one model. This significance test is depicted on the maps using stippling, showing the regions that pass the 100% (i.e., all 14 models agree on the sign of average differences) and 75% (i.e., 11 of the 14 models agree on the sign of average differences) significance test. Since the variation of soil moisture in dry regions is usually very small (Koster et al., 2009), we remove regions with $\overline{SSM}_n$ less than 0.1 (shown as dark gray on the maps), where $\overline{SSM}_n$ is defined as the observational mean SSM after spatiotemporal normalization (Figure S3).

3 Results and Discussion

3.1 Temporal Variabilities of Soil Moisture, Precipitation, and Evapotranspiration from SMAP and ERA5 Data

The temporal variability of SSM (SSM_n) concentrates more in the seasonal to annual frequency band in most regions, with a smaller proportion in the two higher frequency bands, indicating that SSM has a large variability on time scales longer than the seasonal time scale (Figure 3a-3c).

The temporal variability of precipitation (Pr_n) shows different regional distributions over the three frequency bands (Figure 3d-3f). The variability is larger in the lowest frequency band for most tropical regions where the seasonal cycle can be large, and is larger in the highest frequency band for other regions, especially non-tropical regions. The reason is that, in most tropical regions, especially regions with tropical wet and dry climate, like Brazil, India, Northern Australia, and regions between the Sahara Desert and the equator in Africa, although the variation of temperature and radiation are small over a year, rainfall exhibits a strong seasonal cycle – the days with and without rainfall are concentrated so that the boundaries of the wet season and dry season are more distinct. So, precipitation in these regions shows a large seasonal variability. However, in tropical regions with a very wet climate, such as the Democratic Republic of the Congo, Indonesia, and the Philippines, there is no such seasonality because of the more steady rainfall pattern in these regions. On the other hand, there is not an obvious wet and dry season distinction for most non-tropical regions. The occurrence of rainfall is typically more random over a whole year and close to a white noise signal at high frequencies (Katul et al., 2007; Nakai et al., 2014). Therefore, precipitation variability in non-tropical regions is almost all high-frequency variability, except for regions with a Mediterranean climate and monsoonal regions where the monsoon distributes rainfall in a few months, imposing a strong seasonal cycle.

The largest temporal variability of ET (ET_n) in the lowest frequency band means that ET variability is large on time scales longer than seasonal over most regions (Figure 3g-3i), except in regions with a tropical wet climate. The reason is that ET in most regions is driven by either radiation or moisture limitation with high seasonality, except in the wet tropics where the seasonality of radiation and moisture is small but the daily variability can be large. In this way, the results in tropical wet regions, such as in the Amazon Rainforest, Africa's Equator, Indonesia, and the Philippines, are the opposite of other regions in terms of frequency distribution, showing ET variability concentrates on time scales shorter than monthly. This high-frequency radiation variability is mainly due to the variability of clouds on daily to weekly time scales which causes a large variability of ET on these short time scales (Anber et al., 2015). Moreover, this mechanism has the largest influence on regions near the equator because these regions receive more radiation than other regions over a year. Therefore, in these regions, ET variability is mostly located in the highest frequency band. In addition, ET in very dry regions (e.g., desert) does not have a clear seasonal cycle as well due to the strong limitation of moisture.

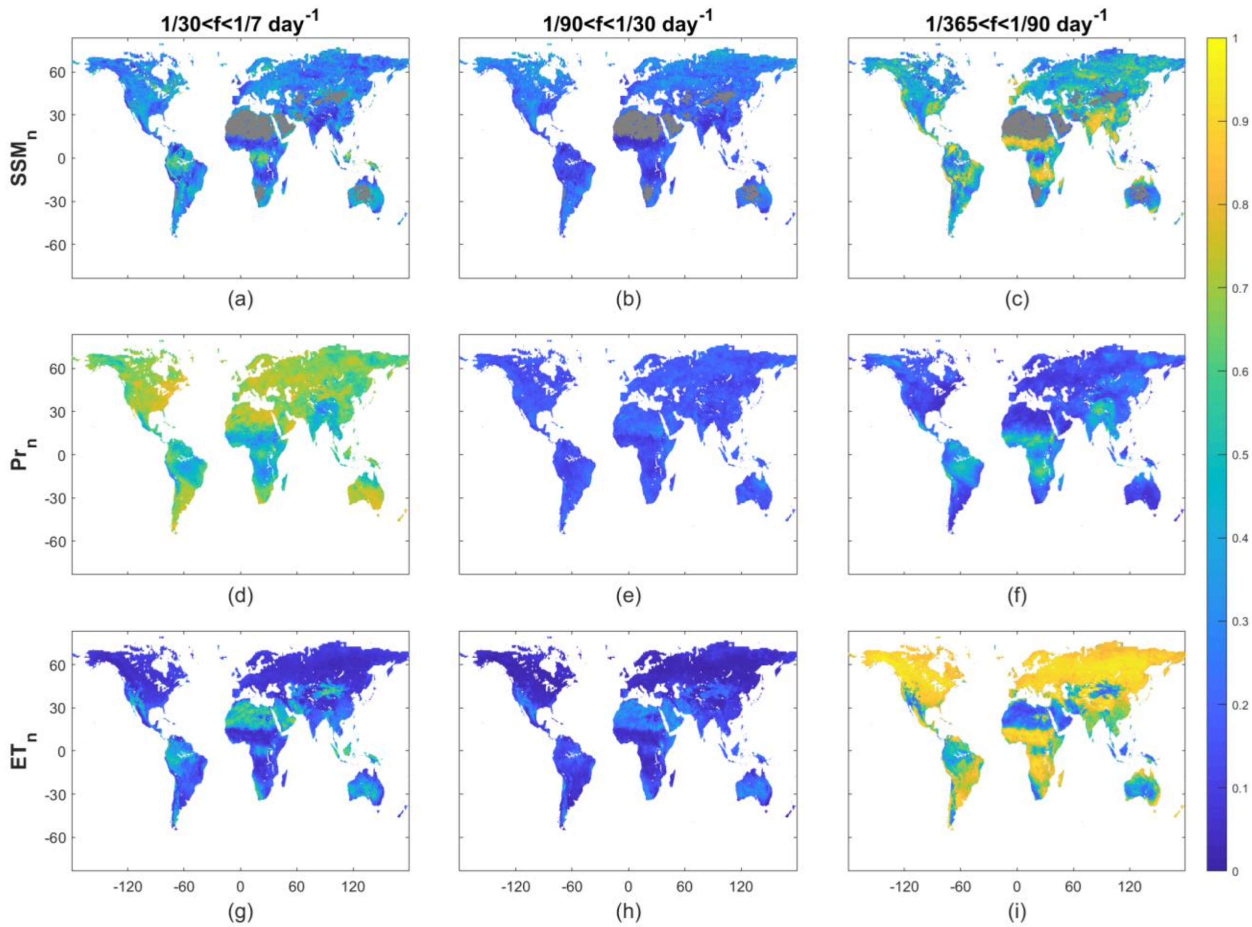

Figure 3. SSM_n (Figure a-c), Pr_n (Figure d-f), and ET_n (Figure g-i) based on SMAP and ERA5 data over the three frequency bands. SSM_n, Pr_n, and ET_n is the normalized variability of SSM, Pr, and ET, respectively, defined in section 2.5. Dark grey parts in Figure a-c are regions with $\overline{SSM}_n$ (observational mean SSM after spatiotemporal normalization) less than 0.1. For all subsequent results, including Figure 3, the three columns from left to right represent the weekly to monthly frequency band ($n = 1$), the monthly to seasonal frequency band ($n = 2$), and the seasonal to annual frequency band ($n = 3$).

The temporal variability of Pr and ET both show an apparent regional distribution (Figure 3). For Pr, the variability in tropical and non-tropical regions is opposite across the three frequency bands – the variability in tropical regions concentrates in the seasonal to annual frequency band, and the variability in non-tropical regions concentrates in the weekly to monthly frequency band. For ET, the variability in most regions concentrates in the seasonal to annual frequency band except for the dry regions and regions near the equator where the variability concentrates in the weekly to monthly frequency band. However, compared to Pr and ET, the temporal variability of SSM is more diverse spatially on a global scale.

Figure 4 shows the global distribution of H_{SEP_n} and H_{EEP_n} based on SMAP and ERA5 data over the three frequency bands (the corresponding values of H_{SEP_n} and H_{EEP_n} in each frequency band see Table S4). In the weekly to monthly frequency band, the total effect of ET and Pr on SSM variability is less than it in the other two frequency bands. Compared to Pr, which is the dominant driver of SSM variability in this frequency band, the fluctuation of ET has limited effects on SSM as ET is a slower process, in part regulated by soil moisture itself (Figures 4a and 4d). On time scales longer than monthly, ET and Pr together have more effects on SSM variability. In the monthly to seasonal frequency band where the total effect of ET and Pr on SSM reaches its largest magnitude, although the proportion of ET variability becomes larger, Pr is still the dominant factor of SSM variability (Figures 4b and 4e). In the seasonal to annual frequency band, the total variability of ET and Pr decreases but is still larger than it in the weekly to monthly frequency band. However, in this frequency band, ET becomes the dominant factor on SSM, especially in the middle and high latitudes. Therefore, Pr variability alone in these regions is no longer able to explain the SSM dynamics, and the seasonality of ET has to be considered (Figures 4c and 4f). Since H_{EEP_n} represents the proportion of ET variability to the total variability of ET and Pr, similar to ET_n shown in Figure 3, H_{EEP_n} patterns are different in tropical wet regions, where ET variability has more effects on SSM on the two higher frequency bands (Figures 4d and 4e), and Pr becomes the dominant factor on the lowest frequency band due to the strong seasonality in rainfall (Figure 4f).

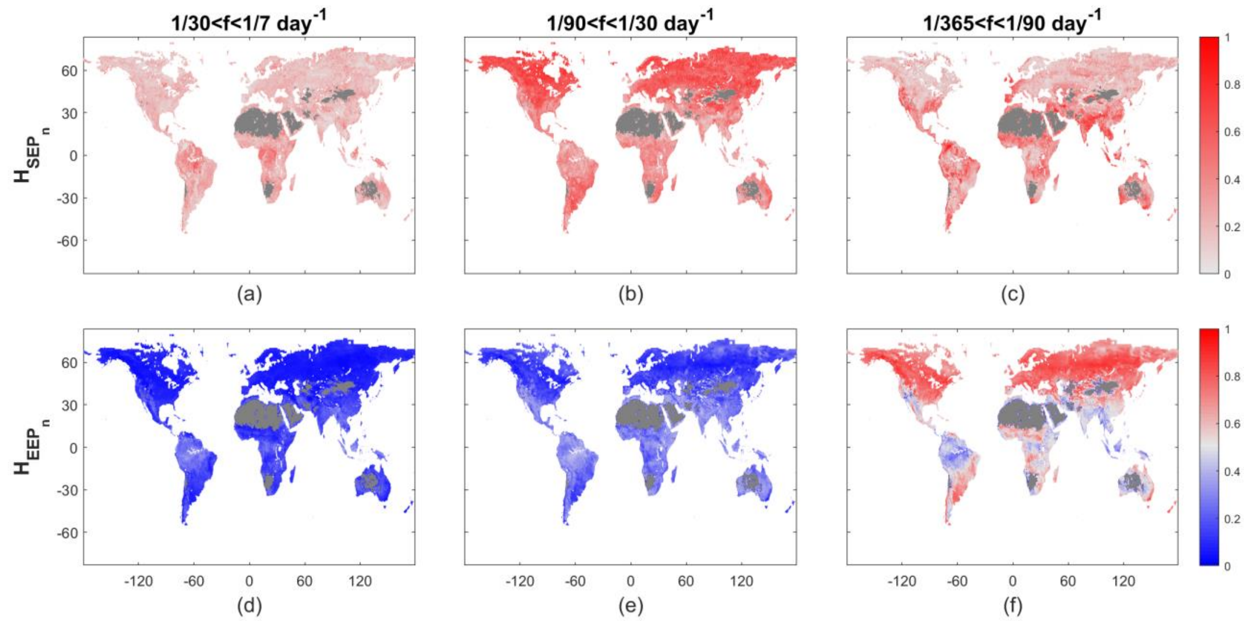

Figure 4. H_{SEP_n} (Figure a-c) and H_{EEP_n} (Figure d-f) based on SMAP and ERA5 data over the three frequency bands. H_{SEP_n} is the ratio of SSM_n to the sum of ET_n and Pr_n, and H_{EEP_n} is the ratio of ET_n to the sum of ET_n and Pr_n, defined in section 2.6. The values within each frequency band are normalized to between zero and one across the three frequency bands. Dark grey parts are regions with $\overline{SSM_n}$ less than 0.1.

To further identify the Pr and ET effects on SSM variability, we evaluate the relationships between their spectral slopes. Figure 5 shows the global distribution of SSM_{kw}, Pr_{kw}, and ET_{kw} expressed in terms of noise color in the three frequency bands based on SMAP and ERA5 data. We also evaluate the spectral slope of potential evaporation (PE_{kw}) from ERA5 to compare it with ET_{kw}.

From a previous study (Xi et al., 2022), we have found that the low-frequency periodic components dominate the contribution to the variance of SSM, and it has more randomness on time scales shorter than monthly and more memory on time scales longer than seasonality. From Figure 5a-5f, we further find that there is a phase shift between SSM and Pr spectra in the two higher frequency bands, especially the highest one, which implicates how Pr variability propagates into the soil moisture system (Katul et al., 2007). In the weekly to monthly frequency band where Pr is the dominant factor on SSM (according to Figure 4d), regions with smaller Pr_{kw} lead to SSM spectra decay more rapidly. In most regions where Pr is similar to a white noise, SSM exhibits a pink noise in the corresponding regions, indicating longer memory induced by soil moisture (Salvucci and Entekhabi, 1994). In regions where Pr exhibits a pink noise, like eastern Africa, Brazil, India, and northern Australia, SSM has a red noise spectrum (Figures 5a and 5d). A similar relationship between SSM and Pr spectra can also be found in the monthly to seasonal frequency band (Figures 5b and 5e), such as in southern North America, southern and north-central Asia, and regions around the Mediterranean, but it is not as evident as that in the highest frequency band since the effect of Pr on SSM variability decreases in this frequency band (according to Figure 4e). In the seasonal to annual frequency band, ET performs more effects on SSM variability than Pr for most regions (according to Figure 4f), so there are no strong correlations between Pr and SSM spectra. In previous studies, soil moisture was found to be similar to a red or black noise corresponding to precipitation having a white or pink noise at high frequency (Katul et al., 2007; Nakai et al., 2014). The SSM_{kw} here is a little larger (Figure 5a). A possible reason is the effect of runoff. Since the only function of runoff is to prevent large positive abnormalities in soil moisture, it may cause the time scale for soil moisture variability to shorten (Delworth and Manabe, 1988). This mechanism will mainly affect the soil moisture variability at high frequency and thus lead to less "redness" of soil moisture spectra.

Unlike between SSM_{kw} and Pr_{kw}, there is no such relationship between SSM_{kw} and ET_{kw}, even at the highest frequency band where ET is dominant on SSM variability (Figures 5a-5c and 5g-5i). It has been found that the sensitivity of soil moisture to precipitation and radiation uncertainty performs differently in seasonality (Wei et al., 2008). Here we also find that Pr and ET exert strong effects on SSM variability in different ways across different time scales. In previous studies, unlike Pr serving as a forcing term, ET was shown to be related to the damping term of soil moisture spectra (Delworth and Manabe, 1988; Katul et al., 2007; Nakai et al., 2014), which modulates potential evaporation (PE). The differences between ET_{kw} and PE_{kw} are mainly due to the variability of soil moisture. PE is an estimate of the maximum evaporation rate from a surface of pure water for given meteorological conditions (Delworth and Manabe, 1988). Weather fluctuations introduce a white or pink noise PE. However, unlike PE, ET is closely related to soil moisture, emphasizing that soil moisture limits and regulates the supply of moisture to the atmosphere on longer time scales. So the SSM dynamics influence ET spectra – leading to a more red noise than PE spectra because SSM has a longer memory. This influence is especially more visible in dry regions. The reason is that, compared to SSM in dry regions, SSM in wet regions mostly tracks the variability of PE. So ET in wet regions will not be strongly affected by SSM variability and thus still shows pink noise. On longer time scales, both ET and PE show obvious seasonality that the low-frequency periodic components dominate the contribution to the variance of signals (Figures 5i and 5l).

To summarize, the effects of Pr and ET on SSM variability are different across time scales. In the two higher frequency bands (especially the weekly to monthly frequency band), Pr, acting as a forcing by averaging the large oscillations to limit high-frequency components, has more effects on SSM variability. In the seasonal to annual frequency band, ET, acting as the dissipative process that prevents SSM anomalies from persisting indefinitely, has more effects on SSM variability.

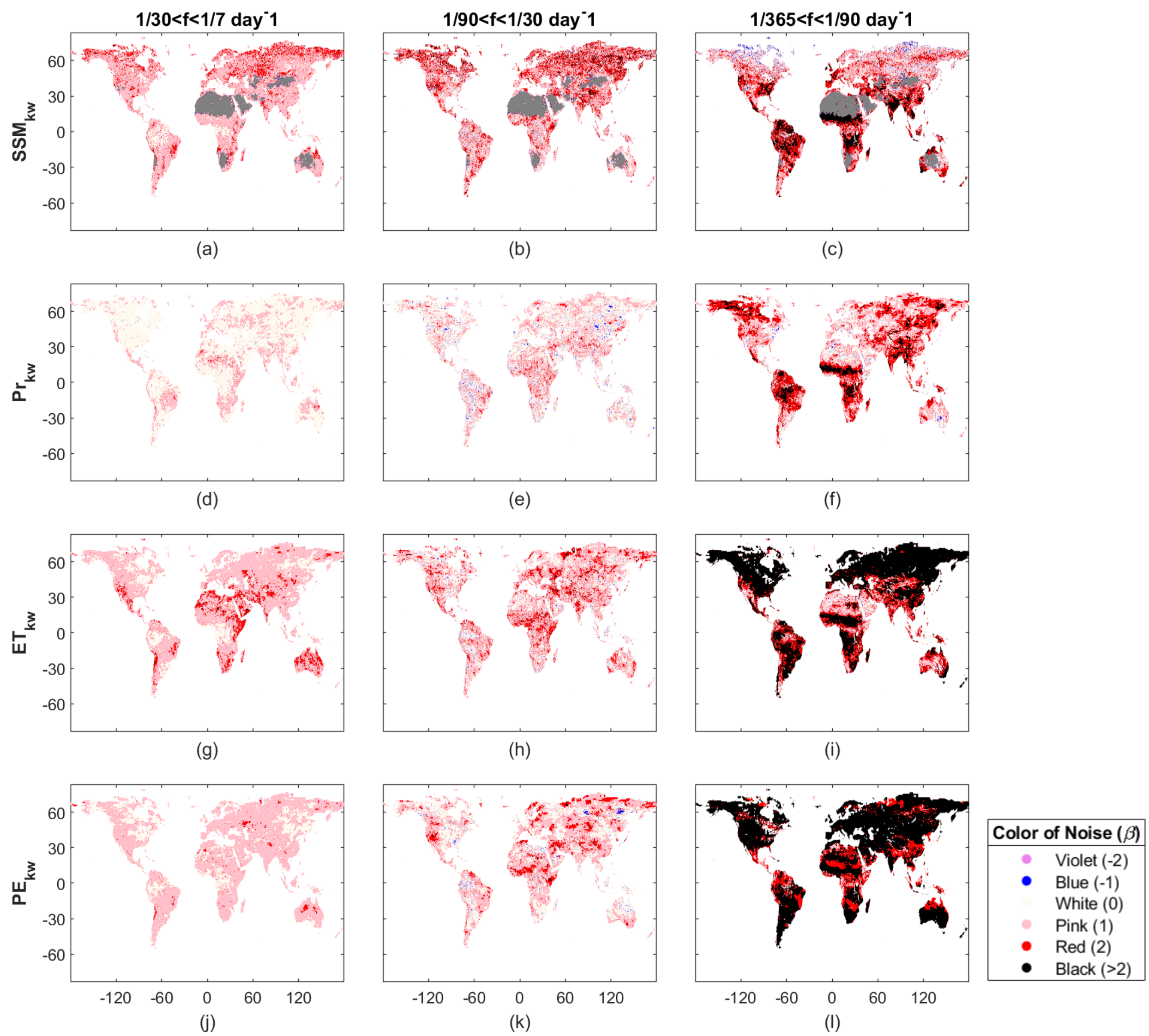

Figure 5. Noise color of SSM (Figure a-c), Pr (Figure d-f), ET (Figure g-i), and PE (Figure j-l) over the three frequency bands according to SSM_{kw}, Pr_{kw}, ET_{kw}, and PE_{kw} based on SMAP and ERA5 data. The colors in each figure represent the corresponding color of noise, referring to the power spectra of SSM, Pr, ET, and PE. The legend shows the color referring to each noise, and the number in brackets is the inverse number of the spectral slope of power-law noise corresponding to each noise color.

3.2 Comparison between CMIP5 simulations and SMAP and ERA5 references

Figure 6a-6c shows the average differences for H_{SEP_n} and H_{EEP_n} of model simulations within CMIP5 compared to SMAP and ERA5 data. A significance test is performed and depicted using stippling. Here, the “+” stippling means the region passes a 100% significance test, and the “.” stippling means the region passes a 75% significance test. Therefore, we only focus on the regions with stippling. For most regions, the multimodel differences of ${H_{SEP}}_n$ are negative in the two higher frequency bands and they are positive in the lowest frequency band, which means that the CMIP5 simulations of the total effect of ET and Pr on SSM variability are smaller on time scales shorter than seasonal and are larger on time scales longer than seasonal, compared to SMAP and ERA5 data (Figure 6a-6c). The average difference of ${H_{SEP}}_n$ is largest in the monthly to seasonal frequency band (-0.6792 and -0.4492 with 100% and 75% significance) and smallest in the weekly to monthly frequency band (-0.3365 and -0.2871 with 100% and 75% significance) (Table 1). For all three frequency bands, the average differences of ${H_{SEP}}_n$ are larger in Central and Northern North America, Central and Eastern Europe, and regions near the equator.

Significance	100% significance test			75% significance test		
Frequency band (day^{-1})	1/7 ~ 1/30	1/30 ~ 1/90	1/90 ~ 1/365	1/7 ~ 1/30	1/30 ~ 1/90	1/90 ~ 1/365
BCC-CSM1.1	-0.2755	-0.5409	0.5554	-0.2249	-0.2797	0.4720
BNU-ESM	-0.3221	-0.6096	0.4718	-0.2740	-0.4001	0.4002
CanESM2	-0.3565	-0.7277	0.5245	-0.3057	-0.4523	0.4311
CNRM-CM5	-0.3323	-0.8718	0.5380	-0.2808	-0.6494	0.4492
CSIRO-Mk3.6	-0.3695	-0.8166	0.4851	-0.3220	-0.5809	0.4208
GFDL-CM3	-0.3262	-0.6595	0.4189	-0.2785	-0.4436	0.3417
GFDL-ESM2G	-0.3243	-0.6493	0.4494	-0.2762	-0.4388	0.3693
GFDL-ESM2M	-0.3250	-0.6566	0.4615	-0.2767	-0.4423	0.3796
MIROC5	-0.3330	-0.5357	0.3897	-0.2846	-0.2995	0.3119
MIROC-ESM	-0.3374	-0.6061	0.4011	-0.2884	-0.3794	0.3257
MIROC-ESM-CHEM	-0.3383	-0.6064	0.4021	-0.2893	-0.3814	0.3271

MRI-CGCM3	-0.3812	-0.8297	0.5968	-0.3321	-0.5989	0.5227
MRI-ESM1	-0.3804	-0.8293	0.5971	-0.3315	-0.5974	0.5225
NorESM1-M	-0.3089	-0.5702	0.4245	-0.2547	-0.3453	0.3421
Average (± 1 SD)	-0.3365 ± 0.0274	-0.6792 ± 0.1110	0.4797 ± 0.0692	-0.2871 ± 0.0280	-0.4492 ± 0.1121	0.4011 ± 0.0685
Observation	0.4734	1.1492	0.4161	0.4305	0.9550	0.4741

Table 1. Observational and multimodel differences of H_{SEP_n} within CMIP5. The observational H_{SEP_n} here is the original value without normalization across the three frequency bands.

From section 3.1, we know that Pr dominates SSM variability in the two higher frequency bands, and ET dominates it in the seasonal to annual frequency band. From Figure 6e-6f, we find that in each frequency band, the effect of the corresponding dominant factor (i.e., Pr or ET) on SSM simulated within the CMIP5 models tends to be smaller than that from ERA5 data. Specifically, in the two higher frequency bands where Pr is the dominant factor, models overestimate the proportion of ET variability to the total variability of ET and Pr. Thus, the effect of Pr on SSM is underestimated by models (Figure 6d-6e). In the lowest frequency band where ET is the dominant factor, models underestimate the effects of ET on SSM. Unlike H_{SEP_n}, the multimodel difference of H_{EEP_n} is largest in the weekly to monthly frequency band (-0.1259 and -0.0770 with 100% and 75% significance) and smallest in the monthly to seasonal frequency band (-0.0677 and -0.0515 with 100% and 75% significance) (Table 2). From Figure 6 (also Table 1 and 2), CMIP5 simulations show larger differences on H_{SEP_n} than H_{EEP_n}, which means that these CMIP5 models perform relatively well in capturing the proportion of ET and Pr variability to their total variability, while they exhibit larger differences in simulating the total effect of ET and Pr on SSM variability compared to SMAP and ERA5 data.

Significance	100% significance test	75% significance test

Frequency band (day^{-1})	1/7 ~ 1/30	1/30 ~ 1/90	1/90 ~ 1/365	1/7 ~ 1/30	1/30 ~ 1/90	1/90 ~ 1/365
BCC-CSM1.1	0.1826	0.1089	-0.1059	0.1374	0.0992	-0.0579
BNU-ESM	0.1818	0.0927	-0.1414	0.1303	0.0751	-0.1141
CanESM2	0.1119	0.1245	-0.0800	0.0675	0.1146	-0.0524
CNRM-CM5	0.0891	0.0235	-0.0732	0.0493	0.0089	-0.0402
CSIRO-Mk3.6	0.0863	0.0575	-0.0793	0.0372	0.0386	-0.0517
GFDL-CM3	0.1216	0.0656	-0.0959	0.0738	0.0519	-0.0693
GFDL-ESM2G	0.1654	0.0988	-0.1087	0.1169	0.0866	-0.0833
GFDL-ESM2M	0.1637	0.0992	-0.1061	0.1133	0.0871	-0.0784
MIROC5	0.0778	0.0028	-0.0553	0.0159	-0.0272	-0.0249
MIROC-ESM	0.1110	0.0429	-0.0743	0.0542	0.0218	-0.0523
MIROC-ESM-CHEM	0.1100	0.0418	-0.0750	0.0527	0.0215	-0.0524
MRI-CGCM3	0.1126	0.0700	-0.0633	0.0715	0.0542	-0.0407
MRI-ESM1	0.1109	0.0711	-0.0624	0.0699	0.0550	-0.0392
NorESM1-M	0.1386	0.0492	-0.1003	0.0880	0.0336	-0.0792
Average (± 1 SD)	0.1259 ± 0.0336	0.0677 ± 0.0332	-0.0872 ± 0.0227	0.0770 ± 0.0347	0.0515 ± 0.0374	-0.0597 ± 0.0222
Observation	0.2406	0.4182	0.7747	0.2236	0.3756	0.7618

Table 2. Observational and multimodel differences of H_{EEP_n} within CMIP5. The observational H_{EEP_n} here is the original value without normalization across the three frequency bands.

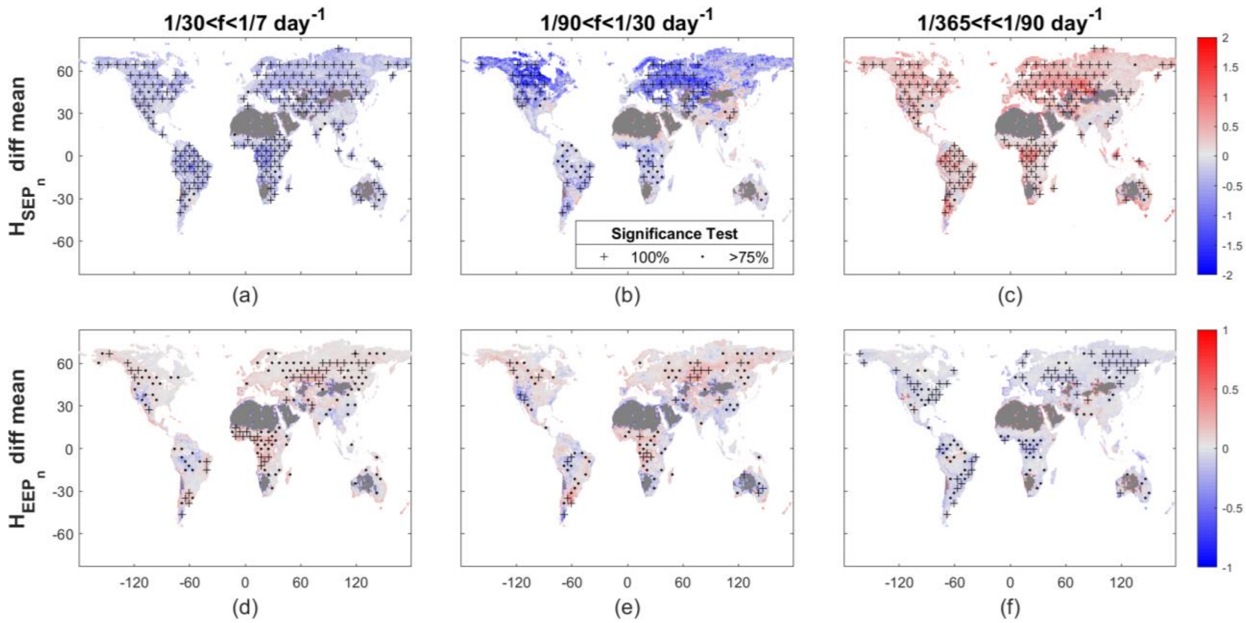

Figure 6. Average differences of H_{SEP_n} (Figure a-c) and H_{EEP_n} (Figure d-f) between CMIP5 models and the observation-based data in the three frequency bands. Dark grey parts are regions with $\overline{SSM_n}$ less than 0.1. For each figure, "+" and "." stippling represents the region that passes a 100% significance test and a 75% significance test, respectively.

In addition to multimodel differences compared to SMAP and ERA5 data, the coefficient of variation (CV) of H_{SEP_n} and H_{EEP_n} across models are also investigated to estimate their statistical variance (Figure S4). For both H_{SEP_n} and H_{EEP_n}, the intermodel spread is larger in the weekly to monthly and monthly to seasonal frequency band and smaller in the seasonal to annual frequency band (also see Table S6). Therefore, for CMIP5 estimations of Pr and ET effect on SSM variability, there is a more extensive intermodel spread on time scales shorter than seasonal and a lower variance among models on time scales longer than seasonal time scale, suggesting an individual deficiency in representing the short-term variability and a systematic deficiency of these CMIP5 models in representing the long-term variability.

The multimodel differences of H_{SEP_n} and H_{EEP_n} are further analyzed with the mean SSM on a global scale. To make a trade-off between high significance and the size of samplings, we use the differences that pass a 75% significance test. Figure S3 shows the global distribution of the mean SMAP SSM after spatiotemporal normalization ($\overline{SSM}_n$). For H_{SEP_n} (Figure 7a-7c), models achieve their best estimates in transitional zones between dry and wet climates, where there is both a strong coupling between soil moisture and Pr (Koster et al., 2004) as well as between soil moisture and ET (Seneviratne et al., 2010). No matter whether $\overline{SSM}_n$ increases or decreases from the intermediate transitional zones, the differences of H_{SEP_n} increase. Therefore, when considering the Pr and ET effect on SSM variability, the CMIP5 models can perform better in regions with strong coupling between these variables, and the differences compared to observation-based data tend to be more apparent in wet and dry regions where interactions are weaker. This finding is particularly evident in the highest and lowest frequency bands where observation-based H_{SEP_n} is smaller. On the other hand, from Figure 7d-7f, H_{EEP_n} differences basically increase with the decrease of $\overline{SSM}_n$ except for extremely dry regions, indicating that the CMIP5 models have difficulties in estimating the interaction between Pr and ET in regions with less soil moisture. When soil moisture is limited, ET is also limited, although sensitive to SSM. Under this condition, ET variation is too small to impact climate variability, and the impact of Pr variation on climate variability is almost independent on SSM as drier soils will lead to lower precipitation likelihood (Seneviratne et al., 2010). Therefore, it is hard for models to capture correct interactions between Pr and ET, shown as larger differences of H_{EEP_n} in drier regions. In regions where $\overline{SSM}_n$ is extremely low (less than 10%), models tend to correctly capture the proportion of Pr and ET variability.

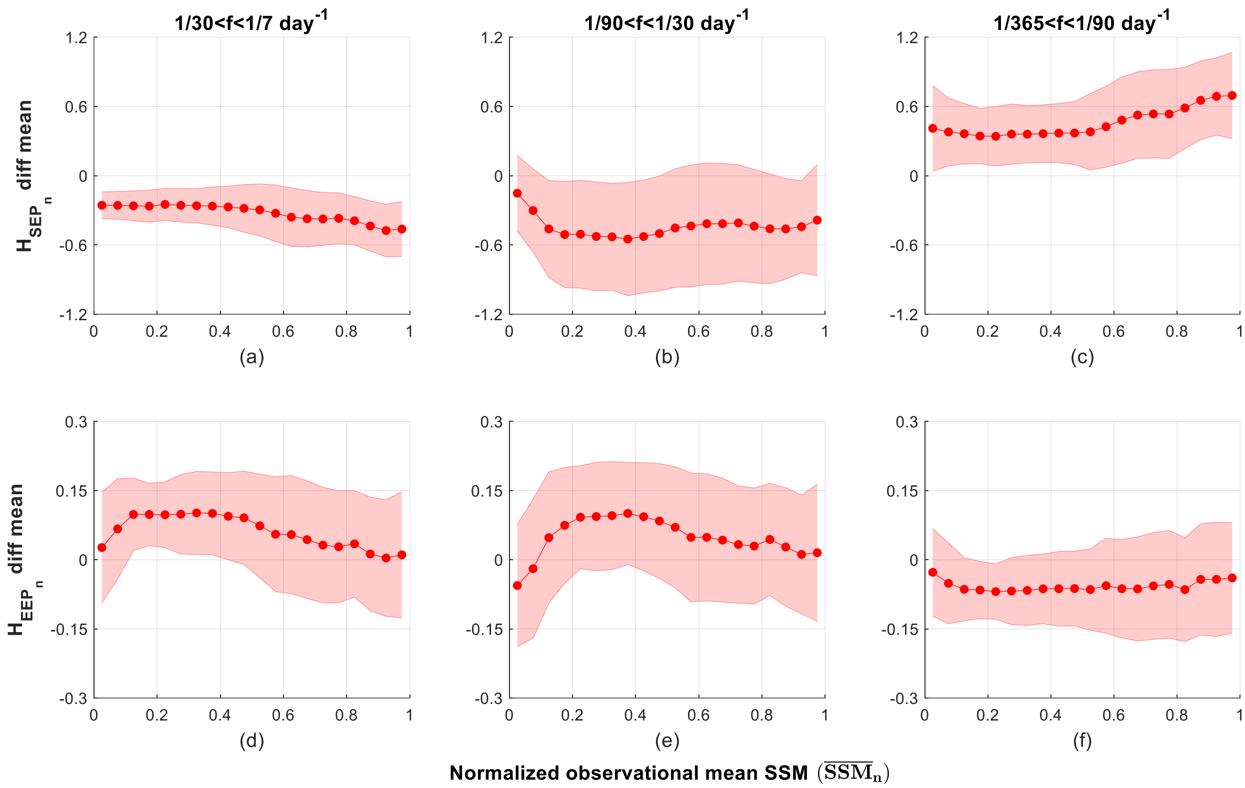

Figure 7. Comparison of average differences of H_{SEP_n} (Figure a-c) and H_{EEP_n} (Figure d-f) between CMIP5 models and observation-based data with $\overline{SSM}_n$ in the three frequency bands. The red shading represents $+/-$ one standard deviation. $\overline{SSM}_n$ is separated into 20 bins of equal size (i.e., 0.05 for each bin), then the mean of H_{SEP_n} and H_{EEP_n} differences located in each bin (corresponding to the range of $\overline{SSM}_n$) were separately calculated for each frequency band. Differences in this figure are the values passing a 75% significance test. All values in the regions with $\overline{SSM}_n$ less than 0.1 are removed.

Finally, we evaluate multimodel differences of the spectral slopes ($SSM_{k\omega}$, $Pr_{k\omega}$, and $ET_{k\omega}$) compared to the SMAP and ERA5 data (Figure 8). Negative differences mean that modeled spectra decay more rapidly and vice versa. Compared to SMAP SSM spectra, CMIP5 SSM spectra decay more rapidly in the two higher frequency bands and less rapidly in the seasonal to annual frequency band in most regions (Figure 8a-8c), indicating that these models underestimate the short-term variability and overestimate the long-term variability of SSM in a non-linear way. For $Pr_{k\omega}$ (Figure 8d-8f) and $ET_{k\omega}$ (Figure 8g-8i), positive differences with high significance in most regions indicate that CMIP5 models underestimate their memory, implying land surface models may not be able to reproduce the correct intensity of Pr and ET variability, especially on time scales longer than seasonal. Our findings are aligned with previous studies (Katul et al., 2007; McColl et al., 2019; Nakai et al., 2014) but with different methods and models. We also find the differences characterizing the memory are not the same across frequencies and are most prominent in the seasonal to annual frequency band. This again suggests that models exhibit deficiency in representing long-term transpiration and soil moisture dynamics.

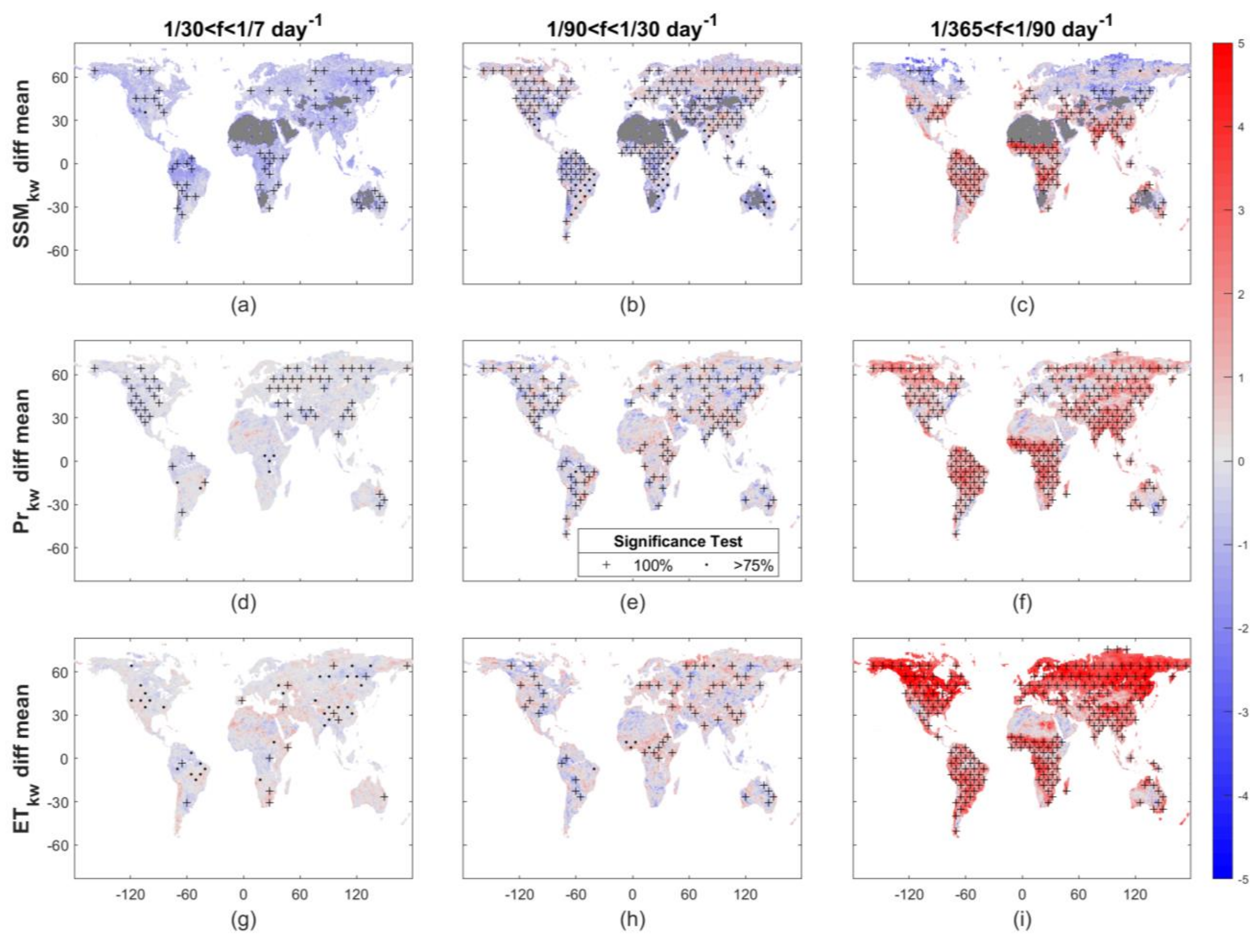

Figure 8. Average differences of $SSM_{k\omega}$ (Figure a-c), $Pr_{k\omega}$ (Figure d-f), and $ET_{k\omega}$ (Figure g-i) between CMIP5 models and the observation-based data in the three frequency bands. Dark grey parts in Figure a-c are regions with $\overline{SSM}_n$ less than 0.1. For each subfigure, "+" and "." stippling represents the region that passes a 100% significance test and a 75% significance test, respectively.

3.3 Uncertainty analysis

Two parts during the data processing could introduce uncertainties to our analysis in this study. First, since the SMAP data is non-continuous on the daily time scale, we fill the missing values before performing Fourier analysis. The gap-filling process is the same as our previous analysis and has been carefully validated using in-situ soil moisture data from International Soil Moisture Network (Xi et al., 2022). Second, the interpolation during intermodel computation may induce uncertainties since the spatial resolution of all these CMIP5 models are much coarser than the "standard" spatial resolution (36 km×36 km, see section 2.7). Apart from comparing the re-gridded results with an intermediate resolution (1°×1°) and finding that the differences are very small (Xi et al., 2022), we also conduct a significance test to constrain the potential uncertainties as much as possible. All statistical correlation analyses in this study are based on the multimodel differences passing a high significance test (more than 75%), ensuring that a systematic performance in land surface models is shown.

Apart from these technical issues, some other aspects may also cause uncertainties. One issue is related to the potential biases of the SMAP data. This study conducts comparative evaluations of ESMs within CMIP5 and uses SMAP products as the observations of SSM. However, even though SMAP meets its performance target and has better performance than other satellite products, its retrievals have been shown to exhibit potential errors in heavily vegetated areas such as forests, with the presence of water bodies, and in frozen soil such as in the Arctic tundra environment (Entekhabi et al., 2014; McColl et al., 2017; Wrona et al., 2017). Therefore, when performing comparative assessments with model simulations, the biases in SMAP data themselves should also be taken into consideration.

Another issue is the linear and time-invariant assumption of the interactions among SSM, Pr, and ET. In this study, we assumed an LTI system of SSM, Pr, and ET and then performed the Fourier analysis based on it. However, the relationships among them may not be linear and time-invariant. For example, in regions with plenty of vegetation, precipitation is first intercepted by the canopy, and then throughfall is further partitioned into surface runoff and infiltration water, which directly affects SSM instead of precipitation. A previous study has also shown that there is a higher linear relationship between soil moisture and precipitation in less-vegetated regions (Sehler et al., 2019). Snow is another factor related to this issue. When the precipitation is snow, it will not interact with SSM immediately. Instead, there is a snow accumulation and melting process, which could take days, weeks, and even months. Thus, the relationship between SSM and Pr may not be time-invariant in high-latitude regions.

Although we have mentioned that there are many complex physical processes involved in the effects on SSM dynamics, and this study aims to only focus on the two elementary variables related to SSM (i.e., Pr and ET), we still want to try to analyze the uncertainties from this aspect and see how much confidence we could have under this LTI assumption. A feasible first step is to mask the regions that could be most affected by these issues and see how the results will change. This approach can also be used to quantify the uncertainties induced by SMAP data mentioned above. We identify the regions with potential uncertainties as dense vegetation cover (vegetation water content > 5 kg/m^2), frozen landscapes (surface temperature < 0°C), and the presence of water bodies (water body fraction > 5% coverage of a pixel) (see Figure S5), which is similar to a previous study (McColl et al., 2017). Then, we recalculate the observation-based H_{SEP_n} and H_{EEP_n}, multimodel differences of H_{SEP_n} and H_{EEP_n}, and the CV of H_{SEP_n} and H_{EEP_n} across the CMIP5 models with these regions being masked (Table S4, S5, and S6). We find that, although being quantitively inconsistent, these results are all qualitative across the three frequency bands, illustrating the feasibility of our analysis on a global scale.

4 Conclusions

This study uses satellite-based observations to evaluate 14 Earth system models within CMIP5 in simulating the effects of Pr and ET on SSM variability across three frequency bands. We find that these models generally underestimate the total effects of Pr and ET on SSM in the high-frequency bands (weekly to monthly and monthly to seasonal) and overestimate it in the low-frequency band (seasonal to annual). Additionally, based on the findings that Pr dominates weekly to seasonal SSM variability and ET dominates seasonal to annual SSM variability, these models underestimate the effects on SSM by Pr or ET that is a dominant factor in each frequency band. Across the three frequency bands, models perform better estimations in regions with strong land-atmosphere interactions between the three variables. For the metrics investigated here, models show an individual deficiency in representing short-term variability and a systematic deficiency of long-term variability.

This study also identifies systematic metrics that can be used to assess model performance and help refine process representation across time scales. Our results highlight that the Earth system models within CMIP5 should improve their representation of precipitation and evapotranspiration effects in modeling soil moisture.

Appendix A: Conceptual LTI systems representing the ET and Pr effects on SSM variability.

A transfer function (also known as system function) (Haykin and Van Veen, 2007) mathematically represents the relationship between the input and output of a system (black-box model). It can usually be used to describe the relationship between the signal excitation and response of a linear time-invariant (LTI) system (Phillips et al., 2003) with the time-frequency as a variable. For an LTI system, even if its specific structure and parameters are not known, its model in the frequency domain can be regarded as a rational polynomial form. Then the properties of the system can be determined by analyzing the input and output of the system.

LTI systems are subject to constraints of linearity and time invariance. The constraint of linearity means that when multiple excitation signals act on the LTI system simultaneously, the total response is equal to the sum of the corresponding individual effects of each excitation. Besides, when the excitation increases by a specific multiple, the response also increases by the same multiples. The constraint of time-invariance means that the response of the LTI system is independent of the time the excitation acts on the system. This means that, regardless of the time sequence of the input signal acting on the system, the output signals are the same. The only difference is the time of their appearances. The constraint of linearity on the LTI system can be expressed as:

$$T[ax_1(n) + bx_2(n)] = ay_1(n) + by_2(n) \tag{A1}$$

where T represents the computational relationship of the system, $x_1(n)$, $y_1(n)$ and $x_2(n)$, $y_2(n)$ are two pairs of excitation and response, respectively. Besides, the constraint of time-invariance on the LTI system can be expressed as:

$$y(n-m) = T[x(n-m)] \tag{A2}$$

which means that when the excitation is delayed for a period of time m, the corresponding response is also delayed for time m.

Subject to the constraints of linearity and time-invariance, if the signal applied to the LTI system is decomposed (as an impulse signal), the response caused by the original excitation signal is obtained by summing the responses generated by each component acting on the system. In this way, the LTI system produces an output signal from any input signal, which can be expressed as (considering the default system as a causal system):

$$y(t) = \int_0^{+\infty} x(\tau)h(t-\tau) = x(t) \otimes h(t) \tag{A3}$$

where $x(t)$ and $y(t)$ are the input and output of the LTI system, respectively, and $h(t)$ is the transfer function of the LTI system. According to the Convolution theorem, the convolution of two signals in the time domain is equivalent to multiplying their corresponding spectra in the frequency domain:

$$x(t) \otimes h(t) = X(k) \cdot H(k) \tag{A4}$$

where $X(k)$ and $H(k)$ are the spectra of $x(t)$ and $h(t)$, respectively.

In the time domain, the terrestrial water balance can be simply expressed as:

$$\frac{\mathrm{d}ssm(t)}{\mathrm{d}t} = pr(t) - et(t) - q(t) \tag{A5}$$

where ssm is surface soil moisture, pr is precipitation, et is evapotranspiration, and q is drainage and runoff. Neglecting drainage and runoff ($q = 0$), this water balance can be further simplified as:

$$\frac{\mathrm{d}ssm(t)}{\mathrm{d}t} = pr(t) - et(t) \tag{A6}$$

where precipitation is the climate input to soil moisture, and evapotranspiration is the water losses relative to soil moisture.

Any system that can be simulated as homogeneous linear differential equations with constant coefficients can be regarded as an LTI system. In this way, the relationships between SSM, ET, and Pr can be described assuming two conceptual LTI systems, where the inputs are $et(t)$ and $pr(t)$, respectively, and the outputs are both $ssm(t)$ (Figure S2). Since the two systems are both single-input and single-output (SISO) systems (Partington, 2004), we can focus on the relationship between their excitations and responses without caring about the internal variations of the systems. In this way, the relationships between excitation and response of the two LTI systems can be expressed as:

$$ssm(t) = et(t) \otimes h_{se}(t) \tag{A7}$$

$$ssm(t) = pr(t) \otimes h_{sp}(t) \tag{A8}$$

where $h_{se}(t)$ and $h_{sp}(t)$ are the transfer function of the "ET-SSM" LTI system (Figure S2(a)) and "Pr-SSM" LTI system (Figure S2(b)), respectively.

It is hard to investigate these two transfer functions in the time domain. However, by applying the convolution operator, equations (A4) and (A5) in the time domain can be converted into the frequency domain as a product:

$$F_{SSM}(k) = F_{ET}(k) \cdot H_{SE}(k) \tag{A9}$$

$$F_{SSM}(k) = F_{Pr}(k) \cdot H_{SP}(k) \tag{A10}$$

where $H_{SE}(k)$ and $H_{SP}(k)$ are the Fourier transforms of the transfer functions $h_{se}(t)$ and $h_{sp}(t)$, respectively. The two LTI systems change the spectra of the input signal by weighting each of its frequency components. This change is completely determined by the transfer functions $H_{SE}(k)$ and $H_{SP}(k)$, which serve as a weighting function transforming the excitation with the spectrum of $F_{ET}(k)$ and $F_{Pr}(k)$ into the response with the spectrum of $F_{SSM}(k)$.

Assuming the input of the two systems is a power signal (i.e., signal power is finite), equations (A6) and (A7) can be read in terms of the power spectrum as:

$$E_{SSM}(k) = E_{ET}(k) \cdot |H_{SE}(k)|^2 \quad \text{(A11)}$$

$$E_{SSM}(k) = E_{Pr}(k) \cdot |H_{SP}(k)|^2 \quad \text{(A12)}$$

In this way, the effects of ET and Pr variability on SSM variability can be identified by $|H_{SE}(k)|^2$ and $|H_{SP}(k)|^2$, respectively.

To consider both ET and Pr effects on SSM, subject to linearity constrains, the two conceptual LTI systems can be combined as (Figure S2(c)):

$$ssm(t) = et(t) \otimes h_{se}(t) + pr(t) \otimes h_{sp}(t) \quad \text{(A13)}$$

If we use an identical transfer function to replace the two cascaded transfer functions as the internal mechanism to capture the total effects of ET and Pr on SSM:

$$ssm(t) = et(t) \otimes h_{sep}(t) + pr(t) \otimes h_{sep}(t) \quad \text{(A14)}$$

where $h_{sep}(t)$ is the transfer function of the LTI system shown in Figure 2 and can be performed by spectral analysis as the two SISO systems discussed above.

Acknowledgments

The land surface models from CMIP5 were available online (https://esgf-node.llnl.gov/). SMAP surface soil moisture data can be obtained on https://earthdata.nasa.gov/. ERA5 data can be obtained on https://cds.climate.copernicus.eu/. Gentine acknowledges funding from NASA 80NSSC18K0998. Zhuang was funded by NASA through a subcontract from JPL #1609311. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table S1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides

coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Data Availability Statement

The codes and data for analysis in this study are available at https://purr.purdue.edu/publications/3999/1.

References

Anber, U., Gentine, P., Wang, S., & Sobel, A. H. (2015). Fog and rain in the Amazon. *Proceedings of the National Academy of Sciences*, *112*(37), 11473-11477.

Akbar, R., Short Gianotti, D., McColl, K. A., Haghighi, E., Salvucci, G. D., & Entekhabi, D. (2018). Hydrological storage length scales represented by remote sensing estimates of soil moisture and precipitation. *Water Resources Research*, *54*(3), 1476-1492.

Bailly-Comte, V., Jourde, H., Roesch, A., Pistre, S., & Batiot-Guilhe, C. (2008). Time series analyses for Karst/River interactions assessment: Case of the Coulazou river (southern France). *Journal of hydrology*, *349*(1-2), 98-114.

Berg, A., & Sheffield, J. (2018). Soil Moisture–Evapotranspiration Coupling in CMIP5 Models: Relationship with Simulated Climate and Projections. *Journal of Climate*, *31*(12), 4865-4878.

Bonan, G. B. (1996). *Land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and users guide. Technical note* (No. PB-97-131494/XAB; NCAR/TN-417-STR). National Center for Atmospheric Research, Boulder, CO

(United States). Climate and Global Dynamics Div..

Bourke, P. (1998). Generating noise with different power spectra laws. *accessed October*, *26*, 2006.

Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., & Savoie, M. H. (2012). EASE-Grid 2.0: Incremental but significant improvements for Earth-gridded data sets. *ISPRS International Journal of Geo-Information*, *1*(1), 32-45.

Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., ... & Yueh, S. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4994-5007.

Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., ... & Kerr, Y. (2018). Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sensing of Environment*, *204*, 931-941.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., & Aida, K. (2018). Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sensing of Environment*, *214*, 1-13.

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., ... & Asanuma, J. (2017). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment*, *191*, 215-231.

Colliander, A., Reichle, R., Crow, W., Cosh, M. H., Chen, F., Chan, S. K., ... & Yueh, S. H. (2021). Validation of soil moisture data products from the NASA SMAP mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data

Store (CDS), *date of access*. https://cds.climate.copernicus.eu/cdsapp#!/home

Delworth, T. L., & Manabe, S. (1988). The influence of potential evaporation on the variabilities of simulated soil wetness and climate. *Journal of Climate*, *1*(5), 523-547.

Dirmeyer, P. A., Jin, Y., Singh, B., & Yan, X. (2013). Trends in land–atmosphere interactions from CMIP5 simulations. *Journal of Hydrometeorology*, *14*(3), 829-849.

Dong, J., Dirmeyer, P. A., Lei, F., Anderson, M. C., Holmes, T. R., Hain, C., & Crow, W. T. (2020). Soil evaporation stress determines soil moisture-evapotranspiration coupling strength in land surface modeling. *Geophysical Research Letters*, *47*(21), e2020GL090391.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., ... & Kimball, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704-716.

Entekhabi, D., Yueh, S., & De Lannoy, G. (2014). SMAP handbook.

Flexas, M. M., Thompson, A. F., Torres, H. S., Klein, P., Farrar, J. T., Zhang, H., & Menemenlis, D. (2019). Global estimates of the power transfer from the wind to the ocean, with emphasis on near-inertial oscillations. *Journal of Geophysical Research: Oceans*.

Ford, T. W., Harris, E., & Quiring, S. M. (2014). Estimating root zone soil moisture using near-surface observations from SMOS. *1foldr Import 2019-10-08 Batch 11*.

Ghannam, K., Nakai, T., Paschalis, A., Oishi, C. A., Kotani, A., Igarashi, Y., ... & Katul, G. G. (2016). Persistence and memory timescales in root-zone soil moisture dynamics. *Water Resources Research*, *52*(2), 1427-1445.

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., & Gentine, P. (2019). Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature*, *565*(7740), 476.

Guillod, B. P., Orlowsky, B., Miralles, D. G., Teuling, A. J., & Seneviratne, S. I. (2015). Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nature communications*, *6*, 6443.

Guo, Z., Dirmeyer, P. A., Koster, R. D., Sud, Y. C., Bonan, G., Oleson, K. W., ... & Xue, Y. (2006). GLACE: the global land–atmosphere coupling experiment. Part II: analysis. *Journal of Hydrometeorology*, *7*(4), 611-625.

Haykin, S., & Van Veen, B. (2007). *Signals and systems*. John Wiley & Sons.

Houska, T., Kraft, P., Chamorro-Chavez, A., & Breuer, L. (2015). SPOTting model parameters using a ready-made python package. *PloS one*, *10*(12), e0145180.

Jiao, D., Xu, N., Yang, F., & Xu, K. (2021). Evaluation of spatial-temporal variation performance of ERA5 precipitation data in China. *Scientific Reports*, *11*(1), 1-13.

Katul, G. G., Porporato, A., Daly, E., Oishi, A. C., Kim, H. S., Stoy, P. C., ... & Siqueira, M. B. (2007). On the spectrum of soil moisture from hourly to interannual scales. *Water Resources Research*, *43*(5).

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., ... & Liu, P. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138-1140.

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J. (2009). On the nature of soil moisture in land surface models. *Journal of Climate*, *22*(16), 4322-4335.

Kumar, S. V., Dirmeyer, P. A., Peters-Lidard, C. D., Bindlish, R., & Bolten, J. (2018). Information theoretic evaluation of satellite soil moisture retrievals. *Remote sensing of environment*, *204*, 392-400.

Levine, P. A., Randerson, J. T., Swenson, S. C., & Lawrence, D. M. (2016). Evaluating the strength of the land–atmosphere moisture feedback in Earth system models using satellite observations. *Hydrology and Earth System Sciences*, *20*(12), 4837-4856.

Martens, B., Schumacher, D. L., Wouters, H., Muñoz-Sabater, J., Verhoest, N. E., & Miralles, D. G. (2020). Evaluating the land-surface energy partitioning in ERA5. *Geoscientific Model Development*, *13*(9), 4159-4181.

McColl, K. A., He, Q., Lu, H., & Entekhabi, D. (2019). Short-Term and Long-Term Surface Soil Moisture Memory Time Scales Are Spatially Anticorrelated at Global Scales. *Journal of Hydrometeorology*, *20*(6), 1165-1182.

McColl, K. A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D. J., Lu, H., ... & Entekhabi, D. (2017). Global characterization of surface soil moisture drydowns. *Geophysical Research Letters*, *44*(8), 3682-3690.

Mudelsee, M. (2013). *Climate time series analysis*. Heidelberg: Springer.

O'Neill, P., Bindlish, R., Chan, S., Njoku, E., & Jackson, T. (2018). Algorithm Theoretical Basis Document. Level 2 & 3 Soil Moisture (Passive) Data Products.

O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, R. Bindlish, and J. Chaubell. 2020. *SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 7*. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/HH4SZ2PXSP6A.

Qiu, J., Crow, W. T., & Nearing, G. S. (2016). The impact of vertical measurement depth on the information content of soil moisture for latent heat flux estimation. *Journal of Hydrometeorology*, *17*(9), 2419-2430.

Partington, J. R. (2004). *Linear operators and linear systems: an analytical approach to control theory* (No. 60). Cambridge University Press.

Phillips, C. L., Parr, J. M., & Riskin, E. A. (2003). *Signals, systems, and transforms* (p. 209). Upper Saddle River: Prentice Hall.

Riegger, J., & Tourian, M. J. (2014). Characterization of runoff-storage relationships by satellite gravimetry and remote sensing. *Water Resources Research*, *50*(4), 3444-3466.

Rivoire, P., Martius, O., & Naveau, P. (2021). A comparison of moderate and extreme ERA-5 daily precipitation with two observational data sets. *Earth and Space Science*, *8*(4), e2020EA001633.

Ruane, A. C., & Roads, J. O. (2007). 6-hour to 1-year variance of five global precipitation sets. *Earth Interactions*, *11*(11), 1-29.

Santanello Jr, J. A., Dirmeyer, P. A., Ferguson, C. R., Findell, K. L., Tawfik, A. B., Berg, A., ... & Wulfmeyer, V. (2018). Land–atmosphere interactions: The LoCo perspective. *Bulletin of the American Meteorological Society*, *99*(6), 1253-1272.

Salvucci, G. D., & Entekhabi, D. (1994). Equivalent steady soil moisture profile and the time compression approximation in water balance modeling. *Water Resources Research*, *30*(10), 2737-2749.

Sehler, R., Li, J., Reager, J. T., & Ye, H. (2019). Investigating relationship between soil moisture and precipitation globally using remote sensing observations. *Journal of Contemporary Water Research & Education*, *168*(1), 106-118.

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

Seneviratne, S. I., Koster, R. D., Guo, Z., Dirmeyer, P. A., Kowalczyk, E., Lawrence, D., ... & Verseghy, D. (2006). Soil moisture memory in AGCM simulations: analysis of global land–

atmosphere coupling experiment (GLACE) data. *Journal of Hydrometeorology*, *7*(5), 1090-1112.

Shangguan, W., Dai, Y., Duan, Q., Liu, B., & Yuan, H. (2014). A global soil data set for earth system modeling. *Journal of Advances in Modeling Earth Systems*, *6*(1), 249-263.

Suni, T., Guenther, A., Hansson, H. C., Kulmala, M., Andreae, M. O., Arneth, A., ... & Seneviratne, S. (2015). The significance of land-atmosphere interactions in the Earth system—iLEAPS achievements and perspectives. *Anthropocene*, *12*, 69-84.

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, *24*(5), 2527-2544.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485-498.

Thomson, R. E., & Emery, W. J. (2014). *Data analysis methods in physical oceanography*. Newnes.

Tuttle, S., & Salvucci, G. (2016). Empirical evidence of contrasting soil moisture–precipitation feedbacks across the United States. *Science*, *352*(6287), 825-828.

Wang, G., Kim, Y., & Wang, D. (2007). Quantifying the strength of soil moisture–precipitation coupling and its sensitivity to changes in surface water budget. *Journal of Hydrometeorology*, *8*(3), 551-570.

Wei, J., Dirmeyer, P. A., & Guo, Z. (2008). Sensitivities of soil wetness simulation to uncertainties in precipitation and radiation. *Geophysical research letters*, *35*(15).

Wei, J., Dirmeyer, P. A., & Guo, Z. (2010). How much do different land models matter for climate simulation? Part II: A decomposed view of the land–atmosphere coupling strength. *Journal of Climate*, *23*(11), 3135-3145.

Wei, J., & Dirmeyer, P. A. (2012). Dissecting soil moisture-precipitation coupling. *Geophysical Research Letters*, *39*(19).

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Willard, J., Jia, X., Xu, S., Steinbach, M., & Kumar, V. (2020). Integrating physics-based modeling with machine learning: A survey. *arXiv preprint arXiv:2003.04919*.

Wrona, E., Rowlandson, T. L., Nambiar, M., Berg, A. A., Colliander, A., & Marsh, P. (2017). Validation of the Soil Moisture Active Passive (SMAP) satellite soil moisture retrieval in an Arctic tundra environment. *Geophysical Research Letters*, *44*(9), 4152-4158.

Wu, W., Geller, M. A., & Dickinson, R. E. (2002). The response of soil moisture to long-term variability of precipitation. *Journal of Hydrometeorology*, *3*(5), 604-613.

Yuan, S., Quiring, S. M., & Leasor, Z. T. (2021). Historical changes in surface soil moisture over the contiguous United States: an assessment of CMIP6. *Geophysical Research Letters*, *48*(1), e2020GL089991.

Zhou, Y., Dong, X., Chen, H., Cao, L., Shao, Q., Sun, S., ... & Rao, J. (2020). Sub-seasonal variability of surface soil moisture over eastern China. *Climate Dynamics*, *55*(11), 3527-3541.

References From the Supporting Information

Cochran, W. T., Cooley, J. W., Favin, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., ... & Welch, P. D. (1967). What is the fast Fourier transform?. *Proceedings of the IEEE*, 55(10), 1664-1674.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90), 297-301.

Gentleman, W. M., & Sande, G. (1966, November). Fast Fourier transforms: for fun and profit. *In Proceedings of the November 7-10, 1966, fall joint computer conference* (pp. 563-578).

Gilman, D. L., Fuglister, F. J., & Mitchell Jr, J. M. (1963). On the power spectrum of “red noise”. *Journal of the Atmospheric Sciences*, *20*(2), 182-184.

Kay, S. M., & Marple, S. L. (1981). Spectrum analysis—a modern perspective. *Proceedings of the IEEE*, 69(11), 1380-1419.

Mandelbrot, B. B. (1982). *The fractal geometry of nature* (Vol. 2). New York: WH freeman.

Steele, J. H. (1985). A comparison of terrestrial and marine ecological systems. *Nature*, *313*(6001), 355-358.

Vasseur, D. A., & Yodzis, P. (2004). The color of environmental noise. *Ecology*, *85*(4), 1146-1152.