ABSTRACT: The interactions between land and atmosphere (with terrestrial and atmospheric coupling segments) play a significant role in weather and climate. A predominant segment of land–atmosphere (LA) feedbacks is the coupling between soil moisture (SM) and surface heat fluxes, the terrestrial coupling leg. The lack of high-quality, long-term, globally distributed observations, however, has hindered a robust, realistic identification of the terrestrial leg strength on a global scale. This exploratory study provides insight into how SM signals are translated into surface flux signals through the construction of a global depiction of the terrestrial leg from several recently developed global, gridded, observationally and satellite-based datasets. The feasibility of producing global gridded estimates of LA coupling metrics is explored. Five weather and climate models used for subseasonal to seasonal forecasting are confronted with the observational estimates to discern discrepancies that may affect their ability to predict phenomena related to LA feedbacks, such as drought or heat waves. The terrestrial feedback leg from observations corroborates the “hot spots” of LA coupling found in modeling studies, but the variances in daily time series of surface fluxes differ markedly. Better agreement and generally higher confidence are seen in metrics using latent heat flux than sensible heat flux. Observational metrics allow for clear stratification of model fidelity that is consistent across seasons, despite observational uncertainty. The results highlight the impact of SM on partitioning available surface energy and illustrate the potential of global observationally based datasets for the assessment of such relationships in weather and climate models.

KEYWORDS: Atmosphere-land interaction; Model evaluation/performance; Surface observations; Seasonal variability

1. Introduction

The interactions between land and atmosphere play a significant role in the climate and weather system. Variations in land states affect the atmosphere through their effects on surface fluxes of sensible and latent heat, and the subsequent impact of those fluxes on the atmosphere vertical structures and mixing processes. Land surface controls on atmospheric boundary layer properties have been well demonstrated (e.g., Ek and Holtslag 2004; Gentine et al. 2013). Recognizing this important role, atmospheric and land surface modeling communities have developed coupled land–atmosphere (LA) models (i.e., weather and climate prediction models), where the land surface models (LSMs) supply lower-boundary conditions to the atmospheric models. In the last two decades, many studies have explored the LA feedbacks due to emerging needs to improve the performance of weather and climate prediction models (e.g., Sippel et al. 2017; Seneviratne et al. 2010; Santanello et al. 2015, 2018; Koster et al. 2011; Dirmeyer et al. 2006; Koster et al. 2004; Dirmeyer 2013).

A predominant segment of LA feedbacks is the coupling between soil moisture (SM) and surface heat fluxes, known as the terrestrial leg of LA interactions and commonly characterized by a temporal correlation between variations in SM and fluxes (e.g., Guo et al. 2006; Lawrence et al. 2007; Dirmeyer 2011). The terrestrial leg identifies when, where, and to what extent soil moisture controls the partitioning of net radiation into sensible and latent heat fluxes. Several modeling studies have assessed the terrestrial leg based on weather and climate models and/or general circulation models (e.g., Dirmeyer et al. 2009; Dirmeyer 2011; Hirsch et al. 2016; Phillips et al. 2017; Williams et al. 2016). However, realistically underpinning terrestrial coupling and evaluating LSMs in terms of the accuracy of their representation of the terrestrial leg have been hampered due to lack of observations providing necessary spatial and temporal coverage.

The emergence of long-term, observationally based, global estimates of surface heat fluxes (e.g., Martens et al. 2017; Jung et al. 2019; Kummerow et al. 2019; Siemann et al. 2018) as well as the availability of independent satellite-based, global measurements of soil moisture (e.g., Dorigo et al. 2017) that span many years enable mapping of observationally based terrestrial coupling metrics on a global scale. Coupling metrics are multivariate by design, statistically relating the variables based on linking processes. Thus, such metrics can provide an understanding of the pathways by which variations at the land surface may affect weather and climate. Consequently, this provides opportunities to benchmark the performance of LSMs within weather and climate models, identify errors in their coupled behavior, and potentially improve their prediction skill.

This study has three main objectives. One is to assess the feasibility of using current global observationally and satellite-based datasets to construct a global depiction of the spatiotemporal patterns of the terrestrial leg of LA coupling.
The second is to provide improved insight into how the translation of a soil moisture signal into a surface flux signal varies by location and season. The third objective is to confront several weather/climate models with coupling metrics derived from these observationally based products, diagnose discrepancies, and attempt to identify the reasons for the discrepancies.

Several approaches have been applied to quantify the LA coupling strength. Here, we use two commonly used correlation-based metrics that can be applied similarly to both observations and models: 1) the Pearson correlation coefficient between variations in soil moisture and fluxes (e.g., Dirmeyer et al. 2009; Denissen et al. 2020), and 2) the terrestrial coupling index, which also takes into account the magnitude of variability in the response term over time (e.g., Guo et al. 2006; Dirmeyer 2011). However, we first address the satellite-based soil moisture product to approximate its random measurement errors and account for their degradation of calculated correlations by applying an approach based on lagged autocorrelation statistics. We analyze spatial patterns and seasonal variations in the coupling metrics between soil moisture and surface fluxes of latent heat and sensible heat. Section 2 describes the observational data, models, and in situ measurements used in the study. Section 3 describes the coupling metrics and the approach used to account for random measurement errors of satellite-based SM. The analysis of the terrestrial leg of LA coupling in observations and models is presented in section 4. Section 5 presents a summary and discussion of the main conclusions of this study.

2. Data

In this study, we use (quasi-)global gridded observationally based datasets and model hindcasts of soil moisture and land latent/sensible surface fluxes at the daily time scale. We also use some in situ data to diagnose differences found in the temporal variability of some of the global datasets. These are described here. All evaluations of global fields are performed at a common 1° spatial resolution and daily temporal resolution. Table 1 summarizes the employed observationally based products and models.

a. Observationally based datasets

Satellite-based observations of near-surface soil moisture are used along with observationally based estimates of land surface fluxes that combine parameterizations of land surface processes, in situ data, satellite-based and/or reanalysis meteorological datasets. Analyzing several observational datasets allows for evaluation of observational uncertainty in identifying the LA coupling strength.

1) LAND SURFACE FLUXES

(i) FLUXCOM

FLUXCOM estimates global gridded land surface fluxes using machine learning algorithms, trained by heat flux measurements from FLUXNET eddy covariance towers (Jung et al. 2019). In this study, we use the remote sensing and meteorological data based (RS+METEO) FLUXCOM product, which uses daily ERA5 meteorological datasets from ECMWF and mean seasonal and mean annual remotely sensed land surface parameters from MODIS. This product is available with a daily temporal and 0.5° spatial resolution from 1979 to 2018. The FLUXCOM product does not include estimates over unvegetated hot (mainly Sahara) and cold (mainly Greenland and Antarctica) deserts.

(ii) GLEAM

The Global Land Evaporation Amsterdam Model (GLEAM) provides estimates of terrestrial evaporation from satellite and reanalysis datasets with a daily temporal and 0.25° spatial resolution from 1980 to 2018. GLEAM uses the Priestley–Taylor equation to calculate potential evaporation (using surface net radiation and near-surface air temperature data) and converts it into actual evaporation using an evaporative stress factor that is a function of vegetation optical depth and soil moisture. In this study, we use version 3.3a of the GLEAM product, which uses reanalysis air temperature and radiation from ERA5 and a combination of gauge-based, reanalysis, and satellite-based precipitation (Martens et al. 2017). GLEAM does not provide estimates of sensible heat flux.

(iii) GEWEX

The GEWEX LandFlux project provides global land sensible and latent heat flux products on a 1° × 1° grid with 3-hourly temporal resolution from 1998 to 2009 based on remotely sensed data and reanalysis meteorology. The GEWEX LandFlux data include several land flux datasets. Here, we use products supplied as part of the GEWEX Integrated Product (GEWEX-IP; Kummerow et al. 2019): two of latent heat flux

<table>
<thead>
<tr>
<th>Dataset Description</th>
<th>Dataset Version</th>
<th>Variable(s)</th>
<th>Temp. coverage</th>
<th>Temp. res.</th>
<th>Spatial res.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observationally based products</td>
<td>FLUXCOM</td>
<td>LS+METEO</td>
<td>LE, H</td>
<td>1979–2018</td>
<td>Daily</td>
</tr>
<tr>
<td>GLEAM</td>
<td>3.3a</td>
<td>E</td>
<td>1980–2018</td>
<td>Daily</td>
<td>0.25° × 0.25°</td>
</tr>
<tr>
<td>GEWEX-IP</td>
<td>2019</td>
<td>LE, H</td>
<td>1998–2009</td>
<td>3-hourly</td>
<td>1° × 1°</td>
</tr>
<tr>
<td>CCI</td>
<td>5.2</td>
<td>SM</td>
<td>1978–2019</td>
<td>Daily</td>
<td>0.25° × 0.25°</td>
</tr>
<tr>
<td>Models</td>
<td>GEFS</td>
<td>v12</td>
<td>SM, LE, H</td>
<td>1999–2016</td>
<td>Daily</td>
</tr>
<tr>
<td>FIM</td>
<td>FIM-iHYCOM</td>
<td>SM, LE, H</td>
<td>1999–2016</td>
<td>Daily</td>
<td>1° × 1°</td>
</tr>
<tr>
<td>GEOS</td>
<td>S2S_2.1</td>
<td>SM, LE, H</td>
<td>1999–2016</td>
<td>Daily</td>
<td>1° × 1°</td>
</tr>
<tr>
<td>CCSM4</td>
<td>4</td>
<td>SM, LE, H</td>
<td>1999–2016</td>
<td>Daily</td>
<td>1° × 1°</td>
</tr>
<tr>
<td>CFSv2</td>
<td>2</td>
<td>SM, LE, H</td>
<td>1999–2010</td>
<td>Daily</td>
<td>1.5° × 1.5°</td>
</tr>
</tbody>
</table>
products obtained by the Priestley–Taylor Jet Propulsion Laboratory (PT-JPL) model and the Penman–Monteith-based Mu model (PM-Mu) (McCabe et al. 2015), as well as one sensible heat flux product parameterized using the Monin–Obukhov similarity theory and near-surface air temperature data from ECMWF Interim reanalysis (Siemann et al. 2018).

The LandFlux products are calculated using precipitation and meteorological data from the National Centers for Environment Prediction (NCEP) Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), net radiation data from the GEWEX Surface Radiation Budget (SRB; Stackhouse et al. 2011), and the land surface temperature product that obtained by merging High-Resolution Infrared Radiation Sounder (HIRS) satellite retrievals with the land surface temperature estimates from CFSR.

Since the introduced observationally based products use different physical/statistical models and observational input data to estimate surface fluxes, they may differ in the temporal variabilities and climatological characteristics. This, of course, results in observational uncertainties in identifying the LA coupling strength. The evaluation of temporal variances of observationally based products using a set of in-site latent heat flux measurements (see section 2c) is presented in section 4d. The potential evaporation, and consequent latent heat flux, calculated using commonly used methods (e.g., Priestly–Taylor equation) are sensitive to temperature changes, the estimation method used, and other controlling variables across methods (McKenney and Rosenberg 1993; Milly and Dunne 2017). This can be a source of uncertainty in identifying the LA coupling strength.

2) SOIL MOISTURE

The ESA Climate Change Initiative (CCI) provides multi-decadal, global satellite-observed soil moisture (ESA CCI SM) datasets (Dorigo et al. 2017). In this study, we use version 5.2 of ESA CCI SM combined product, which is the combination of various single-sensor active [i.e., ERS-1/2 Active Microwave Instrument Wind Scatterometer (AMI WS), ERS-2 AMI WS, and MetOp-A/B ASCAT] and passive (i.e., SMMR, SSM/I, TMI, WindSat, AMSR-E, AMSR2, SMOS, and SMAP) microwave soil moisture products. This is a daily product with a grid spacing of 0.25°. The dataset covers the period from 1978 to 2019. In this product, some areas suffer from seasonal (e.g., northern latitudes) or even continuous (e.g., tropical rain forests) data gaps. The ESA CCI SM products and their underlying merging methodology have been evaluated using ground-based soil moisture observations (e.g., Gruber et al. 2017; Ma et al. 2019; Wang et al. 2021).

b. Model hindcasts

We examine five models including two numerical weather prediction models: 1) GEFS from NOAA/NCEP; 2) the Flow-Following Icosahedral Model (FIM) from NOAA/ESRL; and three seasonal/climate prediction models: 3) GEOS from NASA GMAO, 4) CCSM4 from NCAR run at the University of Miami Rosenstiel School of Marine and Atmospheric Science, and 5) CFSv2 from NCEP. All models include a land surface model coupled to its corresponding global atmospheric and/or oceanic model in a free-running (unconstrained) mode.

The required data (including land fluxes and soil moisture content) of models (except CFSv2) are extracted from the Subseasonal Experimental (SubX) hindcast datasets (Pegion et al. 2019). Even though CFSv2 is a participating model in SubX, it does not provide land fluxes and soil moisture data for this experiment. Therefore, the CFSv2 dataset is extracted from the Subseasonal to Seasonal (S2S) prediction project database (Vitart et al. 2017). SubX data are available on a uniform 1° × 1° longitude–latitude grid with a daily temporal resolution from 1999 to 2016. S2S-CFSv2 data are available with a daily temporal and 1.5° spatial resolution from 1999 to 2010.

The S2S CFSv2 and SubX GEFS datasets provide soil moisture at various depths including shallow soil moisture of top 20 cm and top 10 cm, receptively. The SubX FIM, CCSM4, and GEOS datasets do not provide soil moisture at various depths, but soil moisture vertically integrated from surface to 2 m, 3.43 m, and bedrock, respectively. c. In situ measurements

Daily mean measurements of latent heat flux from FLUXNET2015 at the Central Facility for the ARM Southern Great Plains network (Biraud et al. 2003) are used to evaluate the temporal variances of observationally based products.

3. Methods

a. Metrics

Two well-known metrics are used to investigate the land–atmosphere coupling strength and its spatial and seasonal variability. The most common metric is the temporal Pearson’s correlation coefficient (R) calculated between daily anomalies of land heat flux (latent or sensible) and land state (soil moisture). A positive (negative) correlation between soil moisture and latent (sensible) heat flux implies fluxes are controlled by soil moisture (i.e., the water-limited regime). In contrast, a negative (positive) correlation between soil moisture and latent (sensible) heat flux indicates the energy-limited regime where the atmosphere controls the fluxes (Denissen et al. 2020; Dirmeyer 2011; Dirmeyer et al. 2014). The high correlation between land state and flux may hint at causality but is not necessarily an indication of important feedbacks where the dynamics of variables are not substantial over time.

The second metric is the terrestrial coupling index (I). This coupling metric takes into account the variability in the response term over time and is defined as the product of the standard deviation of the land flux (response) and the Pearson’s correlation between land flux and soil moisture (Guo et al. 2006; Dirmeyer 2011). Unlike the correlation coefficient, the terrestrial coupling index is sensitive to the unit of heat fluxes, therefore, the units among the products are harmonized. The unit of soil moisture has no effect on the coupling metrics. Findell et al. (2015) evaluated the observational data requirements necessary for the robust quantification of LA coupling metrics (including R and I). Their analyses suggest
that 6- and 10-yr surface observation records are required to robustly estimate LA coupling strength using \( R \) and \( I \) metrics, respectively.

We compute the metrics using the daily anomalies of land state and fluxes at each grid point. To obtain anomalies of the variables, the climatology and long-term trends are removed from datasets. A triangular window algorithm (i.e., a smoothing function with weights linearly decreasing with distance from the center point) with a smoothing window of 31 days (±15 days) is applied to remove the effect of noise (as well as sparsity in some cases) when calculating the climatology and long-term trend. Since the ESA CCI product provides measurements of near-surface soil moisture, if a model (e.g., GEFS and CFSv2) provides estimates of soil moisture at various depths, for a fair, consistent comparison, only SM data at the shallowest depth are used for the computation of the coupling metrics. The comparison between observations and models that only provide vertically integrated soil moisture (e.g., GOES, CCSR, and FIM) is performed by acknowledging the different influences of soil moisture depth on surface flux variations. We also believe smoothing the soil moisture time series would make their spectra more compatible with the column soil moisture from models, but that is not done here.

For models, all ensemble members are used for the robust calculation of the metrics. To identify seasonal dynamics in the coupling strength between soil moisture and heat fluxes, the metrics are calculated for seasonal subsets of the daily anomalies described above.

For calculating the climatology and long-term trends, the available temporal coverage of each product is used while for calculating the coupling metrics, the maximum common time period for each pair of soil moisture and surface flux datasets is used.

In addition to the magnitude of the coupling metrics, the spatial correlation between the global maps of coupling metrics obtained from models and observations is analyzed to evaluate the spatial pattern consistency among models and observations. To enable this analysis, CFSv2 data were interpolated bilinearly from its native resolution from the S2S dataset to a regular \( 1\,\degree \times 1\,\degree \) grid.

### b. Observational error

Given that models, by their nature, do not suffer from random errors in estimates of state variables, and in order to consistently compare models with observations, the effect of random measurement errors on coupling metrics should be quantified. The estimates of Pearson’s correlation coefficient between fluxes and measurements of soil moisture are biased low due to the random measurement errors (Findell et al. 2015). Therefore, we first approximate the measurement errors in the satellite-based soil moisture data (ESA CCI SM) and then “correct” the observed coupling metrics to account for those errors.

To estimate observational soil moisture errors, we follow the framework proposed by Vinnikov and Yeserkepova (1991) and successfully applied in recent studies to evaluate measurement errors of in situ soil moisture observational networks (Dirmeyer et al. 2016) and remote sensing soil moisture retrievals (Kumar et al. 2018). The temporal variability of soil moisture corresponds to a statistical model of red noise and therefore can be generated by a first-order Markov process. Based on this recognition, the ratio of measurement error variance \( \delta^2 \) to the temporal variance of measurements \( \sigma^2 \) can be estimated as

\[
\frac{\delta^2}{\sigma^2} = \frac{a}{1 + a},
\]

where \( a \) is the intercept at \( \tau = 0 \) of a linear best fit of \( \ln(\tau) \) versus \( \tau \), where \( \tau \) is the autocorrelation function of observed soil moisture time series at time lag \( \tau \). It should be noted that the parameterization of soil moisture by a first-order Markov process is subject to the resemblance between forcing (e.g., precipitation) and white noise, which is only valid for time scales longer than about one week (Delworth and Manaba 1993). Therefore, we use the 5-day average of ESA CCI SM data for the computation of autocorrelation function and \( a \). In our analysis, the observational errors are estimated using autocorrelations at lags of 1, 2, and 3 (from 5-day averages of SM data).

The correlation coefficient between ESA CCI SM and heat flux, \( R \), can be “corrected” as

\[
R_{\text{corrected}} = R \sqrt{1 + \frac{\delta^2}{\sigma^2}}.
\]

For simplicity, hereafter, the corrected correlation coefficient between soil moisture and flux is used in all cases and is just called the correlation coefficient \( R \).

### 4. Results

#### a. SM observational error

Figure 1 shows the global distribution of relative measurement error (\( \delta/\sigma \)) for the combined soil moisture product from ESA CCI. The magnitude of relative measurement errors coincides with the vegetation density (except for deserts) with larger errors over densely vegetated regions (e.g., eastern United States) and smaller errors over savanna regions (e.g., Sahel). The relative errors are high over deserts (e.g., Sahara and Gobi) due to very small temporal variance of soil moisture (the denominator) in these regions and low accuracy of scatterometer retrievals over dry regions. The spatial pattern of relative measurement error estimates in Fig. 1 is consistent with previous studies (Kumar et al. 2018; Wagner et al. 2012). Corrections tend to be largest over forests and the cores of major deserts. They are small over most agricultural areas, grasslands, and subtropical regions.

The relative measurement error averaged over the globe is 0.38, which is lower than the relative measurement error of single-sensor products participating in the combined ESA CCI product (AMSR-E, ASCAT, SMOS, AMSR2, and SMAP) as reported by Kumar et al. (2018). In addition to the fact that the error variance of blended products is typically smaller than the error variances of any of the individual products, the small domain averaged relative measurement error of ESA CCI SM
may be due to the fact that ESA CCI SM does not include the tropical rain forests, where the other single-sensor products show high measurement errors.

b. Pearson correlation coefficient

Figure 2 shows the global distribution for four different seasons of the Pearson correlation coefficient $R$ between daily anomalies of latent heat flux (LE) from the four observationally based products and SM from CCI, i.e., $R_{(SM, LE)}$. The figure shows the seasonal and regional variations in the strength of SM–LE relations within each product. The histogram on the bottom left corner of each map shows the frequency distribution of $R_{(SM, LE)}$. Bar colors of histograms correspond to color coding of the color bar. In general, the spatial patterns of $R_{(SM, LE)}$ are similar in the different observational products, although the magnitude of $R_{(SM, LE)}$ differs significantly between them. Consistent with previous studies, relatively high $R_{(SM, LE)}$ is observed in desert areas (e.g., the U.S. Great Basin and Kalahari deserts) and relatively low $R_{(SM, LE)}$ is observed in densely vegetated areas [e.g., Southern Cone (the southernmost areas of South America) and the eastern United States]. The FLUXCOM and GEWEX–PM-Mu do not provide estimates of LE, hence $R_{(SM, LE)}$ over deserts like the Sahara, but GLEAM and GEWEX–PT-JPL show relatively high $R_{(SM, LE)}$ over this region. Although GLEAM and GEWEX–PT-JPL both use the Priestley–Taylor relationship to estimate LE, their resulting global fields of $R_{(SM, LE)}$ appear quite different. Several factors are contributing to the observed differences, namely, the sensitivity of Priestley–Taylor-based models to temperature changes (Milly and Dunne 2017) and differences in assumptions and employed methodologies to estimate transpiration and partition water between transpiration and interception (Good et al. 2017; Miralles et al. 2016). Likewise, FLUXCOM and GEWEX–PM-Mu both consider the effects of vegetation on LE, yet they also differ considerably.

Relatively high SM–LE correlation observed over the Great Plains of North America, northern India, the Sahel, and equatorial Africa aligns well with known “hot spots” of land–atmosphere coupling for precipitation and temperature found by Koster et al. (2006), Guo et al. (2006), and Koster et al. (2004). However, strong coupling between SM and LE is also observed over regions such as Australia and South Africa where are not identified as hot spots for precipitation and temperature. On the other hand, the hot spot found over China does not show a high observed SM–LE correlation during boreal summer except for GEWEX–PM-Mu. The SM–LE hot spots in SON are stronger and cover a broader spatial region than in JJA for many of the traditional hot-spot regions. The negative $R_{(SM, LE)}$ over some regions (e.g., at higher latitudes and in Europe during wintertime and Southeast Asia on East Asia summer monsoon) implies atmospheric control of LE.

The switch of correlation from positive to negative in MAM and JJA over east India, which is related to the Indian monsoon onset in mid-June, is well captured by observations (except GEWEX–PM-Mu). Before the Indian monsoon onset, there is little precipitation during MAM and therefore, SM controls LE (i.e., water-limited regime) and a positive correlation is observed between LE and SM. After the monsoon onset during JJA, precipitation increases and soil moisture becomes plentiful for ET. Therefore, available energy controls ET and a negative correlation is observed over this region in JJA.

The spatial average $R_{(SM, LE)}$ over grid cells having a finite value from all four observational products, and the percentage of grid cells with positive significant $R$ ($p$ value $< 0.01$) are shown by $R$ and $P$, respectively, for each map in Fig. 2. Red areas represent hot spots of positive land–atmosphere feedback in the water cycle, as traditionally defined. Among the products, GLEAM (GEWEX–PT-JPL) shows the highest (lowest) values of $R$ and $P$ in each season. Higher values of $R_{(SM, LE)}$ in GLEAM are likely due to the fact that GLEAM estimates evapotranspiration using an explicit dependence of soil moisture. The two GEWEX–IP products have an austral autumn peak in $R$, while the others have a boreal autumn peak. The differences among maps in each season reveal the uncertainty of LA coupling estimates from different observational products. Our investigation (see Figs. S1 and S2 in the...
online supplemental material) also shows FLUXCOM RS + METEO products derived from different meteorological reanalyses produce coupling metrics with slightly different ranges, indicating the role of input data in the representation of LA coupling strength. Among various FLUXCOM RS + METEO products, FLUXCOM-CERES-GPCP, which uses radiation data from CERES and precipitation data from GPCP, shows a low consistency with others.

Figure 3a shows the spatial correlation among the global maps of $R$(SM, LE) from different observational datasets for four different seasons. In each season, only grid cells that have a finite $R$(SM, LE) value in all four observational products are considered for the computation of spatial correlation. The peak correlation for each pair happens in either JJA or SON, coincident with the period of maximum SM–LE coupling across the land areas of the Northern Hemisphere. Among the products, FLUXCOM and GLEAM show the highest spatial correlation, with the peak in JJA, even though they use different methodologies for estimation of daily fluxes (FLUXCOM uses a statistically based algorithm while GLEAM is based on a physical model). However, GLEAM and FLUXCOM share common meteorological inputs, which may contribute to their similarity. As mentioned above, GLEAM and GEWEX–PT-JPL use the same formulation for LE, yet they do not stand out as especially similar. FLUXCOM shows the lowest correlation with other observations in DJF. In comparison to FLUXCOM, GEWEX–PM-Mu overestimates $R$(SM, LE) over Brazil in DJF, which was predominantly recognized as a wet soil moisture regime (i.e., the primary driver of terrestrial evaporation is available energy) in other studies (e.g., Schwingshackl et al. 2017) and low/negative $R$(SM, LE) is expected. In comparison to FLUXCOM and other products, GLEAM overestimates $R$(SM, LE) over the northern, northwestern, and center of Europe where during December–February evaporation dynamics are mostly influenced by the energy supply (Martens et al. 2018).

Figure 4 shows the global distribution of $R$(SM, LE) from five model hindcasts for four different seasons, depicting the seasonal and regional $R$(SM, LE) variations within each model. The intercomparison of models reveals almost the same spatial patterns and seasonal dynamics of $R$(SM, LE), but remarkably different magnitudes. The models (except CCSM4) represent the stronger SM–LE coupling over the Northern Hemisphere during JJA and SON than MAM and DJF. A very low (or negative) $R$(SM, LE) estimate at high latitudes and in tropical rain forests is a common feature of all models. Among the five models, GEFS and CCSM4 respectively show the highest and lowest domain averaged $R$ in each season.
The comparison between forecast models (Fig. 4) and observations (Fig. 2) reveals that GEFS, GEOS, and CFSv2 models characterize the effect of SM variations on LE variations much more strongly than observations, whereas CCSM4 and FIM models have a weaker coupling between SM and LE than observations. The weak correlation between SM and LE from CCSM4 is consistent with previous studies. Mei and Wang (2012) found CCSM4 underestimates land–atmosphere

---

**Fig. 3.** Spatial correlation of R(SM, LE) maps from pairings of (a) observations, (b) models, (c) FLUXCOM and a model, and (d) GEWEX–PM-Mu and a model for four different seasons.

**Fig. 4.** As in Fig. 2, but from the five forecast models.
coupling strength in comparison to observations and reanalysis data. Dirmeyer et al. (2013b) established the weakness in coupling was profound with subsurface soil moisture, but not with surface soil moisture. A possible reason for relatively low $R_{\text{SM, LE}}$ from FIM (and to a lesser extent from CCSM4) may be the influences of different soil moisture depths on magnitudes of $R_{\text{SM, LE}}$. From CCSM4 and FIM we calculate $R_{\text{SM, LE}}$ using vertically integrated SM data, whereas the observational $R_{\text{SM, LE}}$ is calculated using near-surface SM from ESA CCI, and $R_{\text{SM, LE}}$ from GEFS and CFSv2 is calculated using 0–0.1 and 0–0.2 m below surface SM data, respectively. Note that, similar to CCSM4 and FIM, $R_{\text{SM, LE}}$ from GEOS is calculated using vertically integrated SM data. However, $R_{\text{SM, LE}}$ from GEOS is found to be comparable with those of GEFS and CFSv2 models, weakening the latter hypothesis. Our analyses on $R_{\text{SM, LE}}$ calculated using SM at varying depths (i.e., SM from 0–10, 0–40, 0–100, and 0–200 cm layers; results are not shown here) from SubX-GEFS indicate that the magnitude of $R_{\text{SM, LE}}$ weakens using SM in deeper layers. However, regardless of SM depth, $R_{\text{SM, LE}}$ from GEFS is remarkably stronger than $R_{\text{SM, LE}}$ from observations (Fig. 2) and both FIM and CCSM4 models. Likewise, in an independent study, Dirmeyer (2011) found strong consistencies between calculations of a correlation-based SM–LE statistic using soil moisture at varying depths and from four datasets, but not for CCSM4 (Dirmeyer et al. 2013b).

Similar to the observations (Fig. 2), GEFS, GEOS, and CFSv2 show relatively high $R_{\text{SM, LE}}$ in desert areas. CCSM4 and FIM, however, do not show a strong SM–LE coupling over the Sahara and Arabian deserts.

Figure 3b shows the spatial correlation between maps of $R_{\text{SM, LE}}$ from pairings of the models for four seasons. Based on Fig. 3a, the pair of FLUXCOM and GLEAM shows the highest spatial correlation, whereas the pairings of GEWEX–PM-Mu and FLUXCOM, and GEWEX–PM-Mu and GLEAM show the lowest spatial correlation. Therefore, rather than analysis of all observation-model pairs, here we just present the results for spatial comparison between FLUXCOM/GEWEX–PM-Mu (as observations) and the models in Figs. 3c and 3d, respectively. The comparison between Figs. 3a and 3b reveals that in general the global spatial patterns of $R_{\text{SM, LE}}$ within different observations are more consistent (values range from 0.48 to 0.83 with an average of 0.63 over seasons and observations) than among the models (values range from 0.14 to 0.83 with an average of 0.51 over seasons and models). This suggests greater uncertainty between models.

Only one global observational dataset for soil moisture has been used. Our investigations show that including the CCI radiometer-only (PASSIVE) product (which combines only passive microwave soil moisture products) and the CCI scatterometer-only (ACTIVE) product (which combines only single-sensor active soil moisture products), in addition to the COMBINED CCI product, does not change the conclusion on a comparison between consistency of observations versus models. Spatial correlations of both coupling indices and correlations between SM and LE were calculated between all pairings of models and pairings of observational products including the three different CCI products. Seasonal differences in variances among all pairings are as much as 630% larger among model pairings than observational pairings.

Among the models, GEFS and GEOS show the highest spatial similarity for all seasons, but the colors indicate clusters of similarity. CFSv2, GEFS, and GEOS share high pattern correlations. CCSM4 does not resemble well any of the other models, and FIM lies in between. CCSM4 shows the lowest spatial correlation with other models in MAM because of the high discrepancy in estimates of $R_{\text{SM, LE}}$ at high latitudes. GEFS shows the lowest spatial correlation with GEOS and CFSv2 in JJA because of a high discrepancy in estimates of $R_{\text{SM, LE}}$ over high latitudes.

Stratifications in Figs. 3c and 3d are basically identical, suggesting real quality differences among the models regardless of the observational product used. Among the models, the spatial patterns of GEFS, CFSv2, and GEOS show a high degree of consistency with observations, while the spatial patterns of CCSM4 and FIM show the lowest degree of consistency. That GEFS, CFSv2, and GEOS are highly correlated with each other and with the observations, while FIM and CCSM4 are not, suggests FIM and especially CCSM4 are in a different category of quality compared to GEFS, CFSv2, and GEOS. The spatial pattern of GEWEX–PM-Mu shows the lowest degree of consistency not only with other observations (Fig. 3a), but also with models in comparison with FLUXCOM.

Figure 5 shows the global distribution of the Pearson correlation coefficient between anomalies of $H$ from the FLUXCOM and GEWEX products and CCI SM for four seasons, $R_{\text{SM, LE}}$. Clearly, FLUXCOM and GEWEX do not present the same spatial patterns nor ranges of $R_{\text{SM, LE}}$. When energy is available, moisture conditions control the partitioning of heat fluxes, such that the increase of SM, land evaporative fraction [i.e., LE/(LE + $H$)] increases, as explained in the local land–atmosphere coupling (LoCo) process chain by Santanello et al. (2011). Therefore, it is expected that most regions with a pronounced influence of SM variations on latent heat flux, also experience strong negative $R_{\text{SM, LE}}$. This can clearly be seen by comparing the spatial patterns of $R_{\text{SM, LE}}$ (in Fig. 3) and $R_{\text{SM, LE}}$ (in Fig. 5) from FLUXCOM. However, very low absolute $R_{\text{SM, LE}}$ from GEWEX over hot spots and other regions (e.g., Sahara, Arabian, and Kalahari Deserts) suggests the possibility of an insufficient representation of the surface sensible heat flux variability from GEWEX datasets for an accurate representation of SM–H coupling. This issue, in addition to the reported inconsistencies between the GEWEX sensible heat flux products (Siemann et al. 2018) and literature values (Jung et al. 2011), especially at high latitudes and places such as the U.S. Midwest, North American Great Plains, and central Asia, gives us greater confidence in $R_{\text{SM, LE}}$ from FLUXCOM to use as a basis for evaluating models for SM–H coupling estimates.

Figure 6 shows the global distribution of $R_{\text{SM, LE}}$ from five model hindcasts for four different seasons, depicting the seasonal and regional variations within each model. The intermodel comparison reveals almost the same seasonal dynamics of $R_{\text{SM, LE}}$, but remarkably different magnitudes of
FLUXCOM in other regions.

comparable with the observational estimate of R. Differences of different soil moisture depths on magnitudes of compared to shows the frequency distribution of JUNE 2021 ABDOLGHAFOORIAN AND DIRMEYER 1515

reason for relatively low regions. Note that, as discussed for (Fig. 5) reveals that GEFS overestimates the correlation between SM and H variations, especially in South America, Africa, India, and Southeast and East Asia. CCSM4 and FIM, however, underestimate the coupling between SM and H in all regions. Note that, as discussed for R(SM, LE), a possible reason for relatively low R(SM, H) from CCSM4 and FIM compared to R(SM, LE) from FLUXCOM may be the influences of different soil moisture depths on magnitudes of R(SM, H). Even though GEOS and CFSv2 overestimate R(SM, H) at the Southern Hemisphere, their estimates are comparable with the observational estimate of R(SM, H) from FLUXCOM in other regions.

Figure 5. Global distribution of the Pearson correlation coefficient (R) between anomalies of sensible heat flux and soil moisture from two observationally based (left) FLUXCOM and (right) GEWEX products for four different seasons: March–May (MAM), June–August (JJA), September–November (SON), and December–February (DJF). The \( \langle R \rangle \) shows the average of R over grid cells that have a finite value of R in all two observational products. The \( P \) shows the percentage of grid cells with negative significant R (\( p \) value < 0.01). The histogram on the lower left of each map shows the frequency distribution of R.

R(SM, H). The Northern Hemisphere shows a weaker SM–H coupling than the Southern Hemisphere during MAM and JJA. Among the five models, GEFS and CCSM4 respectively show the highest and lowest absolute domain averaged R in all seasons. The GEFS, GEOS, GEOS, and FIM models show comparable spatial patterns especially during MAM, SON, and DJF, while CCSM4 spatial patterns show a low degree of consistency with other models.

The comparison between models (Fig. 6) and FLUXCOM (Fig. 5) reveals that GEFS overestimates the correlation between SM and H variations, especially in South America, Africa, India, and Southeast and East Asia. CCSM4 and FIM, however, underestimate the coupling between SM and H in all regions. Note that, as discussed for R(SM, LE), a possible reason for relatively low R(SM, H) from CCSM4 and FIM compared to R(SM, LE) from FLUXCOM may be the influences of different soil moisture depths on magnitudes of R(SM, H). Even though GEOS and CFSv2 overestimate R(SM, H) at the Southern Hemisphere, their estimates are comparable with the observational estimate of R(SM, H) from FLUXCOM in other regions.

Figure 7a shows the spatial correlation between the maps of \( R(\text{SM}, \text{H}) \) from FLUXCOM and GEWEX for four different seasons. Low values of spatial correlation in all seasons indicate the slight similarity between the spatial pattern of these two observational products. Figure 7b shows the spatial correlation between maps of \( R(\text{SM}, \text{H}) \) from pairings of the five models for four seasons. Contrary to \( R(\text{SM}, \text{LE}) \), the peak spatial correlation of \( R(\text{SM}, \text{H}) \) maps from almost all pairings of models happens in either MAM or DJF. Among the models, the two from NCEP (GEFS and CFSv2) show the highest spatial correlation for all seasons. The low spatial correlation between GEFS/CFSv2 and other models in JJA is probably due to the unreasonable overestimation of \( R(\text{SM}, \text{H}) \) from these models at high latitudes. Similar to \( R(\text{SM}, \text{LE}) \), the spatial patterns of \( R(\text{SM}, \text{H}) \) from CCSM4 again show the lowest consistency with other models.

The spatial comparison between the observational and the model correlations is shown in Figs. 7c and 7d. The models show higher spatial consistency with both observational products in MAM and DJF but, in general, no single model consistently outperforms any other. Poor (or even negative) spatial correlation between GEWEX and models is probably due to the reasons discussed earlier regarding the poor performance of the GEWEX sensible heat flux product.

c. Terrestrial coupling index

The other metric that we use to analyze the land–atmosphere coupling is the terrestrial coupling index I. The terrestrial coupling index is the product of the correlation coefficient and the standard deviation of the flux and accounts for temporal variabilities in the response term of LA linkage (Dirmeyer et al. 2013a). Figure 8 shows the global distribution of terrestrial coupling index calculated from daily anomalies of SM and LE, \( I_{\text{SM:LE}} \), from the four observationally based products for four different seasons. The figure shows the seasonal and regional variations in the strength of SM–LE relations within each product. In general, neither the spatial patterns nor the magnitude of \( I_{\text{SM:LE}} \) across the products are fully consistent, even though they share some similarities. Australia, Southern Africa, Mexico, and savannas regions (e.g., Sahel and east of South America) show a high terrestrial coupling index from all products, especially in JJA and SON. In contrast, high latitudes and southeast Asia show a relatively low \( I_{\text{SM:LE}} \). The distribution of \( I_{\text{SM:LE}} \) (shown in the lower-left corner of each map) is right-skewed, with the mode (maximum occurrence) at the 0–2 W m\(^{-2}\) interval, for all seasons and products. Besides, the comparison between Figs. 2 and 8 reveals that the spatial variability of \( I_{\text{SM:LE}} \) normalized by \( \langle I \rangle \) is smaller than \( R(\text{SM, LE}) \) normalized by \( \langle R \rangle \) for most of products/seasons. These are mostly due to the fact that over desert areas, even though SM and LE are strongly correlated, temporal variations of LE are not large and therefore the terrestrial coupling index over these regions is low. FLUXCOM and GEWEX–PT-JPL show negative \( I_{\text{SM:LE}} \) at high latitudes after snowmelt and wet regions as expected in the energy-limited regime when water is
available and latent heat flux controls the soil moisture variability.

As with $R(SM, LE)$, GLEAM and GEWEX–PT-JPL respectively show the strongest and weakest coupling for $I_{SM:LE}$. Given the comparable ranges of $R(SM, LE)$ from GLEAM, FLUXCOM, and GEWEX–PM-Mu, larger values of $I_{SM:LE}$ from GLEAM are due to larger temporal variabilities of GLEAM LE estimates. This point deserves further investigation, and it is examined in section 4d.

Figure 9a shows the spatial correlation between the global maps of $I_{SM:LE}$ from different observational products for four different seasons. Similar to $R(SM, LE)$, spatial patterns of the $I_{SM:LE}$ maps from FLUXCOM and GLEAM show the highest spatial correlation for all seasons. The maps of $I_{SM:LE}$ from

![Figure 6](image-url)

**FIG. 6.** As in Fig. 5, but for the five forecast models.

![Figure 7](image-url)

**FIG. 7.** As in Fig. 3, but for $R(SM, H)$. 

FLUXCOM and GEWEX–PT-JPL not only show consistency in terms of magnitude but also in terms of spatial correlation, as shown in Fig. 9a.

Figure 10 shows the global distribution of \( I_{\text{SM:LE}} \) from the five models. Among the models, GEOS, followed by GEFS, shows the largest \( I_{\text{SM:LE}} \), whereas CCSM4 shows the smallest \( I_{\text{SM:LE}} \) in all seasons. Based on this figure, GEFS and GEOS show the highest consistency in terms of magnitude, spatial patterns, and seasonal dynamics. CCSM and FIM on the other hand, show the lowest consistency with other models. The relatively large values of \( I_{\text{SM:LE}} \) from models over the tropics and Southern Hemisphere are due to larger variabilities of latent heat flux in these regions.

The comparison between models and observations suggests an overestimation of \( I_{\text{SM:LE}} \) from GEOS and GEFS especially over the Southern Hemisphere, as it is for \( R(\text{SM, LE}) \). Similar to observations, the distribution of \( I_{\text{SM:LE}} \) from models is mostly right-skewed, with the mode at the 0–2 W m\(^{-2}\) interval. In general, the seasonal dynamics of \( I_{\text{SM:LE}} \) from models and observations are comparable, as they present a relatively strong (weak) terrestrial coupling over the Northern Hemisphere in JJA (DJF) as opposed to the Southern Hemisphere where a strong (weak) terrestrial coupling is seen in DJF (JJA).

Figure 9b shows the spatial correlation between maps of \( I_{\text{SM:LE}} \) from pairings of the models for four seasons. The spatial patterns of \( I_{\text{SM:LE}} \) from FLUXCOM and GEWEX–PM-Mu (as observations) and the models are presented in Figs. 9c and 9d, respectively. The spatial patterns of \( I_{\text{SM:LE}} \) from CFSv2, GEOS, and GEFS show the highest degree of consistency [similar to those of \( R(\text{SM, LE}) \)], whereas CCSM4 shows a low degree of consistency with other models and also with GEWEX–PM-Mu. Based on Figs. 9c and 9d, no single model consistently outperforms any other in comparison with observations. All models (except CCSM4) show the lowest spatial correlation with observations in JJA, mainly due to inaccurate representations of \( I_{\text{SM:LE}} \) in densely vegetated areas (e.g., forest areas in Africa, South America) and at high latitudes (e.g., Canada and Russia).

Similar to \( R(\text{SM, LE}) \), the comparison between Figs. 9a and 9b reveals that the global spatial patterns of \( I_{\text{SM:LE}} \) are more consistent across observations (values range from 0.35 to 0.82 with an average of 0.53 over seasons and observations) than across the models (values range from 0.14 to 0.82 with an average of 0.53 over seasons and models). Considering the disparities in the production of the various observational datasets, this strongly indicates greater uncertainty and suggests lower fidelity among models.

Figure S3 shows the global distribution of the terrestrial coupling index between daily anomalies SM and \( H, I_{\text{SM:H}} \), from FLUXCOM and GEWEX for four different seasons. Similar to \( R(\text{SM, H}) \), the \( I_{\text{SM:H}} \) maps from the two products do not share similarity either in terms of spatial patterns or ranges of \( I_{\text{SM:H}} \). Figure S4 shows the global distribution of \( I_{\text{SM:H}} \) from the five models. Among the models, GEFS (CCSM4) shows the highest (lowest) \( I_{\text{SM:H}} \). The seasonal dynamics are almost...
similar across different models, i.e., weaker coupling over the Northern Hemisphere in SON and DJF (as the left-skewed distributions in these seasons suggest) and stronger coupling over the Northern Hemisphere in MAM and JJA. Almost all models show a weak terrestrial coupling index over the Sahara, Arabian, Gobi, and Kalahari Deserts due to the low temporal variability of surface fluxes in these regions. Comparing the maps of $I_{SM:LE}$ from observations (Fig. S3) with those of models, reveals the similarities between FLUXCOM and FIM in terms of magnitude and spatial patterns. However, GEFS, GOES, GEFS, and CFSv2.

![Observation vs Observation](image1)
![Model vs Model](image2)
![FLUXCOM vs Model](image3)
![GEWEX-PM-Mu vs Model](image4)

**FIG. 9.** As in Fig. 3, but for the terrestrial coupling index $I_{SM:LE}$.

**FIG. 10.** As in Fig. 4, but for the terrestrial coupling index $I_{SM:LE}$.
and CFSv2 tend to consistently overestimate $I_{SM:H}$ especially in the Southern Hemisphere, also in other regions (e.g., North America, most parts of Asia) in MAM and JJA.

Figure S5a shows the spatial correlation between the maps of $I_{SM:H}$ from FLUXCOM and GEWEX. Similar to $R(SM, H)$, there is a low degree of consistency between the spatial patterns of $I_{SM:H}$ from these two observational products. Figure S5b shows the spatial correlation between the maps of $I_{SM:H}$ from pairings of the five models for four seasons. The stratification in Fig. 7b and Fig. S5b is almost identical, as the pairing of CFSv2 and GEFS (both from NCEP) shows the highest spatial consistency while the pairings of CCSM4 and these two models show the lowest spatial correlation. Similar to $R(SM, H)$, GEFS and CFSv2 show the lowest spatial correlation with other models in the summer due to inconsistent estimates of $I_{SM:H}$ at high latitudes including Russian and Canada. The spatial comparisons between FLUXCOM (GEWEX) and the models are shown in Fig. S5c (Fig. S5d). The models show higher spatial correlation with observational products in the winter but, in general, no single model consistently outperforms any other.

d. Temporal characteristics of observationally based flux products

Returning to the issue of vastly different temporal variance among the observationally based flux products, example time series and climatologies are shown for LE in Fig. 11. As an example comparison to in situ measurements, the grid cell containing the Central Facility for the Atmospheric Radiation Measurement (ARM) Southern Great Plains network is chosen, and daily mean in situ fluxes from the U.S. ARM site from FLUXNET2015 (Biraud et al. 2003). Time series from the warm season of several years are shown as anomalies from the climatological annual cycle of each product.

It is evident that there are significant differences in the character of the time series, which affect not only the standard deviations but the temporal correlations calculated from these time series. A part of differences between flux tower measurements and observationally based products may be explained by the discrepancy in spatial scale between the product grid boxes and the flux tower footprint, but there are other factors that are contributing to these differences.

Compared to the observed time series, GEWEX–PM-Mu, and especially FLUXCOM, show low temporal variations, while GLEAM and GEWEX–PT-JPL show better agreement in terms of magnitude of standard deviation with the flux tower measurements, even though they differ in dynamics and detecting peaks. The very low variability of FLUXCOM is due to the fact that the FLUXCOM RS+METEO setup uses the mean seasonal cycles of MODIS land products (Jung et al. 2019). Therefore, LE anomaly time series is not a reflection of the daily vegetation dynamic, and its temporal variability is exclusively originated by daily variations of meteorological forcing. Also striking is that only GEWEX–PM-Mu replicates the observed spring peak in LE—other satellite-based products, contrary to in situ observations from FLUXNET, place the peak in summer. At this time, it is not known how widespread such inconsistencies are. Thorough
5. Discussion and conclusions

Despite the well-demonstrated, important role of land–atmosphere (LA) interactions in the Earth and climate system, the lack of long-term globally distributed observations on one hand, and the contrasts among numerical models, on the other hand, hinder a robust, realistic characterization of the LA coupling strength on a global scale.

In this study, we have examined LA coupling strength, specifically the terrestrial leg of LA coupling. We have used several recently produced global gridded observationally based analyses of surface heat fluxes and random error corrected soil moisture to estimate the distribution and seasonality of the terrestrial leg of LA coupling. Furthermore, we have undertaken a comprehensive analysis to identify the controlling effects of soil moisture variations on the surface fluxes of latent heat and sensible heat. We used two correlation-based metrics (i.e., the Pearson’s product-moment correlation coefficient, and the terrestrial coupling index) to measure the strength and investigate spatial patterns and seasonal dynamics of the terrestrial coupling on a global scale. The foundation of our study is formed using four global gridded observationally based surface heat flux products (FLUXCOM, GLEAM, and two from GEWEX–IP) and one satellite-based soil moisture product (ESA CCI). Last, we have confronted five weather and climate models (GEOS, FIM, CFSv2, GEFS, CCSM4) with the observations in order to benchmark their performance in representing the soil FIM, CFSv2, GEFS, CCSM4) with the observations in order to benchmark their performance in representing the soil

1) LA coupling determined from observations

- The regional and seasonal variations of the SM–LE coupling are quite similar in the different observational products. Patterns of temporal correlations (Figs. 2 and 5) are typically very similar among products (Fig. 3a), instilling confidence in these estimates. However, the differing temporal variance of the daily time series of surface fluxes (Fig. 11) leads to very different terrestrial coupling indices (Fig. 8), and weaker correspondence (Fig. 9e).
- Among the three observational products, GLEAM does not provide estimates for sensible heat flux. The comparison between FLUXCOM and GEWEX–IP sensible heat flux products, considering the physics of the system and the reported low quality of GEWEX–IP unconstrained sensible heat flux products (Siemann et al. 2018), gives us greater confidence in SM–H coupling metrics from FLUXCOM as the basis for evaluating models. Consequently, we believe our conception of SM–H coupling resulting from only one observationally based dataset (i.e., FLUXCOM) is less confident than that of SM–LE coupling resulting from among four observationally based datasets.

2) LA coupling in models

- The model assessment shows that GEFS, CFSv2, and GEOS generally overrepresent the connection between SM and LE, while FIM and CCSM4 are too weak. In terms of spatial patterns, GEFS, CFSv2, and GEOS show a high degree of consistency with one another and with the observations while FIM and particularly CCSM4 do not, suggesting FIM and CCSM4 are in a different category of quality compared to GEFS, CFSv2, and GEOS. CCSM4 shows the highest spatial correlation with other models in JJA and DJF. It is speculated spurious periodicity in vegetation phenology that arises at the start of the dry season over regions that have large amounts of C4 grass (Dirmeyer et al. 2013b) contributes to this peak.
- For SM and $H$, the comparison reveals similar seasonal dynamics and comparable spatial patterns (especially during MAM, SON, and DJF), but remarkably different magnitudes of the SM–$H$ coupling. The low spatial correlation between GEFS/CFSv2 and other models in JJA is probably due to the unreasonable overestimation of the SM–$H$ coupling from these models at high latitudes. The comparison between models and FLUXCOM reveals that GEFS overestimates the SM–$H$ coupling in all seasons over the globe, while CCSM4 underestimates it. Models show higher spatial consistency with FLUXCOM in MAM and DJF but, in general, no single model consistently outperforms any other.
- We have not seen a systematic difference in the performance of weather prediction models versus climate/seasonal prediction models. CCSM4 (a climate/seasonal prediction model) and FIM (a weather prediction model) show a weaker coupling between SM and LE than observations and the lowest degree of consistency in terms of spatial patterns with observations. On the other hand, GEFS (a weather prediction model) and GEOS and CFSv2 (climate/seasonal prediction models) show a stronger coupling between SM and LE than observations and a high degree of spatial consistency with observations. We also have not seen a clear difference in the performance of operational models (GEFS and CFSv2) versus research models (GEOS, FIM, and CCSM4). In this study, we have evaluated the performance of only five models, therefore such a conclusion based limited number of models might not be robust.

3) Global patterns and seasonal variability of LA coupling

- The global spatial patterns of the SM–LE coupling are more consistent across observations than across the models. Considering the disparities in the production of the various observational datasets, this strongly indicates greater uncertainty and suggests lower fidelity for some models. The areas with strong SM–LE coupling from both observations and models align well with known hot spots of land–atmosphere coupling from other studies (e.g., Koster et al. 2006; Guo et al. 2006; Koster et al. 2004) including seasonal variations. The stronger SM–flux coupling from both observations and models over the Southern Hemisphere compared to the Northern
Hemisphere is consistent with dominant temporal SM–precipitation feedbacks found in the CMIP5 GCMs (Moon et al. 2019). More research is needed to delve into details of the differences among the various datasets, their regional and seasonal dependencies, and how best to apply the information to improve the representation of the physical processes involved in the terrestrial leg of LA coupling.

- The two coupling metrics notably differ in detecting the SM–LE coupling over desert areas, where SM and LE are strongly correlated, but temporal variations of LE are small and therefore the terrestrial coupling index is low.

The results highlight the impact of SM on surface fluxes of latent and sensible heat and provide an advanced benchmark for the assessment of such relationships in weather and climate models. However, this study is limited by several factors. As mentioned earlier, CCI soil moisture is representative of water in shallow soil, and plant water content where vegetation canopies are thick, while some of the SubX model hindcasts report total soil column water (tcw). Even though our investigations show the coupling strength in GEOS (calculated from tcw) is found to be comparable with those of GEFS and CFSv2 models (calculated from a shallow top layer), and the coupling strength in GEFS, regardless of SM depth, is much stronger than observations, a possible reason for relatively low LA coupling from the CCSM4 and FIM models may be the influences of different SM depths. A fairer assessment of models could be had with more output fields from the models. The relative measurement error in the satellite-based SM data product (shown in Fig. 1) is high over densely vegetated and very dry regions. This hampers an authentic comparison between models and observations in these particular regions. Further comparison between satellite and in situ based LA metrics could be used to inform applicability and improve quality of global observational coupling assessments.

In this study, we investigated the LA coupling on a daily time scale. Soil moisture–surface flux relationships may vary with temporal scales as a longer time scale acts like a low-pass time filter and removes higher frequencies from consideration. The LA coupling has been investigated at multiple time scales from subdaily to seasonal (Wei and Dirmeyer 2012; Koster et al. 2003; Zhang et al. 2008). For instance, Zeng and Yuan (2018) found the coupling between SM and LE is more significant as time scales increase from daily to monthly. Evaluation of LA coupling at a longer time scale using the observationally based products introduced in this study would be useful for monthly–seasonal time scale given SM is the key land parameter that controls the monthly to seasonal variability and affects the predictability of the atmosphere.

While we see a great promise from the use of global gridded observational products to discern global patterns and seasonal variations in land–atmosphere coupling, such products should not be used “off the shelf.” It is evident that careful assessment and validation of each product is needed to boost confidence. Although outside the scope of this study, information from multiple data sources could be brought to bear to improve quality and reduce uncertainty. Methods such as triple collocation that was used to produce the blended CCI soil moisture product (Gruber et al. 2017) could be used to blend and improve flux products. Duration of the datasets is also an issue, especially for the GEWEX products; but as noted by Findell et al. (2015), 10–12 years is likely adequate for such an assessment, especially when the random noise in the soil moisture retrievals is accounted for.

There is also an intrinsic limitation of the two correlation-based statistical metrics; correlation only reveals the linear component of any relationship between two variables and cannot imply causation. Therefore, further studies applying approaches such as information-theory-based approaches that are useful for finding linear and nonlinear relationships between variables and are suitable for the analysis of causal relationships in nonlinear systems, would be helpful for the better understanding of land–atmosphere interaction and evaluation of weather and climate models.

This is a pathfinder study—as global observationally based datasets for LA coupling metrics mature and improve, their utility will increase. We hope to motivate model developers to begin considering validation, calibration, and ultimately development using such newly emerging observationally based datasets on continental to global scales. Extension of model validation beyond the limited distribution of in situ soil moisture and flux sites as satellite-based datasets become more reliable will be a boon to improving weather and climate model fidelity.

Acknowledgments. This research has been supported by National Aeronautics and Space Administration (NASA) Grant NNX14AM19G. We thank Kathy Pegion for her help with SubX datasets. All datasets used in this study are publicly available: FLUXCOM at http://wwwfluxcomorg; GLEAM at https://www.gleam.eu; GEWEX IP at ftp://rain.atmos.colostate.edu/pub/pbrown/GEWEX_IP_2019; ESA CCI SM at https://www.esa-soilmoisture-cci.org/; SubX hindcasts at https://iridl.ldeo.columbia.edu/SOURCES/.Models/SubX; S2S reforecasts at http://apps.ecmwf.int/datasets/data/s2s; and FLUXNET2015 at https://fluxnet.org doi/FLUXNET2015/.

REFERENCES


