**LETTER • OPEN ACCESS**

# Diagnosing the representation of surface and layered soil moisture in Earth system models

To cite this article: Elias C Massoud *et al* 2026 *Environ. Res. Lett.* **21** 024032

View the article online for updates and enhancements.

## You may also like

# ENVIRONMENTAL RESEARCH
## LETTERS

**LETTER**

# Diagnosing the representation of surface and layered soil moisture in Earth system models

Elias C Massoud[1,]*, Nathan O Collier[1], Min Xu[1], Mingjie Shi[2] and Forrest M Hoffman[1]

[1] Integrated Computational Earth Sciences Group, Oak Ridge National Laboratory, 1 Bethel Valley Rd, Oak Ridge, TN 37830, United States of America
[2] Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA 99354, United States of America
* Author to whom any correspondence should be addressed.

E-mail: massoudec@ornl.gov

## Abstract

Surface soil moisture (*mrsos*) and vertically integrated soil moisture (*mrsol*) over the top 10 cm should, by definition, be physically consistent in Earth System Models (ESMs). However, an evaluation of nine CMIP6 models reveals substantial inconsistencies: in some models, *mrsos* and integrated *mrsol* agree globally; in others, they align only in specific regions; and in a few, they diverge across all grid cells. These discrepancies arise from a combination of factors, including metadata errors, inconsistent variable definitions, or diagnostic sequencing within the model. We demonstrate how such issues can lead to significant biases, even when both variables are present and seemingly well-defined. As model complexity increases and multi-model comparisons become more common, assumptions about variable equivalence may lead to flawed conclusions. This study highlights the need for routine consistency checks, improved metadata standards, and community-wide practices that ensure reliability of derived variables across ESM outputs, particularly in preparation for CMIP7.

## 1. Introduction

In Earth System Models (ESMs), soil moisture (SM) plays a crucial role in regulating land-atmosphere interactions (Zhou *et al* 2019), including the energy balance (Sedighi *et al* 2016), evapotranspiration (Purdy *et al* 2018), and the hydrologic cycle (Western *et al* 2002, Massoud *et al* 2020). In agriculture, it impacts crop growth and yield by affecting soil aeration (Jin *et al* 2023) and root water uptake (Goet *et al* 2024). In hydrology, it influences runoff generation (Singh *et al* 2021), infiltration (Daly *et al* 2005), and groundwater recharge (Zhang *et al* 1999). In the cryosphere, it plays a crucial role in the water and energy balance, as it influences thermal insulation, energy fluxes, hydrological processes and feedback mechanisms (Hageman *et al* 2016). For drought monitoring, SM is used to assess drought conditions (Sehgal *et al* 2021) and predict agricultural impacts (Champagne *et al* 2019). In Earth system studies, changes in SM are analyzed to understand its effects on water resources (Gleick 1989, Massoud *et al* 2020) and ecosystem

dynamics (Holsten *et al* 2009, Massoud *et al* 2022). For hydrological forecasting, accurate representation of SM improves hydrological models used for flood forecasting (Wanders *et al* 2014) and water resource management (Long *et al* 2019). Therefore, understanding and accurately modeling SM is essential for predicting and managing environmental changes across the globe.

Observing and measuring SM globally has historically been difficult, with high uncertainties being present between different data products (Wanders *et al* 2012, Massoud *et al* 2021, Wang *et al* 2021, 2023). Remote sensing products from various satellite missions like SMOS (SM and Ocean Salinity) and SMAP (SM Active Passive) provide SM data through microwave remote sensing (Kerr *et al* 2016, Suman *et al* 2020). Additionally, *in situ* networks, such as the International SM Network, and flux networks, such as FLUXNET, offer valuable ground-based measurements, enhancing the accuracy and reliability of SM observations (Pan *et al* 2019, Dorigo *et al* 2021). A recent study evaluated various satellite

and model-based SM products using *in-situ* measurements, and provided a ranking of skill for the various products considered (Beck *et al* 2021). Although these different data products contribute to a better understanding of SM dynamics, there are still high uncertainties in the quantification of SM globally, making it particularly difficult to benchmark and evaluate model simulations of SM.

Historically, the representation of SM in ESMs has been inconsistent (Clark *et al* 2015), with this variability contributing significantly to uncertainties in SM estimation (Cheng *et al* 2017, Massoud *et al* 2020, 2023). This inconsistency in simulated SM affects land-atmosphere interactions, influencing surface fluxes and boundary layer characteristics, which in turn amplifies uncertainties (Huang and Margulis 2013, Lee and Hohenegger 2024). As a result, comparing and benchmarking model simulations of SM becomes challenging (Qiao *et al* 2022, Wang *et al* 2022). ESMs simulate SM using different soil-vegetation-atmosphere transfer schemes (Avissar 1998) which are influenced by factors such as soil texture, vegetation cover, precipitation, and temperature. Additionally, variations in model resolution, soil layer structures, and parameterizations further contribute to discrepancies in simulated SM. Differences in the representation of soil properties and hydrological processes across models lead to significant uncertainties in SM tracking and quantification (Mirus 2015, Massoud *et al* 2020).

By standardizing the definition and measurement of SM, the 6th Coupled Model Intercomparison Project (CMIP6) aims to provide consistent and comparable data across different models, facilitating SM-related analyzes and projections. This is directly possible with the definition of the surface SM variable, *mrsos*, which models have defined as the average SM content in the top 10 cm of the soil. However, when more detailed vertical information of SM is required, one must utilize the layered SM variable, *mrsol*, which is defined differently in most models. This discrepancy in the definition of layered SM in models is problematic for studies that aim to investigate deeper layers in the soil profile (see Qiao *et al* 2022, Wang *et al* 2022), and warrants further investigation.

In this paper, we evaluate the consistency between surface SM (*mrsos*)—which is defined by CMIP6 as the average SM over the top 10 cm, and the equivalent value computed by integrating the layered SM variable (*mrsol*) to 10 cm depth. If both variables follow protocol and represent the same physical quantity, they should be numerically equivalent. However, we find that this is not the case for many models. Our analysis reveals a spectrum of discrepancies: in some models, *mrsos* and integrated *mrsol* agree closely; in others, they diverge in specific regions or seasons; and in a few, they differ globally. These differences reflect a range of potential issues, from metadata inconsistencies and post-processing errors to model-specific implementations of soil hydrology. We argue that such discrepancies, though sometimes subtle, can propagate into multi-model analyzes and introduce uncertainty into downstream applications. By documenting and diagnosing these issues, our study highlights the importance of transparent documentation, robust metadata, and standardized quality control procedures across model development teams. The lessons presented here extend beyond SM and serve as a broader call for the community to scrutinize derived variables in ESMs, particularly when those variables are reported alongside their source components.

## 2. Materials and methods

### 2.1. Surface and layered SM in CMIP models

In CMIP6 models, the variable *mrsos* represents the surface SM, defined as the total mass of water in the top 10 cm of the soil (typically expressed in units of kg m$^{-2}$). This variable captures the amount of water stored in the uppermost soil layer, which is crucial for understanding surface-atmosphere interactions, plant water availability, and other hydrological processes. In CMIP6 models, *mrsol* specifically denotes the layered SM content over a defined depth range. This variable is essential for modeling soil water dynamics within ESMs, contributing crucial information to simulations of land surface processes, such as evapotranspiration (ET), runoff, and SM availability. The estimation of *mrsos* depends on *mrsol*, and is typically defined as the integration of *mrsol* in the top 10 cm of the soil profile. Model outputs of *mrsol* are used in some research applications (Qiao *et al* 2022, Wang *et al* 2022), however, other studies have used the *mrsos* variable (Gu *et al* 2019, Qiao *et al* 2022, Wang *et al* 2022), since SM information in the top 10 cm is usually standardized between different products. In this study, we use both *mrsos* and *mrsol* outputs from the models, at monthly temporal resolution.

*The mrsos variable* is shown in figure 1, depicting surface SM values for a selection of CMIP6 models: ACCESS-ESM1-5, BCC-CSM2-MR, CESM2, CNRM-ESM2-1, CanESM5, MIROC6, MPI-ESM1-2-LR, NorESM2-LM, and UKESM1-0-LL. Table 1 lists the models, their latitude and longitude grid sizes, land model used, and the depth of the soil layers for both the *mrsos* and *mrsol* variables. For consistency, we used the first available ensemble member variant for each model (e.g. r1i1p1f1). These specific models were selected to illustrate the various outcomes of these discrepancies. In other words, we selected these models to show how some might have no discrepancy between *mrsos* and *mrsol* to 10 cm, some might have it in certain grid cells, while a selection of models might have this bias globally. Our aim
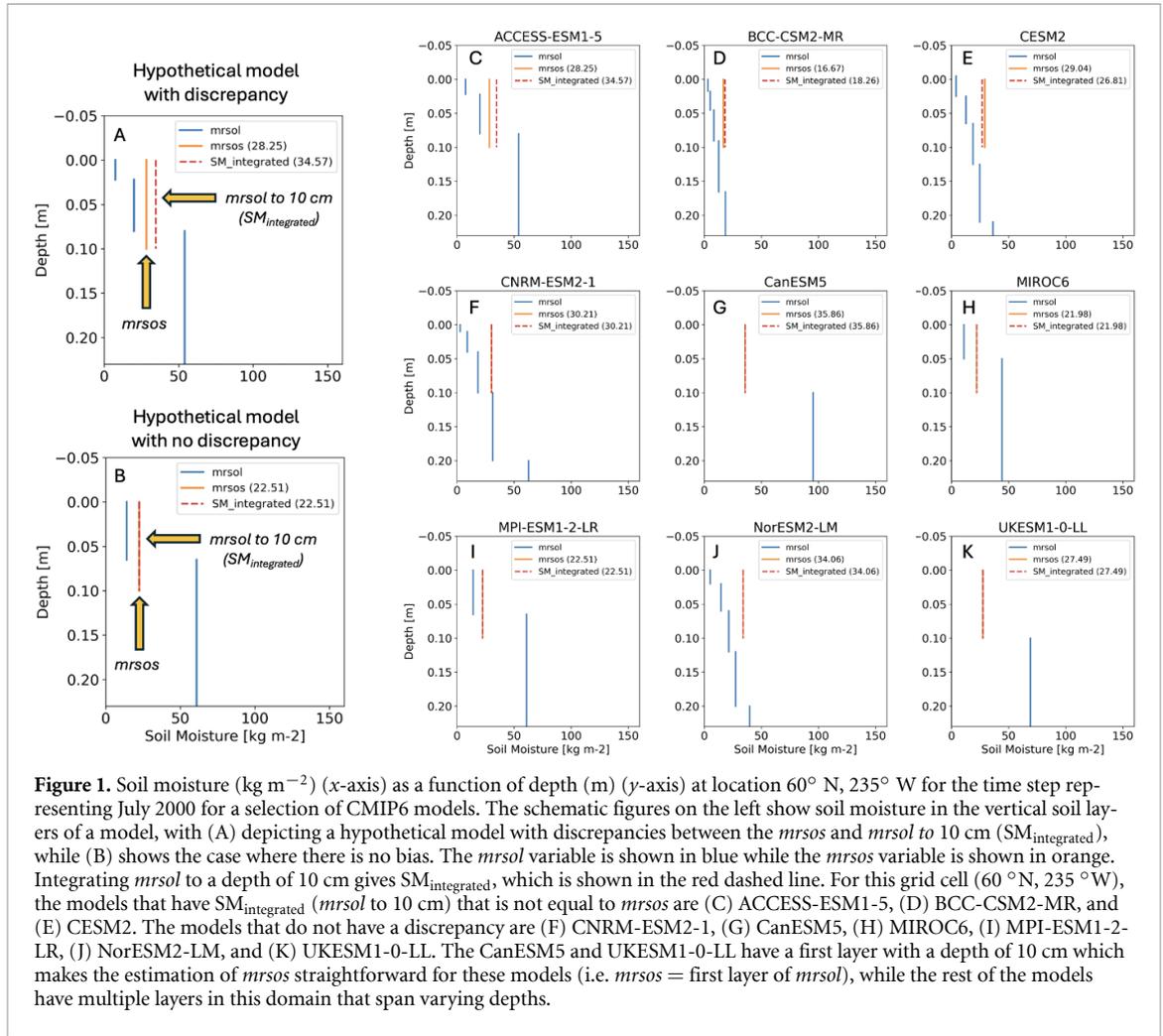
**Figure 1.** Soil moisture (kg m$^{-2}$) (*x*-axis) as a function of depth (m) (*y*-axis) at location 60° N, 235° W for the time step representing July 2000 for a selection of CMIP6 models. The schematic figures on the left show soil moisture in the vertical soil layers of a model, with (A) depicting a hypothetical model with discrepancies between the *mrsos* and *mrsol to* 10 cm (SM$_{integrated}$), while (B) shows the case where there is no bias. The *mrsol* variable is shown in blue while the *mrsos* variable is shown in orange. Integrating *mrsol* to a depth of 10 cm gives SM$_{integrated}$, which is shown in the red dashed line. For this grid cell (60 °N, 235 °W), the models that have SM$_{integrated}$ (*mrsol* to 10 cm) that is not equal to *mrsos* are (C) ACCESS-ESM1-5, (D) BCC-CSM2-MR, and (E) CESM2. The models that do not have a discrepancy are (F) CNRM-ESM2-1, (G) CanESM5, (H) MIROC6, (I) MPI-ESM1-2-LR, (J) NorESM2-LM, and (K) UKESM1-0-LL. The CanESM5 and UKESM1-0-LL have a first layer with a depth of 10 cm which makes the estimation of *mrsos* straightforward for these models (i.e. *mrsos* = first layer of *mrsol*), while the rest of the models have multiple layers in this domain that span varying depths.

in this paper is to show multiple models from each category of bias, resulting in the 9 selected models. The models shown here with minimal or no global bias between *mrsos* and *mrsol* (figure 2) are CanESM5, CNRM-ESM2-1, and MIROC6; models with regionally confined or grid-level bias (figure 3) are MPI-ESM1-2-LR and NorESM2-LM; and models with widespread discrepancies globally (figure 4) are ACCESS-ESM1-5, BCC-CSM2-MR, and CESM2. The plots in figure 1 display values at a representative grid cell located at 60 °N, 235 °W, and include both *mrsos* and the integrated *mrsol* to 10 cm depth (referred to as SM$_{integrated}$, as defined in the section 2.2). In this grid cell, several models show agreement between *mrsos* and SM$_{integrated}$, while others diverge. In the following section, we extend this comparison to the global scale.

## 2.2. Calculating the discrepancies

We utilized SM data from various CMIP6 models to compare the *mrsos* output with our own calculation of *mrsol* integrated in the top 10 cm (SM$_{integrated}$). Similar to Qiao *et al* (2022) and Wang *et al* (2022), SM in the top 10 cm is computed as the integrated SM

content (SM$_{integrated}$ in equation (1)) down to a depth of 10 cm, calculated using the following integration:

$$\text{SM}_{\text{Integrated}} = \left[ \sum_{\text{layer } i=1}^{\text{layer } n-1} mrsol(i) \right] + \left[ \frac{mrsol(n)}{\text{d}z(n)} * z_{\text{remaining}} \right] \quad (1)$$

where the soil profile of the given model consists of *n* layers, *mrsol*(*i*) represents the SM value in the *i*th layer (in units kg m$^{-2}$), *mrsol*(*n*) corresponds to the SM in the final layer that extends to include the 10 cm depth, d$z$(*n*) is the thickness of this final layer, and *z_remaining* denotes the portion of the final layer's depth required to reach exactly 10 cm (expressed in meters). For the cases where *mrsos* is truly equivalent to the vertical integration of *mrsol* to 10 cm (i.e. *mrsos* = SM$_{integrated}$, and the BIAS$_{\text{SM top 10 cm}}$ in equation (2) has non-zero values) then the difference between the two should be zero,

$$\text{BIAS}_{\text{SM top 10 cm}} (\text{lat, lon, time})$$
$$= mrsos (\text{lat, lon, time})$$
$$- \text{SM}_{\text{Integrated}} (\text{lat, lon, time}). \quad (2)$$

**Table 1.** CMIP6 models used in this study, along with their latitude and longitude grid sizes and the land model used for that model. The 'X' marks whether a model has a difference in the SM calculation and indicates if it is global or only in certain areas. The depth of the soil layers in each model are also listed, for both the *mrsos* and *mrsol* variables.

| Model | Lat | Lon | Land model | Equal all grids | Equal some grids | Equal no grids | *mrsos* depth bands | *mrsol* depth bands |
|---|---|---|---|---|---|---|---|---|
| ACCESS-ESM1-5 | 145 | 192 | CABLE | | | X | 0.022[a] | 0.0, 0.022, 0.08, 0.234, 0.643, 1.72, 4.6 |
| BCC-CSM2-MR | 160 | 320 | BCC-AVIM2 | | | X | 0.1 | 0.0, 0.017 512 82, 0.045 091 79, 0.090 561 82, 0.165 529 23, 0.289 1296, 0.492 912 14, 0.828 892 77, 1.382 831 16, 2.296 121 18, 3.433 093 01 |
| CanESM5 | 64 | 128 | CLASS | X | | | 0.1 | 0.0, 0.1, 0.35, 4.1 |
| CESM2 | 192 | 288 | CLM5 | | | X | N/A | −0.005, 0.025, 0.065, 0.125, 0.21, 0.33, 0.49, 0.69, 0.93, 1.21, 1.53, 1.89, 2.29, 2.745, 3.285, 3.925, 4.665, 5.505, 6.445, 7.485, 8.575 |
| CNRM-ESM2-1 | 128 | 256 | ISBA | X | | | 0.1 | 0.0, 0.01, 0.04, 0.1, 0.2, 0.4, 0.8, 1, 1.5, 2, 3, 5, 8, 12 |
| MIROC6 | 128 | 256 | MATSIRO | X | | | 0.1 | 0.0, 0.05, 0.25, 1, 2, 4, 14 |
| MPI-ESM1-2-LR | 96 | 192 | JSBACH | | X | | 0.1 | 0.0, 0.065, 0.319, 1.232, 4.134, 9.834 |
| NorESM2-LM | 96 | 144 | CLM5 | | X | | 0.1 | 0.0, 0.02, 0.06, 0.12, 0.2, 0.32, 0.48, 0.68, 0.92, 1.2, 1.52, 1.88, 2.28, 2.72, 3.26, 3.9, 4.64, 5.48, 6.42, 7.46, 8.6 |
| UKESM1-0-LL | 144 | 192 | JULES-ES-1.0 | X | | | 0.1 | 0.0, 0.1, 0.35, 1.0, 3.0 |

[a] ACCESS-ESM1-5 does not provide a depth band of 10 cm for *mrsos*.

**Figure 2.** Models with no discrepancy between *mrsos* and *mrsol* to 10 cm (SM$_{integrated}$). The plots show the long-term mean simulations of (left) *mrsos*, (center) vertically integrated *mrsol* to a depth of 10 cm (SM$_{integrated}$), and (right) their difference for the CanESM5, CNRM-ESM2-1, MIROC6, and UKESM1-0-LL models. As shown in the right panels, the difference for these models is zero throughout the globe, indicating that the integration of *mrsol* to a 10 cm depth is all that is necessary to define the *mrsos* variable in CanESM5, CNRM-ESM2-1, MIROC6, and UKESM1-0-LL. Therefore, there is no discrepancy for these models. Some color bars include arrows, indicating that values may extend beyond the displayed range.
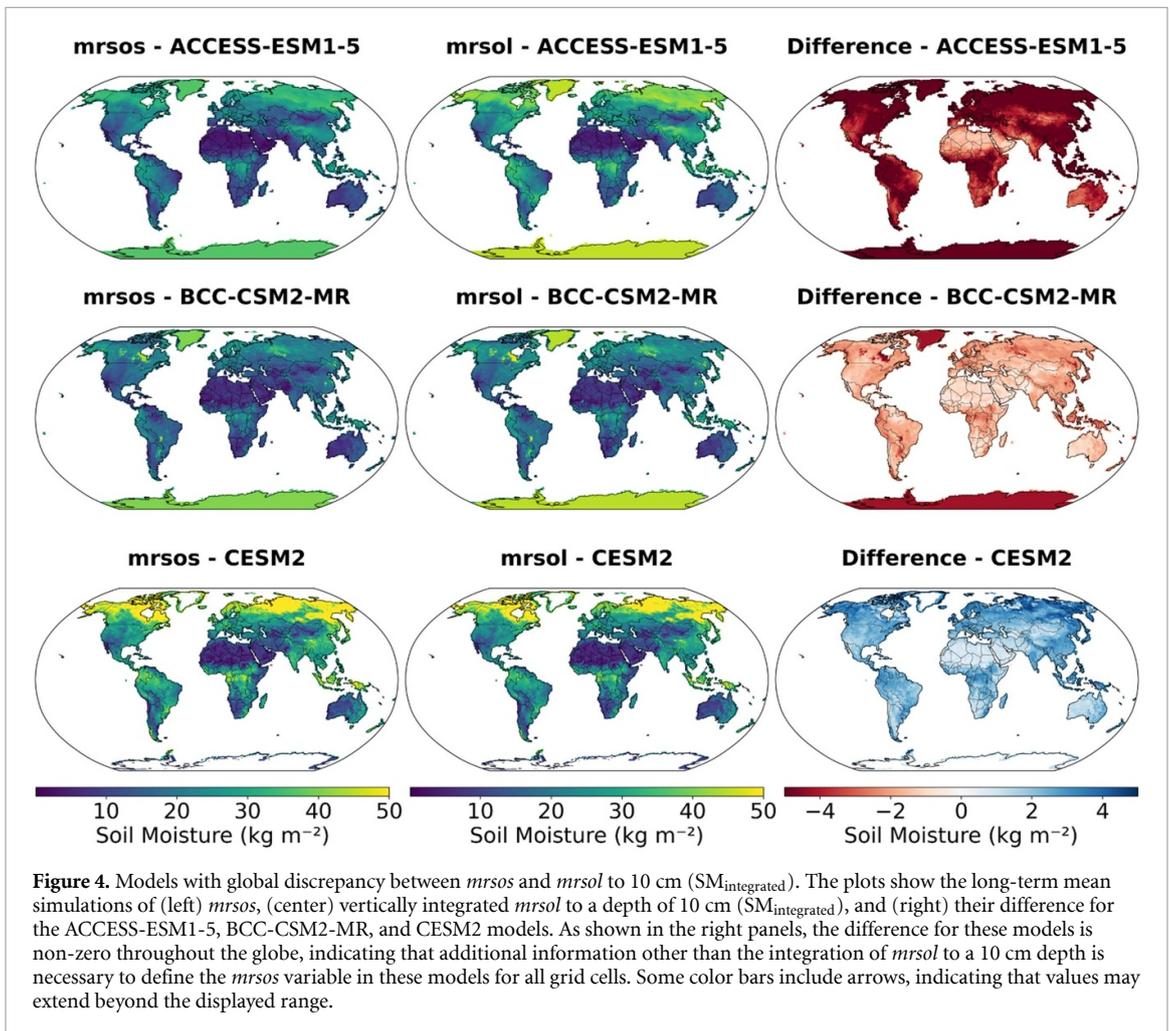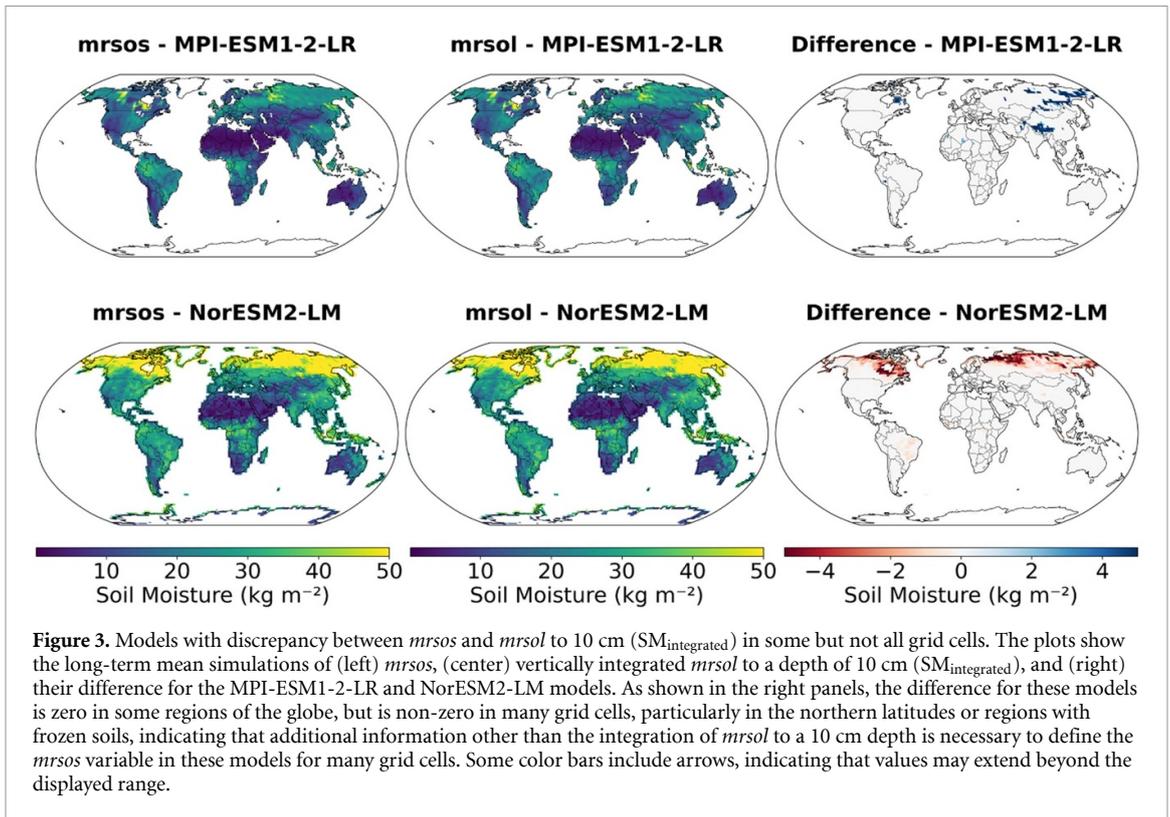
This outcome suggests that the *mrsos* variable was properly diagnosed from the model's layered SM and correctly processed according to CMIP standards. However, when this subtraction yields non-zero values, it signals a problem, either *mrsos* was calculated or processed incorrectly, or the layered *mrsol* values were modified after *mrsos* was diagnosed. In such cases, unknown or undocumented model-specific factors appear to influence the top 10 cm SM estimate. This makes it difficult to know which variable, *mrsos* or integrated *mrsol*, provides a more accurate representation of surface SM. Moreover, since this bias can vary spatially and seasonally, it complicates efforts to analyze or compare SM profiles across depths, models, and time. This dilemma is central to the present study and highlights the need for greater transparency and standardization in how SM variables are diagnosed and reported in ESMs.

## 3. Results

### 3.1. Inconsistent integration in CMIP6 models

We compared the vertically integrated SM (SM$_{integrated}$) with the surface SM (*mrsos*) variable across the CMIP6 models considered in this study. Looking at this in global simulations reveals an important variability in the consistency of this integration process among the models. For CanESM5, CNRM-ESM2-1, MIROC6, and UKESM1-0-LL (figure 2), the difference between SM$_{integrated}$ and *mrsos* is consistently zero across the globe. This finding suggests that the integration adequately defines the *mrsos* variable in these models without the need for additional information. In contrast, MPI-ESM1-2-LR and NorESM2-LM exhibit a mixed pattern (figure 3). While the difference between SM$_{integrated}$ and *mrsos* is zero in some regions of the globe for

**Figure 3.** Models with discrepancy between *mrsos* and *mrsol* to 10 cm (SM$_{integrated}$) in some but not all grid cells. The plots show the long-term mean simulations of (left) *mrsos*, (center) vertically integrated *mrsol* to a depth of 10 cm (SM$_{integrated}$), and (right) their difference for the MPI-ESM1-2-LR and NorESM2-LM models. As shown in the right panels, the difference for these models is zero in some regions of the globe, but is non-zero in many grid cells, particularly in the northern latitudes or regions with frozen soils, indicating that additional information other than the integration of *mrsol* to a 10 cm depth is necessary to define the *mrsos* variable in these models for many grid cells. Some color bars include arrows, indicating that values may extend beyond the displayed range.



**Figure 4.** Models with global discrepancy between *mrsos* and *mrsol* to 10 cm (SM$_{integrated}$). The plots show the long-term mean simulations of (left) *mrsos*, (center) vertically integrated *mrsol* to a depth of 10 cm (SM$_{integrated}$), and (right) their difference for the ACCESS-ESM1-5, BCC-CSM2-MR, and CESM2 models. As shown in the right panels, the difference for these models is non-zero throughout the globe, indicating that additional information other than the integration of *mrsol* to a 10 cm depth is necessary to define the *mrsos* variable in these models for all grid cells. Some color bars include arrows, indicating that values may extend beyond the displayed range.

these models, it is non-zero in many grid cells, particularly in northern latitudes (e.g. NorESM2-LM) or regions with perenially frozen soils such as the Tibet Plateau (e.g. MPI-ESM1-2-LR). For ACCESS-ESM1-5, BCC-CSM2-MR, and CESM2 (figure 4), the difference between $SM_{integrated}$ and *mrsos* is consistently non-zero across the entire globe. These results suggest that the definition of *mrsos* in these models requires additional information beyond simply integrating *mrsol* to a 10 cm depth.

These findings highlight the need for careful consideration of how *mrsos* is defined and interpreted across CMIP6 models. While there are models that demonstrate uniformity in their integration approach, other models highlight the variability and complexity in defining *mrsos*, particularly in regions with distinct SM dynamics, such as frozen soils. Addressing these inconsistencies is crucial for enhancing the reliability and comparability of SM data across different ESMs.

## 3.2. Model-specific investigation and analysis

To explore these possibilities, we conducted a deeper investigation of specific models, focusing particularly on CESM2, which uses CLM5 as its land model, and ACCESS-ESM1-5, which uses the CABLE. The next subsections detail these findings.

### 3.2.1. Additional CESM2 analysis: diagnosing the source of SM discrepancies

To further investigate the causes of discrepancies between *mrsos* and integrated *mrsol* ($SM_{integrated}$) for CESM2, we conducted a targeted analysis leveraging its open-source land model, CLM5. Using raw outputs from CESM2 and source code from CLM5 (the land model used by both CESM2 and NorESM2-LM) we identified two primary sources of bias. First, the CMIP6 CESM2 data on the Earth System Grid Federation (ESGF) includes incorrect depth bounds (depth_bnds) introduced during CMORization. In this model, soil is divided into 20 layers (totaling 8 meters), with defined layer boundaries from initVerticalMod.F90. For example, the lower bounds of the top four layers should be [0, 20, 60, 120 cm]. However, in the CMIP6 CESM2 data obtained from ESGF, the depth_bnds coordinate reports these bounds as [−5, 25, 65, 125 cm], indicating a misrepresentation introduced during CMORization. Replacing these with the correct values from CESM source code significantly reduces global *mrsos*–*mrsol* discrepancies, as seen in the CESM2fix comparison in figure 5(A). Figure 5(A) compares the original CESM2 output from ESGF with a corrected version using the depth bounds from the CLM5 source code ('CESM2fix'). The corrected depth_bnds significantly reduces global discrepancies, confirming that part of the bias stems from metadata misalignment. Second, we found that

layered SM (*mrsol*) is altered after *mrsos* is diagnosed, due to updates in the hydrology scheme that affect runoff computation. The source code inspection reveals that after the 10 cm SM (*mrsos*) is calculated, the underlying SM layers (*mrsol*) are updated in the HydrologyDrainageMod.F90 routine to compute subsurface runoff. As a result, *mrsol* no longer reflects the same state as when *mrsos* was diagnosed. This sequencing issue contributes to regional biases, particularly in high-latitude regions and during thaw seasons (figure 5(B)).

This analysis reveals that CESM2's apparent classification, as a model with widespread bias in figure 4, shifts to one with more regionally confined bias once the depth metadata is corrected (e.g. CESM2fix in figure 5(A)), making it similar to the models shown in figure 3. In fact, the spatial pattern of the discrepancy maps for CESM2fix in figure 5(A) becomes nearly identical to that of NorESM2-LM in figure 3, consistent with their shared use of CLM5. These findings highlight two main causes of bias in CESM2:

1. Incorrect layer bounds (depth_bnds) assigned during CMORization;
2. Post-diagnosis updates to layered SM that alter *mrsol* independently of *mrsos*.

Such in-depth diagnoses are only possible when source code and pre-CMOR raw data are available. For closed-source models, users may not be able to identify these discrepancies, reinforcing the need for modeling centers to implement stronger quality control protocols and provide metadata transparency.

### 3.2.2. Discrepancies in model layer definitions: the case of ACCESS-ESM1-5

An especially notable case is ACCESS-ESM1-5, which shows the largest and most widespread discrepancies between *mrsos* and integrated *mrsol* across the globe (figure 4). According to table 1, the vertical depth associated with the *mrsos* variable in this model is 2.2 cm, whereas most other models define this depth at 10 cm, corresponding to the nominal 0–10 cm layer (CESM2 is an exception, as its *mrsos* depth metadata is missing entirely). This difference in vertical representation suggests that ACCESS-ESM1-5 is reporting surface SM from a shallower layer than intended, which may partly explain the large biases observed in comparison with *mrsol* (figure 4).

Importantly, this issue has been acknowledged by the modeling group. An official erratum for the ACCESS-ESM1-5 *mrsos* variable was submitted to the CMIP6 community via the IPSL errata service on 29 November 2023 (https://errata.ipsl.fr/static/view.html?uid=73381777-6fab-0944-94d3-641cd9b07670). The report clarifies that *mrsos* was incorrectly calculated due to a misapplied scaling
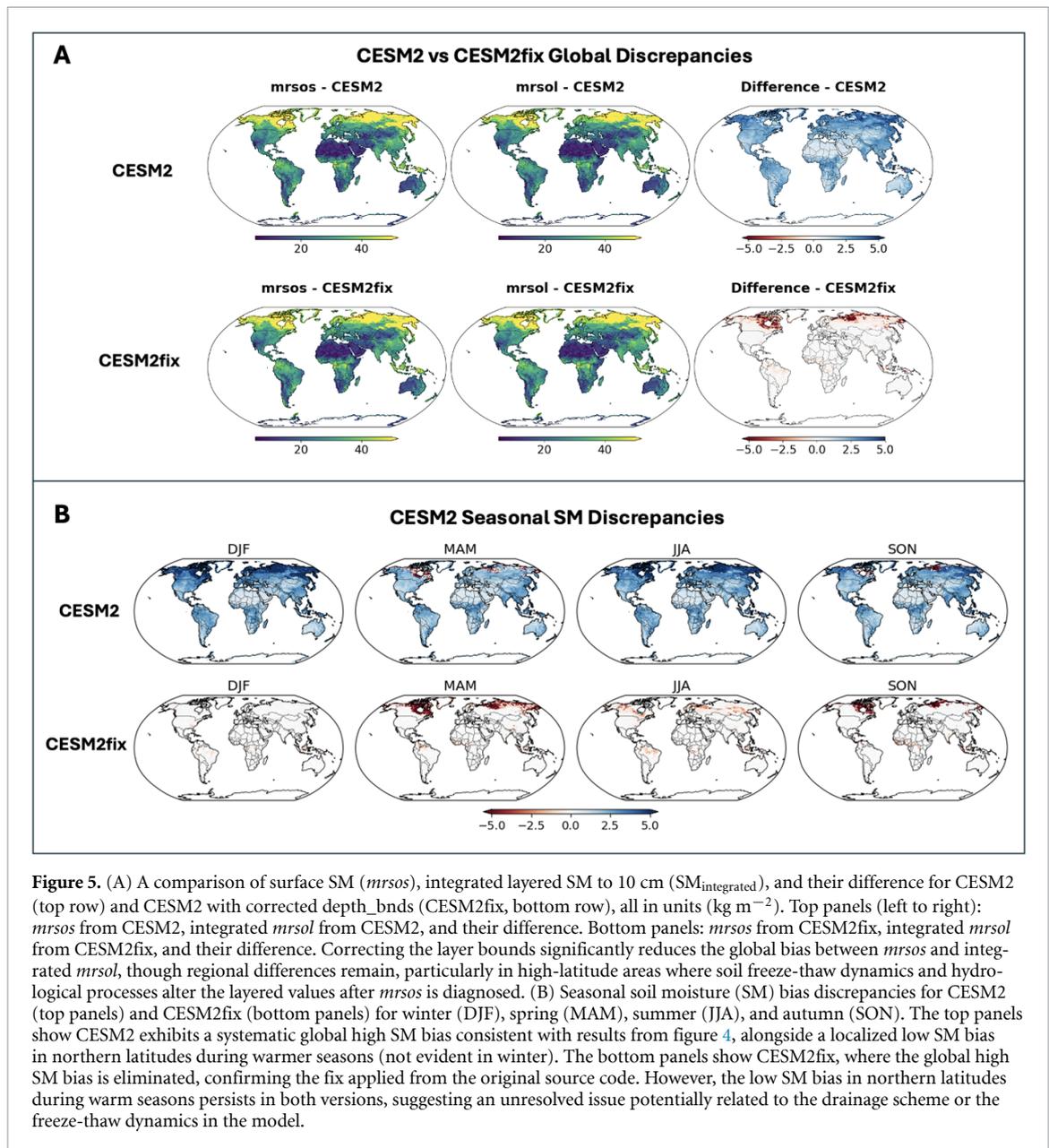
**Figure 5.** (A) A comparison of surface SM (*mrsos*), integrated layered SM to 10 cm (SM$_{integrated}$), and their difference for CESM2 (top row) and CESM2 with corrected depth_bnds (CESM2fix, bottom row), all in units (kg m$^{-2}$). Top panels (left to right): *mrsos* from CESM2, integrated *mrsol* from CESM2, and their difference. Bottom panels: *mrsos* from CESM2fix, integrated *mrsol* from CESM2fix, and their difference. Correcting the layer bounds significantly reduces the global bias between *mrsos* and integrated *mrsol*, though regional differences remain, particularly in high-latitude areas where soil freeze-thaw dynamics and hydrological processes alter the layered values after *mrsos* is diagnosed. (B) Seasonal soil moisture (SM) bias discrepancies for CESM2 (top panels) and CESM2fix (bottom panels) for winter (DJF), spring (MAM), summer (JJA), and autumn (SON). The top panels show CESM2 exhibits a systematic global high SM bias consistent with results from figure 4, alongside a localized low SM bias in northern latitudes during warmer seasons (not evident in winter). The bottom panels show CESM2fix, where the global high SM bias is eliminated, confirming the fix applied from the original source code. However, the low SM bias in northern latitudes during warm seasons persists in both versions, suggesting an unresolved issue potentially related to the drainage scheme or the freeze-thaw dynamics in the model.

factor in the third soil layer of the CABLE land model, leading to *mrsos* values that are approximately 20% too low. Although the erroneous data have been flagged, they are not being replaced due to resource limitations. Instead, users are advised to recomputed *mrsos* from layered SM values (*mrsol*) using a correction formula.

This case illustrates a broader issue: although *mrsos* is supposed to represent a standard 0–10 cm depth, its actual implementation can vary across models due to differences in depth metadata and post-processing workflows. Without consistent documentation and automated checks, such discrepancies can go unnoticed by downstream users. Moreover, while the errata system is an important mechanism for transparency, its effectiveness is limited when corrections are not propagated into updated datasets and

redistributed, for example to ESGF. The ACCESS-ESM1-5 example reinforces the need for stronger quality control in CMIP workflows and highlights the consequences of metadata misalignment, both for internal model fidelity and for multi-model comparison efforts like this one.

### 3.3. Impacts on real-world applications: land-atmosphere interactions

To explore how the discrepancies discussed in this paper would impact applications in Earth Science or other fields, we investigated how the use of *mrsos* or *mrsol* to 10 cm (SM$_{integrated}$) would affect results from an evaluation of a few select variables. We chose the ACCESS-ESM1-5 model for this example since it exhibited the largest apparent discrepancy (based on results in figure 4).
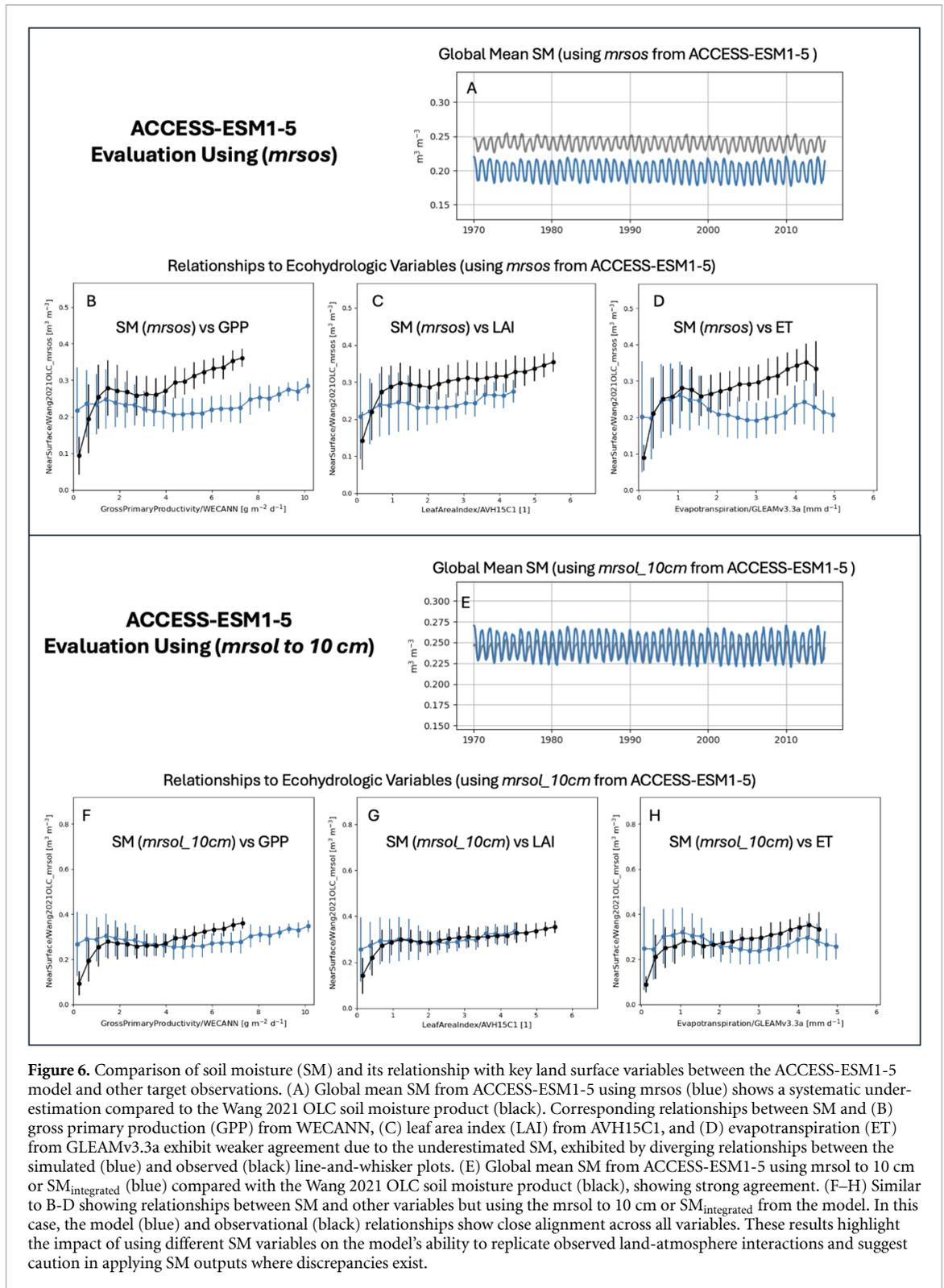
**Figure 6.** Comparison of soil moisture (SM) and its relationship with key land surface variables between the ACCESS-ESM1-5 model and other target observations. (A) Global mean SM from ACCESS-ESM1-5 using mrsos (blue) shows a systematic underestimation compared to the Wang 2021 OLC soil moisture product (black). Corresponding relationships between SM and (B) gross primary production (GPP) from WECANN, (C) leaf area index (LAI) from AVH15C1, and (D) evapotranspiration (ET) from GLEAMv3.3a exhibit weaker agreement due to the underestimated SM, exhibited by diverging relationships between the simulated (blue) and observed (black) line-and-whisker plots. (E) Global mean SM from ACCESS-ESM1-5 using mrsol to 10 cm or $SM_{integrated}$ (blue) compared with the Wang 2021 OLC soil moisture product (black), showing strong agreement. (F–H) Similar to B-D showing relationships between SM and other variables but using the mrsol to 10 cm or $SM_{integrated}$ from the model. In this case, the model (blue) and observational (black) relationships show close alignment across all variables. These results highlight the impact of using different SM variables on the model's ability to replicate observed land-atmosphere interactions and suggest caution in applying SM outputs where discrepancies exist.

First, we looked into how the global mean SM from the model would compare against that of a target product (in this case we chose the Wang 2021 OLC SM product from Wang *et al* 2021). Figure 6(A) shows global mean SM values in the model when *mrsos* is used (blue) compared to the target data (black), with a systematic underestimation of SM in the model. When looking at how the SM impacts land-atmosphere interactions and other vegetation processes, we plotted the relationship of SM to gross primary production (GPP), leaf area index (LAI), and ET, to explore how the relationship of one variable maps onto the other. The target products used here were WECANN for GPP (g m$^{-2}$ d$^{-1}$) (Alemohammad *et al* 2017), AVH15C1 for LAI (unitless) (Clevarie *et al* 2016), and GLEAMv3.3a for ET (mm d$^{-1}$) (Martens *et al* 2017). figures 6(B)–(D) show, again, an underestimation of the simulated

(blue) compared to the observed (black) relationships, exhibited by diverging line-and-whisker plots.

Conversely, figures 6(E)–(H) show similar plots to figures 6(A)–(D), but for when *mrsol* to 10 cm (SM$_{integrated}$) is used in the model, and we see that the values line up very closely in all subplots. Figure 6(E) shows the alignment in the global mean SM from the model (blue) and the target product (black), and figures 6(F)–(H) show the strong agreement between the considered relationships. Therefore, using *mrsos* in the model in this case caused a mismatch between the simulated dynamics and the target data product used to evaluate these dynamics, which was not the case when using *mrsol* to 10 cm (SM$_{integrated}$). For models with no discrepancies, the evaluation would be the same whether *mrsos* or *mrsol* to 10 cm (SM$_{integrated}$) is used, however in cases where discrepancies do exist (such as the case for ACCESS-ESM1-5), caution must be taken when applying the SM values in real-world applications, such as evaluating the model's skill in simulating SM and its impact on land-atmosphere interactions.

## 4. Explaining the discrepancies

### 4.1. Possible factors for discrepancies in CMIP6 models

The discrepancies between the surface SM (*mrsos*) and the vertically integrated layered SM to 10 cm (*mrsol* to 10 cm or SM$_{integrated}$) across different CMIP6 models can be attributed to several factors. One key reason is related to incorrect variable mapping between the raw model outputs and the standardized CMIP6 SM variables. If the translation from native model variables to CMIP6 conventions is not handled consistently (Thor *et al* 2023), this could result in mismatches between *mrsos* and the integration of *mrsol*. Another possibility involves interpolation errors that occur when layered SM is aggregated to represent the top 10 cm. Bugs in the models themselves or in the post-processing codes used to prepare CMIP6 outputs could introduce discrepancies (Feng *et al* 2022), particularly in models with complex soil layering structures. Additionally, misinterpretations of the variable definitions could contribute to these differences. For example, the exclusion of ice content in the calculation of SM, or the unintended inclusion of factors like surface snow or drainage water, could lead to inconsistencies between *mrsos* and the integrated *mrsol*. This is the reason for excluding frozen grid cells in various SM products (Reichle *et al* 2019, Massoud *et al* 2023). Notably, our results, shown in figure 3, reveal that MPI-ESM1-2-LR and NorESM2-LM (as well as the CESM2fix plots in figure 5) display a distinct bias in cold regions, where SM may be frozen. This suggests that it is essential to account for both liquid and frozen water when estimating SM in different layers.

Beyond these data and processing issues, structural differences in model design also play a major role. The number and thickness of soil layers, as well as the hydrological schemes that govern vertical and lateral water movement, can strongly influence the consistency between *mrsos* and *mrsol*. In models where the uppermost soil layer is exactly 10 cm thick (e.g. CanESM5 and UKESM1-0-LL), *mrsos* and *mrsol* (integrated to 10 cm) are typically consistent and exactly equal to each other, as no vertical interpolation is required. However, when the uppermost layer is thinner or thicker than 10 cm, interpolation or partial-layer weighting introduces numerical errors that can propagate into systematic biases. These structural differences thus determine whether *mrsos* is directly comparable to *mrsol*.

To further investigate the source of these discrepancies, we examined the CLM5 land model source code, which underlies CESM2 and NorESM2-LM, and identified a specific process-based cause for the negative bias observed in these models. In CLM5, *mrsos* is calculated before execution of the *HydrologyDrainage* module, which implements numerical adjustments to maintain water conservation across soil layers. According to the CLM Technical Note (https://escomp.github.io/CTSM/tech_note/Hydrology/CLM50_Tech_Note_Hydrology.html), this module redistributes excess liquid water from deeper layers upward and removes remaining excess through drainage. These adjustments increase water content in upper layers of *mrsol* after *mrsos* has already been diagnosed, leading to a systematic underestimation of *mrsos* relative to the updated *mrsol*. This explains the negative *mrsos*–*mrsol* bias found in CESM2 and NorESM2-LM.

Furthermore, frozen-soil processes may amplify these inconsistencies. The *HydrologyDrainage* module also simulates lateral drainage from perched saturated zones that can form above frozen layers (section 2.7.4 of the CLM Technical Note). When soils freeze, hydraulic conductivity decreases sharply due to the ice impedance factor, potentially trapping liquid water in upper layers. Subsequent thawing can redistribute this water nonlinearly, altering the post-drainage water balance. These processes affect *mrsol* but not *mrsos*, since the latter is diagnosed earlier in the time-stepping sequence, prior to freeze–thaw and drainage adjustments. Together, these factors indicate that discrepancies in CESM2 and NorESM2-LM stem not only from data or metadata issues but also from inherent structural and process-level characteristics of the land model.

### 4.2. Related studies and model comparisons

SM in CMIP models has been the focus of numerous studies. In a recent analysis, Qiao *et al* (2022) evaluated the performance of a range of CMIP6 models in simulating SM, with a particular focus on

two depths: surface SM (0–10 cm) and deeper SM (0–2 m). The vertical integration method employed to estimate deeper SM in their study was also applied here (i.e. equation (1)). However, this study did not address the discrepancies in how different models represent SM. For instance, Qiao *et al* (2022) found that CESM2 showed relatively low skill in simulating surface SM but performed better at simulating deeper SM in northern latitudes, such as Siberia and Eastern North America. However, this conclusion rests on an inconsistent definition of deep SM across models. Similarly, Wang *et al* (2022) analyzed SM in CMIP6 models over China, using methods comparable to those in Qiao *et al* (2022) and this study, but defining their depth ranges as 0–10 cm (surface) and 0–100 cm (root zone). Interestingly, the authors did not report any issues related to incorrect depth_bnds in CESM2, and the results for *mrsos* in ACCESS-ESM1-5 appeared normal despite the known misreporting of this variable. Therefore, how can models be evaluated if the very variable being assessed is not consistently defined?

Our findings in section 3.3 demonstrate that inconsistencies between *mrsos* and the vertically integrated *mrsol* can propagate into significant differences in the evaluation of simulated land–atmosphere coupling, affecting relationships with variables such as GPP, LAI, and ET. These results align with recent studies that emphasize the sensitivity of ecosystem–climate feedbacks to SM representation. For instance, Zhang *et al* (2024) showed that biases in SM dynamics can alter vegetation–climate feedbacks and destabilize terrestrial ecosystem functioning under warming scenarios. Likewise, Zhang *et al* (2025) found that errors in distinguishing surface versus subsurface SM substantially affect ET and land–atmosphere coupling strength. Further, a recent study by Qiao *et al* (2024) highlighted how CMIP6 models exhibit regional inconsistencies in the sensitivity of air temperature to surface soil temperature, suggesting that biases in soil water availability and precipitation– ET balance may lead to overestimation of land–atmosphere energy exchanges in some regions. Together with our results, these studies highlight that ensuring physical consistency between *mrsos* and *mrsol* is not merely a technical concern but is critical for accurately representing feedback processes and for the credibility of model-based assessments of how SM impacts other processes, like floods and its impacts or analyzes on land–atmosphere interactions.

### 4.3. CMOR, CF-compliance, and data quality control

CMOR (climate model output rewriter) is essential for ensuring ESM model outputs comply with CMIP standards. It formats raw data into a consistent structure for multi-model comparisons, with strict requirements for variables and metadata attributes, such as the *sdepth1* attribute, which specifies the top 10 cm of SM for the *mrsos* variable in CMIP6. In addition to CMOR, CF-compliance (climate and forecast standard names convention) ensures standardized variable names, coordinates, and attributes, promoting data consistency across models and facilitating easier comparison. It helps maintain interoperability between climate data and analysis tools. However, issues arose in models like CESM2, which failed to include sdepth1 in its metadata, and ACCESS-ESM1-5, which did not follow the correct definition of sdepth1. These discrepancies violate both CMIP6 and CF standards, complicating data consistency. These issues highlight the need for stronger CMOR quality control and better CF-compliance, ensuring that models follow standardized definitions for improved multi-model comparisons, which is especially important for forthcoming model intercomparison projects, like CMIP7.

## 5. Implications and recommendations for the modeling community

These findings raise several important considerations for the broader modeling community. First, it is essential to assess whether the discrepancies we document are large enough to affect surface SM research, particularly in studies relying on multi-model comparisons or benchmarking against observational datasets (Massoud *et al* 2022, Qiao *et al* 2022, Wang *et al* 2022, 2023). Second, when both *mrsos* and *mrsol* are available, researchers must decide which to prioritize. Since *mrsos* is often a diagnostic variable derived from *mrsol*, the latter may offer greater physical fidelity by more directly reflecting model processes. Finally, to improve consistency and reliability in future model intercomparison projects (c.f., Massoud *et al* 2025), we recommend more rigorous quality assurance protocols that can detect and resolve inconsistencies like those documented here. Efforts to address these challenges are already underway through initiatives such as the rapid evaluation framework, launched by the CMIP Climate Model Benchmarking Task Team, which aims to establish a standardized, open, and systematic approach for evaluating model output in CMIP7 (Hoffman *et al* 2025).

These issues extend beyond SM. Many variables in atmospheric, oceanic, and terrestrial models are derived from layered outputs through vertical integration, averaging, or interpolation, for example, near-surface air temperature, ocean heat content, root-zone soil carbon (*cSoilAbove1m*), and root-zone water content (*rwzc*). Others must satisfy physical constraints, such as energy conservation in surface radiation or mass balance in the carbon cycle. For instance, net ecosystem exchange (*NEE*) over land and ocean should align with the net atmosphere-land

CO₂ flux (*netAtmosLandCO2Flux*) and the air-sea flux (*fgco2*), respectively. When these derived variables are reported without validation against their source fields, inconsistencies can propagate through scientific analyzes and assessments.

This study highlights the importance of routinely checking the consistency between original and derived variables. As ESMs become more complex, and as the community prepares for CMIP7, transparent documentation, accessible diagnostics, and systematic quality control will be essential to ensure that model outputs remain scientifically sound and suitable for downstream applications.

## Data availability statement

The CMIP6 model data used in this study are publicly available and can be accessed through the Earth System Grid Federation (ESGF) nodes. The data can be retrieved from the ESGF portal (https://esgf-node.llnl.gov/projects/cmip6/) (Earth System Grid Federation (ESGF) 2025) by searching for the relevant model output datasets. These datasets are distributed under the WCRP Coupled Model Intercomparison Project (CMIP6) protocols and are available to the public for research purposes. Specific details on accessing the data, including model names, experiment IDs, and variable names, can be found in the corresponding metadata files on the ESGF website. All of the codes needed to run the analysis and everything required to reproduce this work are available on GitHub: https://github.com/EliasMassoud1/Paper_mrsos_mrsol (Zenodo DOI: http://doi.org/10.5281/zenodo.14967814) (Massoud 2025).

## Acknowledgments

## Author contributions

Elias C Massoud ⬛ 0000-0002-1772-5361
Conceptualization (equal), Data curation (equal), Formal analysis (lead), Investigation (lead), Software (equal), Validation (equal), Visualization (equal), Writing – original draft (lead), Writing – review & editing (lead)

Nathan O Collier
Conceptualization (equal), Formal analysis (supporting), Investigation (supporting), Methodology (equal), Software (equal), Validation (equal), Writing – review & editing (supporting)

Min Xu
Conceptualization (supporting), Formal analysis (supporting), Investigation (supporting), Validation (supporting), Writing – review & editing (supporting)

Mingjie Shi ⬛ 0000-0002-2469-4831
Conceptualization (supporting), Investigation (supporting), Writing – review & editing (supporting)

Forrest M Hoffman
Conceptualization (supporting), Funding acquisition (lead), Writing – review & editing (supporting)

## References

Alemohammad S H, Fang B, Konings A G, Aires F, Green J K, Kolassa J, Miralles D, Prigent C and Gentine P 2017 Water, energy, and carbon with artificial neural networks (WECANN): a statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence *Biogeosciences* **14** 4101–24

Avissar R 1998 Which type of soil–vegetation–atmosphere transfer scheme is needed for general circulation models: a proposal for a higher–order scheme *J. Hydrol.* **212** 136–54

Beck H E *et al* 2021 Evaluation of 18 satellite- and model-based soil moisture products using *in situ* measurements from 826 sensors, Hydrol *Earth Syst. Sci.* **25** 17–40

Champagne C, White J, Berg A, Belair S and Carrera M 2019 Impact of soil moisture data characteristics on the sensitivity to crop yields under drought and excess moisture conditions *Remote Sens.* **11** 372

Cheng S, Huang J, Ji F and Lin L 2017 Uncertainties of soil moisture in historical simulations and future projections *J. Geophys. Res. Atmos.* **122** 2239–53

Clark M P *et al* 2015 Improving the representation of hydrologic processes in Earth system models *Water Resour. Res.* **51** 5929–56

Claverie M, Matthews J, Vermote E and Justice C 2016 A 30+ year AVHRR LAI and FAPAR climate data record: algorithm description and validation *Remote Sens.* **8** 263

Daly E and Porporato A 2005 A review of soil moisture dynamics: from rainfall infiltration to ecosystem response *Environ. Eng. Sci.* **22** 9–24

Dorigo W, Himmelbauer I, Aberer D, Schremmer L, Petrakovic I, Zappa L and Preimesberger W 2021 The international soil moisture network: serving Earth system science for over a decade *Hydrol. Earth Syst. Sci.* **25** 5749–83

Earth System Grid Federation (ESGF) 2025 *CMIP6 model output data archive* (Earth System Grid Federation) (available at: https://esgf-node.llnl.gov/projects/cmip6/)

Feng D, Wang G, Wei X, Amankwah S O Y, Hu Y, Luo Z, Hagan D F T and Ullah W 2022 Merging and downscaling soil moisture data from CMIP6 projections using deep learning method *Front. Environ. Sci.* **10** 847475

Gleick P H 1989 Climate change, hydrology, and water resources *Rev. Geophys.* **27** 329–44

Goet G, Sonkar I, Kumar S, Hari Prasad K S and Ojha C S P 2024 Effect of salinity on crop growth and soil moisture dynamics: a study with root water uptake model *J. Hazard. Toxic Radioact. Waste* **28** 04024009

Gu X, Li J, Chen Y D, Kong D and Liu J 2019 Consistency and discrepancy of global surface soil moisture changes from multiple model-based data sets against satellite observations *J. Geophys. Res. Atmos.* **124** 1474–95

Hagemann S, Blome T, Ekici A and Beer C 2016 Soil-frost-enabled soil-moisture–precipitation feedback over northern high latitudes *Earth Syst. Dyn.* **7** 611–25

Hoffman F *et al* 2025 Rapid Evaluation Framework for the CMIP7 Assessment Fast Track, EGUsphere [preprint] (https://doi.org/10.5194/egusphere-2025-2685)

Holsten A, Vetter T, Vohland K and Krysanova V 2009 Impact of climate change on soil moisture dynamics in Brandenburg with a focus on nature conservation areas *Ecol. Modelling* **220** 2076–87

Huang H-Y and Margulis S A 2013 Impact of soil moisture heterogeneity length scale and gradients on daytime coupled land-cloudy boundary layer interactions *Hydrol. Process.* **27** 1988–2003

Jin C, Lei H, Chen J, Xiao Z, Leghari S J, Yuan T and Pan H 2023 Effect of soil aeration and root morphology on yield under aerated irrigation *Agronomy* **13** 369

Kerr Y H *et al* 2016 Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation *Remote Sens. Environ.* **180** 40–63

Lee J and Hohenegger C 2024 Weaker land–atmosphere coupling in global storm-resolving simulation *Proc. Natl Acad. Sci. USA* **121** e2314265121

Long Y, Wang H, Jiang C and Ling S 2019 Seasonal inflow forecasts using gridded precipitation and soil moisture information: implications for reservoir operation *Water Resour. Manag.* **33** 3743–57

Martens B, Miralles D G, Lievens H, van der Schalie R, de Jeu R A M, Fernández-Prieto D, Beck H E, Dorigo W A and Verhoest N E C 2017 GLEAM v3: satellite-based land evaporation and root-zone soil moisture *Geosci. Model Dev.* **10** 1903–25

Massoud E C *et al* 2025 Benchmarking soil moisture and its relationship to ecohydrologic variables in Earth System Models *EGUsphere [preprint]* (https://doi.org/10.5194/egusphere-2025-3517)

Massoud E C 2025 *Code and workflows for "Paper_mrsos_mrsol"* Zenodo (https://doi.org/10.5281/zenodo.14967814)

Massoud E C, Andrews L, Reichle R, Molod A, Park J, Ruehr S and Girotto M 2023 Seasonal forecasting skill for the high mountain Asia region in the Goddard Earth observing system *Earth Syst. Dyn.* **14** 147–71

Massoud E C, Bloom A A, Longo M, Reager J T, Levine P A and Worden J R 2022 Information content of soil hydrology in a west Amazon watershed as informed by GRACE *Hydrol. Earth Syst. Sci.* **26** 1407–23

Massoud E C, Liu Z, Shaban A and El Hage M 2021 Groundwater depletion signals in the Beqaa plain, Lebanon: evidence from GRACE and Sentinel-1 data *Remote Sens.* **13** 915

Massoud E, Turmon M, Reager J, Hobbs J, Liu Z and David C H 2020 Cascading dynamics of the hydrologic cycle in California explored through observations and model simulations *Geosciences* **10** 71

Mirus B B 2015 Evaluating the importance of characterizing soil structure and horizons in parameterizing a hydrologic process model *Hydrol. Process.* **29** 4611–23

Pan J, Shangguan W, Li L, Yuan H, Zhang S, Lu X, Wei N and Dai Y 2019 Using data-driven methods to explore the predictability of surface soil moisture with FLUXNET site data *Hydrol. Process.* **33** 2978–96

Purdy A J, Fisher J B, Goulden M L, Colliander A, Halverson G, Tu K and Famiglietti J S 2018 SMAP soil moisture improves global evapotranspiration *Remote Sens. Environ.* **219** 1–14

Qiao L, Zuo Z and Xiao D 2022 Evaluation of soil moisture in CMIP6 simulations *J. Clim.* **35** 779–800

Qiao Y, Wang G, Hagan D F T, Lim Kam Sian K T C, Chen L, Aalto J, Li S, Zou X and Lu J 2024 Contrasting sensitivity of air temperature trends to surface soil temperature trends between climate models and reanalyses *npj Clim. Atmos. Sci.* **7** 43

Reichle R H *et al* 2019 Version 4 of the SMAP level-4 soil moisture algorithm and data product *J. Adv. Model. Earth Syst.* **11** 3106–30

Sedighi M, Hepburn B D P, Thomas H R and Vardon P J 2016 Energy balance at the soil atmospheric interface *Environ. Geotech.* **5** 146–57

Sehgal V, Gaur N and Mohanty B P 2021 Global flash drought monitoring using surface soil moisture *Water Resour. Res.* **57** e2021WR029901

Singh N K, Emanuel R E, McGlynn B L and Miniat C F 2021 Soil moisture responses to rainfall: implications for runoff generation *Water Resour. Res.* **57** e2020WR028827

Suman S, Srivastava P K, Petropoulos G P, Pandey D K and O'Neill P E 2020 Appraisal of SMAP operational soil moisture product from a global perspective *Remote Sens.* **12** 1977

Thor R N, Mertens M, Matthes S, Righi M, Hendricks J, Brinkop S, Graf P, Grewe V, Jöckel P and Smith S 2023 An inconsistency in aviation emissions between CMIP5 and CMIP6 and the implications for short-lived species and their radiative forcing *Geosci. Model Dev.* **16** 1459–66

Wanders N, Karssenberg D, Bierkens M, Parinussa R, de Jeu R, van Dam J and de Jong S 2012 Observation uncertainty of satellite soil moisture products determined with physically-based modeling *Remote Sens. Environ.* **127** 341–56

Wanders N, Karssenberg D, De Roo A, De Jong S M and Bierkens M F P 2014 The suitability of remotely sensed soil moisture for improving operational flood forecasting *Hydrol. Earth Syst. Sci.* **18** 2343–57

Wang A, Kong X, Chen Y and Ma X 2022 Evaluation of soil moisture in CMIP6 multi-model simulations over conterminous China *J. Geophys. Res. Atmos.* **127** e2022JD037072

Wang Y, Mao J, Jin M, Hoffman F M, Shi X, Wullschleger S D and Dai Y 2021 Development of observation-based global multilayer soil moisture products for 1970–2016 *Earth Syst. Sci. Data* **13** 4385–405

Western A W, Grayson R B and Blöschl G 2002 Scaling of soil moisture: a hydrologic perspective *Annu. Rev. Earth Planet. Sci.* **30** 149–80

Zhang L, Dawes W R, Hatton T J, Reece P H, Beale G T H and Packer I 1999 Estimation of soil moisture and groundwater recharge using the TOPOG_IRM model *Water Resour. Res.* **35** 149–61

Zhang Q *et al* 2024 Water limitation regulates positive feedback of increased ecosystem respiration *Nat. Ecol. Evol.* **8** 1870–6

Zhang Q, Chuixiang Y, Destouni G, Wohlfahrt G, Kuzyakov Y, Runze L and Kutter E 2025 How Michaelis–Menten kinetics can represent ecosystem-scale respiration: scale and applicability *Environ. Res. Lett.* **20** 041003

Zhou S, Williams A P, Berg A M, Cook B I, Zhang Y, Hagemann S, Lorenz R, Seneviratne S I and Gentine P 2019 Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity *Proc. Natl Acad. Sci. USA* **116** 18848–53