

RESEARCH ARTICLE OPEN ACCESS

Scaling and Uncertainty in Soil Moisture Modelling: A Probabilistic Deep Learning Perspective

Balazs Bischof  | Erwin Zehe  | Ralf Loritz 

Institute of Water and Environment, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Correspondence: Balazs Bischof (balazs.bischof@kit.edu)

Received: 28 July 2026 | **Revised:** 28 July 2026 | **Accepted:** 3 August 2026

Keywords: Gaussian mixture models | large-scale | long-short term memory networks | probabilistic deep learning | small-scale | soil moisture | uncertainty | variability

ABSTRACT

Soil moisture plays a central role in terrestrial water and energy exchanges, yet its representation across spatial scales remains challenging due to strong heterogeneity, measurement uncertainty, and limited transferability of soil parameters. While deep learning models have shown skill in reproducing soil moisture dynamics at large scales, they are commonly applied deterministically, providing limited insight into predictive uncertainty and variability. Here, we apply a probabilistic deep learning framework based on Gaussian Mixture Long Short-Term Memory networks (GM-LSTMs) to model soil moisture dynamics and uncertainty across the contiguous United States using in situ observations from the International Soil Moisture Network. The model is trained and evaluated in a cross-validation setting on ungauged locations and forced with multiple meteorological datasets, with DayMet emerging as the most effective driver. Rather than focusing primarily on predictive performance, we use regional learning to examine how soil moisture dynamics and variability emerge across climatic, physiographic, and soil-textural gradients. We analyse the structure of predictive uncertainty using mixture entropy and Jensen–Shannon divergence to distinguish dispersion from distributional complexity. The model reproduces temporal dynamics and rank structure of soil moisture and outperforms ERA5-Land and SMAP benchmarks, while revealing systematic biases in absolute volumetric water content. Predictive uncertainty exhibits coherent spatial organization controlled by physiography and soil texture, and distinct moisture-dependent regimes consistent with established hydrological theory. Variability peaks at intermediate soil moisture states, in agreement with catchment-scale observations, indicating that signatures of soil moisture organization persist across scales. The results demonstrate that probabilistic data-driven modelling can provide physically interpretable information on soil moisture variability and uncertainty, and offer new perspectives on scale-robust patterns of land-surface hydrological behaviour in the absence of explicit small-scale process representation.

1 | Introduction

Soil moisture is an important hydrological variable that plays a central role in the terrestrial water and energy cycles, governing processes such as evapotranspiration, infiltration, runoff generation, and plant water availability (Seneviratne et al. 2010; Brocca et al. 2017). Its spatiotemporal variability has a major influence on land-atmospheric interactions, affecting boundary-layer development, precipitation formation,

and temperature extremes (Koster et al. 2004; Seneviratne et al. 2006; Hauser et al. 2016). Through these feedbacks, soil moisture modulates climate predictability on sub-seasonal to seasonal timescales (Guo et al. 2011) and contributes to variations in carbon uptake and drought intensity (Humphrey et al. 2021). At the same time, it is a key variable for hydrological and agricultural applications, including flood and drought forecasting, irrigation scheduling, and water-resource management (Brocca et al. 2017; Entekhabi et al. 2010). Despite

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Hydrological Processes* published by John Wiley & Sons Ltd.

its important role, accurately measuring and modelling soil moisture remains a long-standing challenge due to its strong heterogeneity across spatial and temporal scales, substantial measurement uncertainty, and dependence on local soil, vegetation, and climatic conditions (Mälicke et al. 2020; Zehe et al. 2010; Famiglietti et al. 2008; Robinson et al. 2008).

A key challenge arises from the strong spatial and temporal variability of soil moisture. While satellite-based observations provide spatially continuous soil moisture retrievals at footprints of several hundred square meters, these products require ground-based “truth” data for calibration and validation (Bronstert et al. 2012). Ground-based networks rely on numerous techniques, each with distinct advantages and limitations. Gravimetric methods offer high accuracy but are destructive and not suitable for repeated large-scale monitoring. Neutron probes can capture volumetric soil moisture over large depths but are costly and pose safety challenges. Among available methods, Time-Domain Reflectometry (TDR) has become the predominant approach for distributed, in situ soil moisture monitoring (Calamita et al. 2012; Hillel 1982). TDR sensors measure the dielectric permittivity of the soil-water-air mixture along a small transmission line, which is then converted into a soil moisture estimate. However, these measurements represent only a proxy for soil water content, and even sensors installed within a few decimetres can yield notably different readings (Jackisch et al. 2017). In other terms, the spatial extent covered by a single TDR observation is smaller than the representative elementary volume (REV) of soil water storage (Famiglietti et al. 2008) that could lead to strong uncertainties.

Explaining such small-scale variability in a modelling exercise would require high-resolution information about local properties at the 1 m scale, which are, even in densely observed research catchments, not at hand. At small scales, soil moisture variability arises from micro-topographic features, soil texture, vegetation cover which control water retention. Field studies have shown that TDR observations can spread across a range of $0.23 \text{ m}^3 \text{ m}^{-3}$ (around 50% of porosity) and more, even when taken at a spatial extent of 20 by 20 square meters (Zehe et al. 2010), reflecting local fluctuations in soil texture and soil water retention. Such variability can be as large as that observed across an entire catchment (Mälicke et al. 2020). While point-scale measurements may not represent absolute storage, they are often rank-stable, meaning that wet and dry locations preserve their relative order over time (Mälicke et al. 2020). Various geostatistical studies have investigated soil moisture variability at small catchment scales, typically ranging from 1 to 5 km^2 . In an examination by Bárdossy and Lehmann (1998) in southwest Germany, soil moisture measurements at 60 locations within a 6.3 km^2 catchment were analysed. The study explored different geostatistical interpolation techniques, such as ordinary kriging, indicator kriging, external drift indicator kriging, and Bayes-Markov updating. The findings highlighted the advantages and drawbacks of each method, offering insights into their applicability for small-scale soil moisture variability assessment. Herbst and Diekkrüger (2003) took a different approach by developing a model process design that enables the analysis of both spatial and temporal soil moisture structures. Their work emphasized the challenge of using point measurements to validate the modelled spatial structure of soil moisture due to differing spatial

scales. Additionally, the study by Perry and Niemann (2008) contributed valuable insights into soil moisture variability on a larger scale, presenting interpolation strategies for extrapolating point measurements across entire catchments. These findings show how small-scale variability of soil moisture complicates its representation in models operating at larger spatial scales.

The mismatch between process, observation, and modelling scales has been recognized as a central challenge in hydrology (Blöschl and Sivapalan 1995). Traditional physics-based models, which rely on the numerical solutions of the Darcy-Richards equation, provide process-level understanding but often fail to generalise across scales. Parameters such as hydraulic conductivity and soil water retention curves are not easily transferable, and effective large-scale representations remain uncertain. As one moves from well-instrumented research catchments (e.g., Attert; Pfister et al. 2017) to continental domains such as the contiguous United States (CONUS), data limitations and heterogeneity highlight such scaling issues, making it difficult to maintain both process fidelity and predictive skill.

Data-driven modelling approaches have emerged as powerful tools to estimate soil moisture at large scales (Orth 2021; Roberts et al. 2022; Liu et al. 2022; Han et al. 2023). Among these, Long Short-Term Memory networks (LSTMs) have shown considerable success due to their capacity to learn temporal dependencies from sequential data, making them well-suited for the task. Recent papers highlighted the importance of not only predicting the absolute values of hydrological variables but also characterising the associated uncertainty and variability (Li et al. 2023; Fang et al. 2020). Despite this growing emphasis on uncertainty, LSTMs are still most commonly applied in a deterministic way, producing single-value predictions without expressing model confidence. This lack of predictive uncertainty limits their usefulness in risk-aware decision-making contexts. Probabilistic extensions, such as Gaussian Mixture LSTMs (GM-LSTM; Klotz et al. 2022), address this shortcoming by outputting a mixture of Gaussian distributions rather than point estimates. In doing so, they capture both the central tendency and the inherent uncertainty (and small-scale variability; see Bischof et al. 2026) of soil moisture predictions, thereby enabling a richer and more realistic representation of spatiotemporal dynamics. In practical terms, knowing how confident a model is can be as critical as the prediction itself.

In this study, we explore how probabilistic deep learning captures the multiscale variability of soil moisture and produces meaningful uncertainty estimates across diverse spatial and hydroclimatic conditions. To this end, we employ an already developed GM-LSTM framework to predict soil moisture dynamics and variability from the local scale (Attert; Bischof et al. 2026) to the continental scale (CONUS), examining its capacity to generalize beyond the scale limitations that constrain traditional process-based models.

The specific objectives of this work are to:

1. Quantify the ability of the GM-LSTM model to reproduce soil moisture variability and dynamics over the entire CONUS

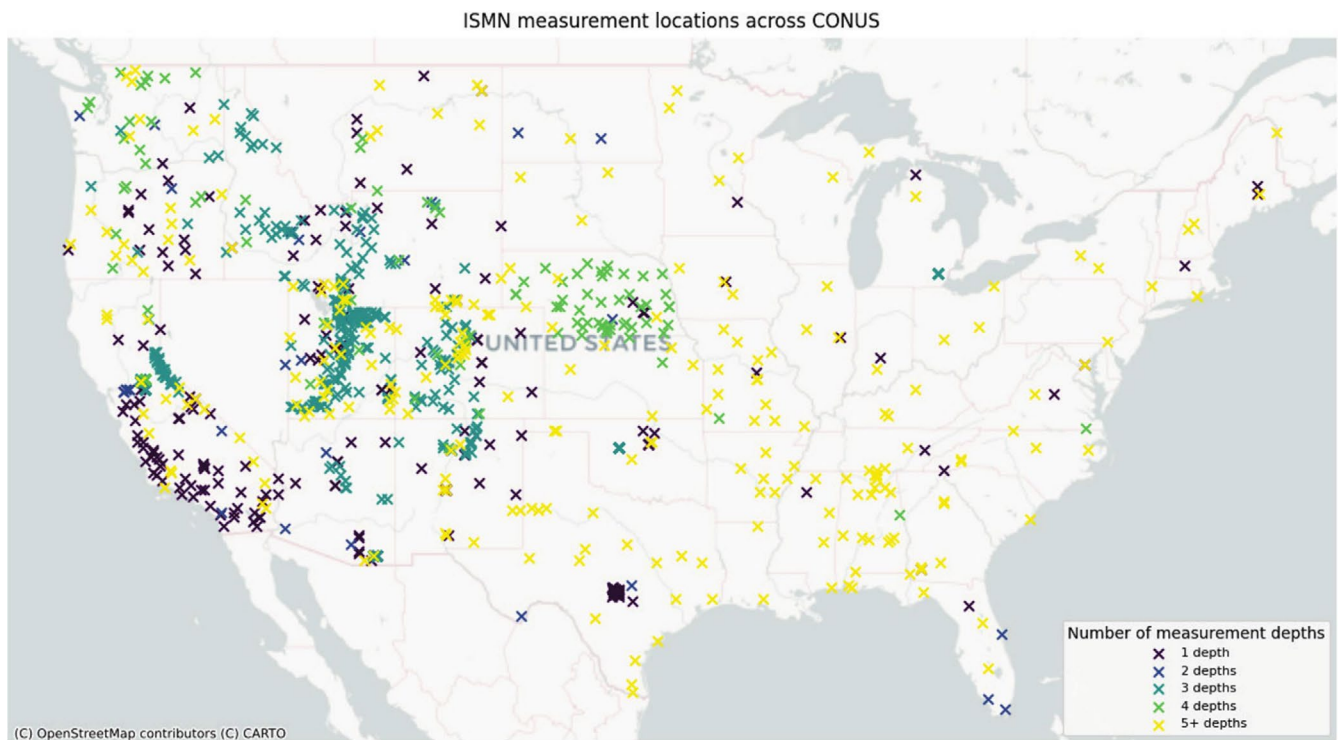


FIGURE 1 | Spatial distribution of ISMN measurement sites across CONUS with colour-coded depth availability. The distribution of time series across depth ranges is as follows: 0–5 cm (868), 5–10 cm (415), 10–20 cm (737), 20–40 cm (180), and > 40 cm (999).

2. Evaluate how predictive uncertainty of the GM-LSTM corresponds to known physical and observational sources of variability
3. Determine whether probabilistic data-driven models can reveal scale-independent patterns, offering new insights into the organization and predictability of soil moisture

By addressing these questions, this study aims to advance understanding of how probabilistic, data-driven approaches can complement and extend process-based modelling. The findings help bridge the gap between small-scale process knowledge and robust large-scale predictive capability, highlighting how regional probabilistic modelling can reveal emergent soil moisture dynamics across climatic and physiographic gradients. This supports the development of hydrological models that are less constrained by scaling issues and better able to represent the intrinsic variability of soil moisture.

2 | Domain and Data

2.1 | Study Area

The International Soil Moisture Network (ISMN; Dorigo et al. 2011) is a global database that compiles and quality-controls in situ soil moisture observations from a wide range of monitoring networks. This study focuses on sites within the CONUS, where the ISMN offers broad coverage across diverse environmental conditions, leveraging the region's extensive network availability and wide range of climates, land covers, soils, and geological settings. This environmental diversity provides an ideal testbed for training models under different

hydrometeorological conditions, enhancing the generalizability of the results. In total, 1017 locations were selected across CONUS, yielding approximately 3200 in situ soil moisture measurement time series (Figure 1). These time series span multiple monitoring depths (0.05–1 m), with some locations providing single-depth measurements, while others offering multi-depth profiles of up to five distinct soil layers. The selected temporal coverage of the data extends from January 1, 2000 to December 31, 2020. The length of individual time series varies by location, with some sites providing nearly continuous measurements across the full 21-year period, while others offer only short observational windows of 3–6 months.

A quality control (QC) procedure was implemented to identify and correct implausible behaviour in soil moisture time series. For each record a set of physically and statistically grounded criteria was applied. A series was considered valid if it exhibited: (1) sufficient temporal coverage (> 365 days with valid data), (2) a low proportion of exact zero values (< 20%), (3) realistic variability (standard deviation > $0.02 \text{ m}^3 \text{ m}^{-3}$), (4) no excessively long periods of consistency (maximum run of changes that are less than 10^{-3} lasting < 90 consecutive days), and (5) no physically implausible daily fluctuations (> $0.4 \text{ m}^3 \text{ m}^{-3}$). Observations meeting all criteria retained unchanged, whereas those failing any test underwent targeted correction that removed unreliable segments, specifically timesteps with values of 0 and samples belonging to days identified as containing large daily jumps or extended flat runs. The reduced series were then re-evaluated under the same QC framework; those that subsequently met all criteria were accepted and kept, while series that continued to fail (e.g., due to insufficient coverage) were excluded. After the QC procedure,

2765 time series were kept, equaling a 14% reduction in available data.

2.2 | Dynamic and Static Input Features

We trained multiple model configurations to assess the influence of different meteorological forcing datasets on prediction performance. Dynamic input variables (Tables A1–A3) were derived from: ERA5-Land (Muñoz-Sabater et al. 2021; 9km, hourly), DayMet (Thornton et al. 2021; 1 km, daily), and NLDAS (Xia et al. 2012; 12km, hourly), as well as a combined setup using all forcings. These datasets differ in spatial resolution, temporal frequency, and data sources (reanalysis, ground-based, or blended), allowing us to test how such characteristics affect model skill and uncertainty. Static physiographic attributes (Table A4) were taken from the HydroATLAS database (Linke et al. 2019), providing information on topography, soils, land cover, and climate zones.

To efficiently gather input data, we used a modified version (<https://github.com/BalazsBis/earthengine-exporter>) of the open-source Caravan dataset pipeline (Kratzert et al. 2021), designed to collect data for rainfall-runoff modelling. While the original framework aggregates gridded data over watershed boundaries, such spatial aggregation is not necessary for point-based soil moisture modelling. Therefore, the pipeline was optimized to extract only the required gridded data from Google Earth Engine (GEE) for exact point coordinates corresponding to the ISMN measurement locations.

3 | Model and Methods

3.1 | LSTMs Meet GMMs: What They Are and Why We Use Them?

Traditional deep learning (DL) models in hydrology produce deterministic point estimates, capturing temporal dependencies, but neglecting predictive uncertainty. Such uncertainty is often approximated through model ensembles. In contrast, the GM-LSTM used here outputs the parameters of a time-varying probabilistic distribution rather than a single value. The model generates multiple means (μ), standard deviations (σ), and mixture weights (π) for a number of Gaussian components, representing the relative contribution of each mode. This formulation enables the model to capture multimodal and non-Gaussian characteristics of soil moisture dynamics arising from interacting processes such as precipitation, evapotranspiration, infiltration, and spatial heterogeneity— processes that often keep soil moisture in multiple, locally quasi-equilibrium states throughout the year.

We apply a model architecture from our previous work (Bischof et al. 2026) that has been tested in a 250 km² highly monitored experimental catchment. Here, we extend the application to the continental scale, adapting the framework for large-scale soil moisture modelling. To improve generalization, Gaussian noise with a standard deviation of 0.05 was added to the input during training, and mixture weights were temperature-scaled via a soft-max function for numerical stability. Given

the large scale of the application, computational optimizations were required to make training and inference feasible across the extensive dataset. All hyperparameters are listed in Table B1.

3.2 | Benchmarking Model Performance

Predictions were systematically compared against volumetric water content (VWC) estimates from the ERA5-Land reanalysis and the Soil Moisture Active Passive (SMAP; Reichle et al. 2025) satellite products. To support robust evaluation, a five-fold cross validation was implemented: in each fold, 80% of the locations and the corresponding soil moisture time series were used for training, and a non-overlapping 20% were held out for testing, such that the model was evaluated on entirely unseen locations. This procedure was repeated five times for each fold. We have no validation period as we did not adjust our model hyperparameters and used the same setup as reported in Bischof et al. (2026).

3.3 | Deterministic and Probabilistic Evaluation Metrics

Model performance was evaluated using the Kling-Gupta efficiency (KGE), the Spearman Rank Correlation (Spearman's ρ), and the Root Mean Squared Error (RMSE) metrics. However, in multimodal predictive distributions, the mean may sit in an unlikely space between distinct modes. Such scenarios might be problematic for the deterministic metrics to represent since the model with multimodal distribution effectively says that “the outcome could be this or that, but not something in-between.” Therefore, especially for multimodal predictions, calculating the metrics with weighted averages fails to represent the actual meaning of predicted distributions and a metric like the Continuous Rank Probability Score (CRPS) is required to have a more fair and realistic evaluation compared to simply using a central value.

3.4 | Continuous Rank Probability Score

CRPS measures the discrepancy between the predicted cumulative distribution function (CDF) of a forecast and a step function representing the true outcome. It quantifies the spread of the predicted distribution around the observed value. The CRPS is defined as follows:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} [F(x) - 1(x \geq y)]^2 dx \quad (1)$$

$F(x)$ is the cumulative distribution function (CDF) of the prediction, evaluated at x , $1(x \geq y)$ is the indicator function that equals 1 if $1(x \geq y)$, 0 otherwise, and dx the infinitesimal change in x , indicating integration over the entire real number line. Values near zero indicate excellent probabilistic accuracy (e.g., 0.0–0.05 very good, 0.1–0.3 moderate), while larger values (e.g., >0.5) signal increasingly poor sharpness or calibration.

3.5 | Analysing Uncertainty

To characterize the structure of predictive uncertainty, we compute two complementary quantities from the gridded GM-LSTM predictions for a pre-selected date: (1) the overall dispersion of the predictive distribution, and (2) the internal complexity arising from disagreement or separation between mixture components. All predictive mixtures are evaluated on the fixed physically meaningful interval of $[0.02, 0.65] \text{ m}^3 \text{ m}^{-3}$, divided into bins of width 0.02 (equaling measurement uncertainty; Robinson et al. 2008). Bins that receive no probability mass are assigned a very small value only to ensure numerical stability and avoid undefined logarithmic terms. This allows all locations to be evaluated on the same fixed support, so differences in the metrics reflect genuine differences in predictive uncertainty rather than artefacts of empty bins (Darscheid et al. 2018). The lower bound represents residual soil water content, below which liquid water is largely immobile, while the upper bound corresponds to soil porosity, which defines the maximum volumetric water content under saturation.

3.5.1 | Mixture Entropy: Overall Dispersion

The mixture entropy measures how widely the predictive probability spreads across the predefined interval. It is computed as the Shannon entropy (Shannon 1948) of the binned mixture distribution,

$$H_{\text{mix}} = - \sum_i p_i \log_2 p_i \quad (2)$$

High values indicate low-confidence predictions, while low values indicate narrow, confident ones. With $M = 32$ bins (plausible range of 0.02 to 0.65 $\text{m}^3 \text{ m}^{-3}$ with bin size of 0.02 $\text{m}^3 \text{ m}^{-3}$), the possible range of entropy values are $0 \leq H_{\text{mix}} \leq \log_2 32 \approx 5$ bits.

3.5.2 | Shape Complexity: Jensen–Shannon Divergence From the Dominant Gaussian Component

H_{mix} alone does not allow one to distinguish whether the uncertainty arises from a genuinely broad distribution or from the presence of multiple distinct modes in the mixture. To quantify how different the predicted mixture distribution $p(x)$ is from its dominant Gaussian component $q(x)$, we compute the Jensen–Shannon (JS) divergence

$$\text{JS}(p, q) = \frac{1}{2} \text{KL}(p \parallel m) + \frac{1}{2} \text{KL}(q \parallel m), \quad m = \frac{1}{2}(p + q) \quad (3)$$

For each grid cell and time step, the model produces a predictive distribution in the form of a Gaussian mixture, obtained from the ensemble of model folds. The predictive distribution p , is discretized into probabilities p_i over a physically meaningful soil moisture range. To assess how much of the predictive uncertainty is captured by the most probable mixture component, we construct a reference unimodal distribution by selecting the Gaussian component with the largest mixture weight. This dominant Gaussian represents the single most likely soil moisture state predicted by the model. It is discretized over the same bins, yielding probabilities q_i . The similarity (or difference) between

the full predictive mixture and its dominant Gaussian component is calculated using the JS divergence. This is computed by using the midpoint distribution $m = \frac{1}{2}(p + q)$ and the average of the two Kullback–Leibler (KL) divergences, $\text{KL}(p \parallel m)$, describing how different the predicted distribution is compared to the midpoint and $\text{KL}(q \parallel m)$, showing how different the single Gaussian approximation is from the midpoint. Each KL divergence term measures how much information is lost when replacing one distribution with the midpoint (for more details see Loritz et al. 2019). Low JS values indicate that the predictive uncertainty is well represented by a single dominant Gaussian component, whereas high JS values reflect predictive distributions with substantial probability mass in secondary modes, indicating multiple competing plausible soil moisture states and pronounced multimodal or non-Gaussian behaviour.

3.5.3 | Distributional Structures

We analyse how predictive behaviour transitions between simple and complex structures, as well as between confident and uncertain regimes, across different soil moisture intervals. For each interval, grid-level predictive distributions are categorized based on their entropy (H_{mix}) and JS divergence, using the median values across all grid cells as thresholds to distinguish lower versus higher uncertainty and simpler versus more complex distributional structure. To classify the structure, the median of H_{mix} and JS was used across all grid cells.

$$H_{\text{med}} = \text{median}(H_{\text{mix}}) \text{ and } \text{JS}_{\text{med}} = \text{median}(\text{JS})$$

Each grid cell is assigned to one of the four regimes based on whether it lies above or below these thresholds:

- Simple—confident: $H_{\text{mix}} \leq H_{\text{med}}$ and $\text{JS} \leq \text{JS}_{\text{med}}$
- Simple—uncertain: $H_{\text{mix}} > H_{\text{med}}$ and $\text{JS} \leq \text{JS}_{\text{med}}$
- Complex—confident: $H_{\text{mix}} \leq H_{\text{med}}$ and $\text{JS} > \text{JS}_{\text{med}}$
- Complex—uncertain: $H_{\text{mix}} > H_{\text{med}}$ and $\text{JS} > \text{JS}_{\text{med}}$

4 | Results

4.1 | Performance, Accuracy, and Bias of the Model

We evaluated different model setups using ERA5-Land, NLDAS, DayMet, and a combination of them (all forcings) as dynamic input variables. The outcomes showed that using DayMet and all forcings outperforms the other two setups. Using NLDAS as input data performs the weakest, while employing ERA5-Land lies in between. This highlights DayMet's high-resolution, gauge-based measurements proving to be more effective at capturing soil moisture dynamics than other, lower resolution products. Although combining all the selected variables from all products produced comparative results, we chose DayMet as the preferred input because it consistently performed well across both metrics, while keeping the setup simpler and more efficient. In comparison with the benchmarks, Table 1 shows that the median of the model driven by DayMet outperforms the VWC estimates from the ERA5-Land reanalysis and of SMAP. Although

ERA5-Land achieves a similar level of performance in terms of rank correlation, its median KGE and RMSE values are significantly worse. The model achieves a median rank correlation of 0.74 (for shallow depths), indicating that it captures the timing and ordering of wetting-drying events accurately. In contrast, its KGE has median values around 0.42, reflecting systematic level bias. Table 1 also shows that the model's accuracy declines with increasing depths. While shallow soil moisture shows direct response to meteorological forcings, deeper layers respond due to percolation and storage, after increasingly long lags and with a stronger dampening. Subsoil soil moisture dynamics are also governed by processes and properties that are weakly or not at all represented in the used features (e.g., vertical heterogeneity in saturated hydraulic conductivity, rooting depth and uptake profiles, and lateral flow). In addition, measurement challenges are typically larger for deeper sensors, worsening signal to noise ratios and limiting the achievable correlation. Variance in deep soil moisture is lower, so identical absolute errors yield poorer

KGE via its variability (α) and bias (β) terms, while longer lags decrease rank correlations.

Figure 2 presents the 1 km soil moisture predictions across the CONUS. Figure 2a shows an example gridded prediction for a winter day, chosen to get an exemplary impression of the model's large-scale behaviour and to visually assess physical realism. The model reproduces the Mississippi moisture corridor and captures the orographic influence of the Pacific Northwest, where mountains act as a barrier to incoming precipitation. The map also shows a distinct dry zone in the New Mexico–Colorado four corners region, matching the area of drought that persisted during that time period (US Drought Monitor 2019); in addition, a wet band appears along coastal California and the Sierra Nevada foothills, consistent with atmospheric river storms that struck during early and mid-January 2019, resulting in heavy rainfall and flooding (in between 3 and 10 January 2019 the Coastal and Sierra Nevada

TABLE 1 | Summary table of Kling-Gupta efficiency (KGE), root mean square error (RMSE), Spearman's rank correlation (Spearman's ρ), and continuous rank probability score (CRPS) metrics for the Gaussian Mixture LSTM (GM-LSTM) model and the ERA5-Land and SMAP soil moisture product benchmarks.

Model/depth bin	KGE		RMSE		Spearman's rho		CRPS	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
GM-LSTM/0–0.2 m	0.287	0.422	0.092	0.080	0.683	0.737	0.062	0.049
GM-LSTM/0.2–0.5 m	0.235	0.403	0.096	0.083	0.683	0.725	0.067	0.052
GM-LSTM/0.5–1.0 m	0.004	0.223	0.115	0.099	0.529	0.573	0.085	0.069
SMAP	0.106	0.205	0.105	0.091	0.572	0.637	N/A	N/A
ERA5-Land	−0.090	0.229	0.125	0.111	0.613	0.659	N/A	N/A

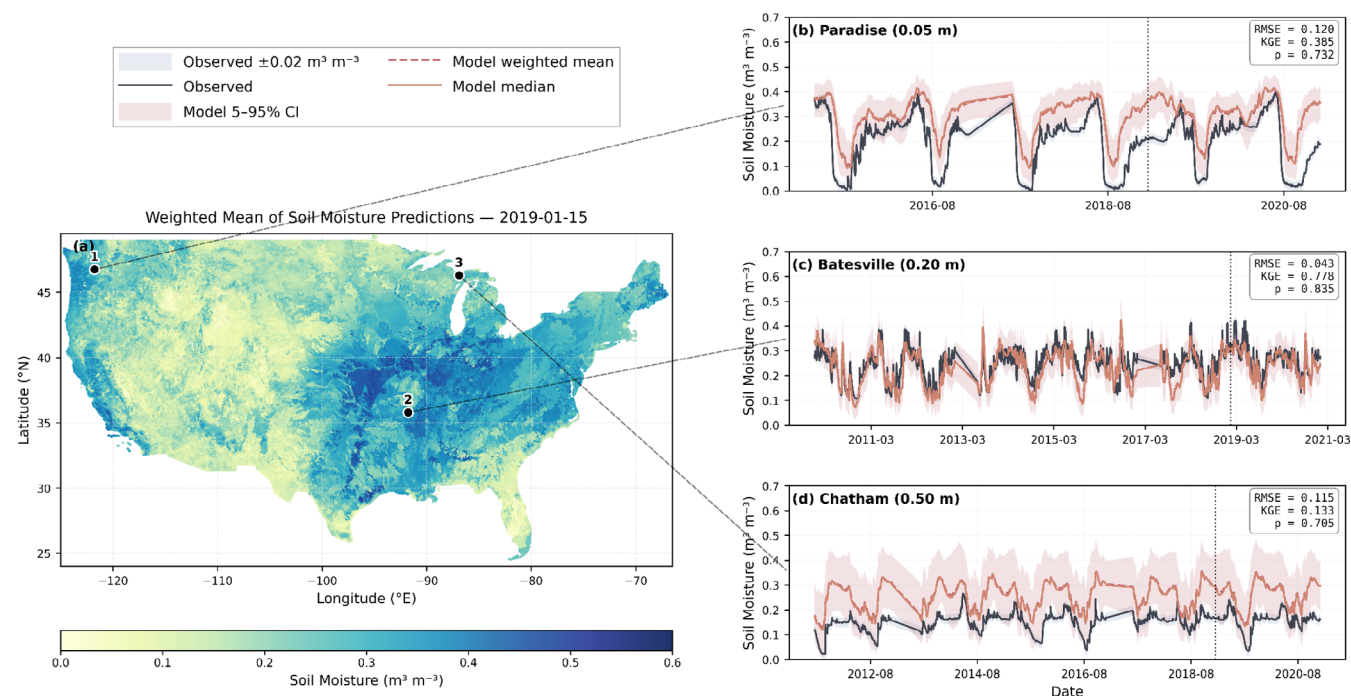


FIGURE 2 | Predictions of the Gaussian Mixture LSTM (GM-LSTM) model. (a) Weighted mean soil moisture predictions for the entire CONUS in 1 km resolution for a predefined date (15 January 2019). (b–d): Three locations for three different depths (0.05, 0.20, and 0.50 m) comparing actual measurements (black line with an uncertainty band of $0.02 \text{ m}^3 \text{ m}^{-3}$) with GM-LSTM predictions (red line with a 5% and 95% confidence interval).

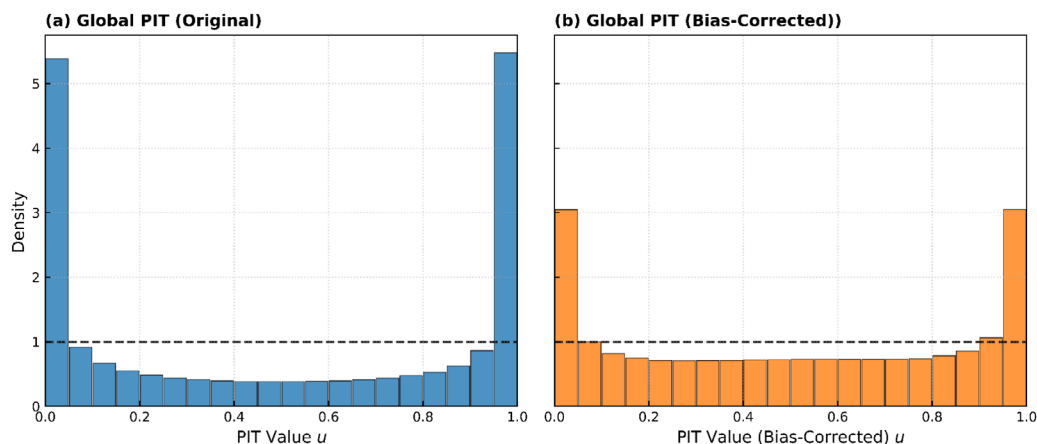


FIGURE 3 | Original (Panel A) and bias-corrected (Panel B) PIT diagrams of model predictions averaged across all folds.

mountains in California received more than 200 mm precipitation; CW3E 2019).

The time series in Figure 2b–d illustrate the behaviour supported by the PIT diagram (Figure 3; explanation can be found in Appendix C): the predictive intervals are systematically too narrow, and this under-dispersion is closely tied to bias in the mean predictions. When the predicted central tendency is shifted, the confidence intervals cannot compensate, causing observations to fall outside the predicted ranges and producing PIT patterns that deviate from uniformity. The example time series show the same mechanism: although temporal dynamics are mostly captured well, the prediction band frequently sits above or below the measurements. Importantly, this does not imply that simply inflating the predictive intervals would resolve the issue; widening the distribution without correcting the bias would only reduce sharpness and practical usefulness. In applications such as drought early-warning, too wide uncertainty bands would make the signal less informative. The summary uncertainty metrics are consistent with this interpretation (Table 1). A median CRPS of 0.049 (Table 1) indicates that the predictive distributions are sharp and capture substantial temporal structure. The CRPS-RMSE comparison (Figure C1) further supports this conclusion: across sites, CRPS is consistently lower than RMSE, showing that the probabilistic distributions contain meaningful information beyond the deterministic mean. The locally weighted scatterplot smoothing (LOWESS) trend shows that as RMSE increases, the CRPS-RMSE ratio approaches one, meaning that the probabilistic skill collapses toward deterministic error when the mean prediction becomes biased. This pattern is common in under-dispersed models whose uncertainty does not adapt when the mean is misplaced, meaning that the predictive spread is not inherently too narrow, but it is unable to expand in response to systematic mean error.

Together the PIT shape, time series behaviour, and RMSE-CRPS relationship form a coherent picture: the model is sharp and internally consistent, but biased and overconfident stemming mainly from the mean bias rather than insufficient variance in the mixture distribution. Addressing this bias would require modifying the predictive uncertainty structure rather than simply broadening predictive intervals, which is both technically challenging without introducing artificial corrections and

beyond the scope of this study. In the context of soil moisture, this behaviour is expected: absolute VWC is strongly location-dependent (limited by porosity and the residual soil water content and affected by the interplay of macropore flow, matrix flow and root water uptake– Grayson et al. 1997; Western et al. 2002; Teuling and Troch 2005), and multi-site, ungauged models frequently miss the mean while still capturing timing and relative changes reliably. Since drought detection, onset alerts, and irrigation scheduling depend on anomalies, thresholds, and short-term changes rather than absolute VWC, accurate dynamics and temporal ranking remain important determinants of practical usefulness.

4.2 | Physical Consistency in Model Behaviour

Figure 4 illustrates the relationships between predicted weighted soil moisture and various static attributes. The general trend shows that with increasing elevation (Figure 4b) and slope (Figure 4a), soil moisture declines. Higher elevations and steeper slopes are often associated with cooler temperatures, increased runoff potential, and thinner soil layers. These conditions reduce the soil's capacity to retain moisture and enhance drainage, leading to drier soil profiles. A maximum in predicted soil moisture near 300 m is reasonable, since mid-elevations often coincide with cooler microclimates, higher precipitation, and deeper or more developed soils compared to lowlands, which can produce higher moisture availability. At higher elevations, declining temperatures, thinner soils, and increased runoff typically reduce soil water retention, creating the observed hump-shaped relationship. The soil texture fraction curves display relatively clear and physically consistent patterns. For silt (Figure 4c), soil moisture increases with higher percentages, which can be attributed to the mixed pore structure of silt soils that allows for effective water retention while maintaining moderate drainage and infiltration. A similar trend is observed for clay (Figure 4e), where higher fractions correspond to higher soil moisture. This reflects the fine pore structure of clay soils, which promotes strong water retention and slower percolation. However, depending on the type of clay minerals and whether those may swell/shrink or not, infiltration behaviour might be largely different– ranging from almost no infiltration in case of absent or closed cracks

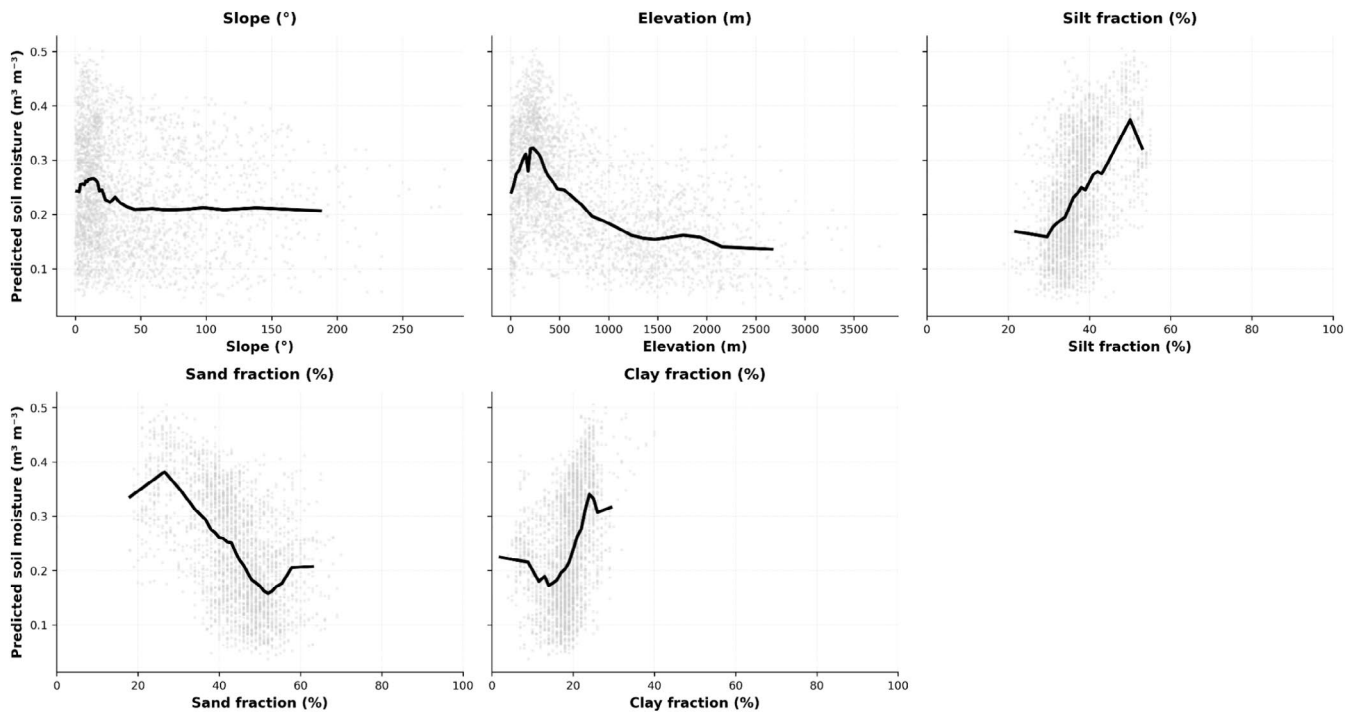


FIGURE 4 | Relationships between the predicted weighted mean soil moisture and different static attributes (slope degree (°), elevation (m), silt fraction (%), sand fraction (%), clay fraction (%)).

to preferential infiltration into cracked clay soils. In contrast, the sand fraction (Figure 4d) shows an inverse relationship: soil moisture decreases as sand content increases. This is expected, as sandy soils have larger pores that facilitate rapid drainage and limit their ability to retain water due to almost neglectable capillary forces.

4.3 | Linking Predictive Uncertainty to Environmental Controls

The entropy maps (Figure 5a,b) summarize the information content of the probabilistic predictions. Figure 5a displays the spread (H_{mix}) of the predicted distributions, and Figure 5b shows the spatial structure of the JS divergence. Both maps indicate that underlying attributes strongly influence the spread and shape of the predicted distributions. To get a cleaner interpretation, we analysed how environmental characteristics (i.e., static attributes) shape the structure of predictive uncertainty by examining the joint behaviour of H_{mix} and JS divergence across a range of physiographic and soil-texture settings (Figure 5c–e and Figure D1). We assessed their relationships with slope degree, elevation, sand, silt, clay fractions, and predicted soil moisture.

H_{mix} increases with predicted soil moisture content, being consistent with storage-release dynamics: wetter soils accommodate larger changes, producing broader predictive distributions. Topographic patterns show a contrasting pattern. Steep and high-elevation areas exhibit low H_{mix} , reflecting rapid drainage, more surface runoff (thus less infiltration/recharge), shallow or rocky profiles, and limited storage. Many of these regions also show elevated JS divergence, indicating structural uncertainty despite the relatively narrow spreads. Snow influence, frozen-thaw cycles, and heterogeneous mountain soils can produce

differing hydrological trajectories, leading to different plausible moisture states even when the variance is low. Sandy soils displayed low H_{mix} consistent with their limited storage capacity and fast drainage. Predictions here are more likely narrow but uncertain in structure. Mid-range entropies combined with higher JS divergence suggest threshold-like infiltration and preferential flow that create divergent moisture trajectories after rainfall. Silt contents of around 40%–50% cluster at higher H_{mix} . These soils have moderate retention and permeability, promoting relatively large moisture variability. A moderate rise in JS divergence in this range indicates that transitional textures can support differing hydrological responses. Clay-rich soils show higher H_{mix} but lower JS divergence on average. Their strong retention and slow drainage broaden the distribution but typically produce smoother, more consistent moisture dynamics, leading to decreased multimodality.

Across all attributes, a coherent pattern emerges: H_{mix} reflects water-storage capacity and drainage behaviour; therefore, it tends to be higher in clay and silt but lower in sandy environments. Dispersion is higher in lowlands, whereas it is lower in steeper and higher elevation terrains. Shape complexity (JS) identifies regions where hydrological dynamics support multiple plausible states, namely high-elevation and snow-influenced areas, coarse-textured sandy soils, and intermediate silt regimes.

4.4 | Understanding the Structure of Uncertainty

We examined broader patterns linking uncertainty measures with the predicted soil moisture values themselves and with seasonal variation. Higher H_{mix} values mainly occur under wet conditions (Figure 5a). This aligns with the expectation that wetter

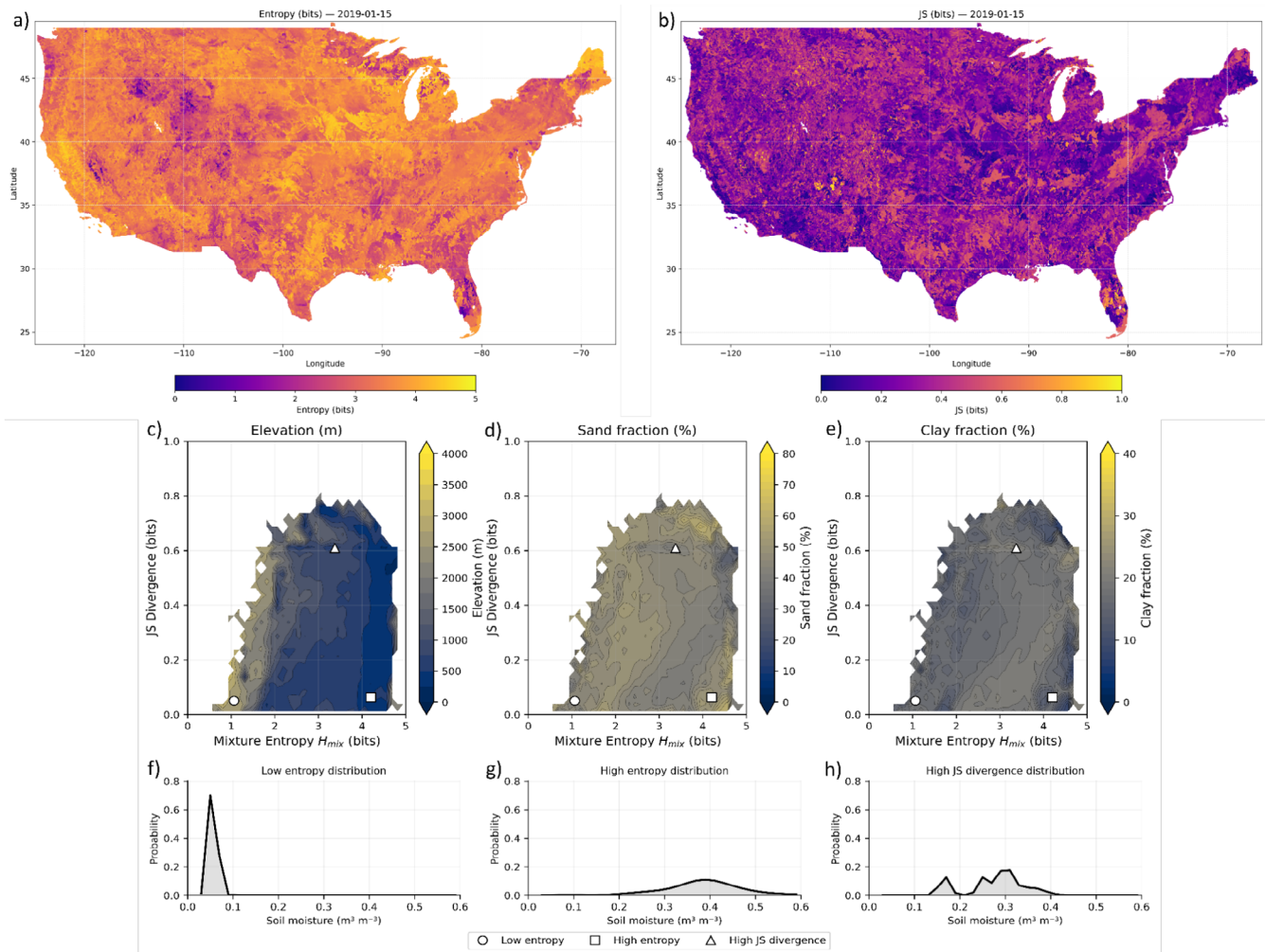


FIGURE 5 | CONUS maps of the two uncertainty measures (Panels A and B). Further plots (Panels C–E) show the relationships between mixture dispersion (H_{mix}), distributional complexity (JS divergence), and different environmental attributes. Panels (F–H) depict example distributions for three different parts of the plot (i.e., low entropy, high entropy, and high JS divergence). The figure summarizes predictions for a chosen date (15 January 2019) for all 1 km grid cells within the CONUS.

soils have greater storage capacity and exhibit broader variability, which widens the predictive distribution. JS divergence does not follow a monotonic trend with soil moisture, as the highest JS values cluster around intermediate entropy levels. This pattern reflects a statistical property of mixture distributions, where it is the easiest to express multimodality when the distribution has enough spread to separate the components but not too much so differences become smoothed out. Figure 5c confirms the positive correlation between moisture levels and dispersion, while it also shows that, besides being frequent at intermediate moisture levels, high JS divergence is more likely to occur at the lower ends compared to very wet states. This suggests that the model expresses more frequent multimodal behaviour not only when the system has moderate variability but also when soil moisture approaches lower bounds, where nonlinear processes and measurement noise can increase structural ambiguity.

Entropy peaks at intermediate soil moisture levels (Figure 5b), when the model must choose between differing soil moisture trajectories controlled by mid-summer convective rainstorms or winter dry anomalies. Similarly, in the hydrological process perspective, variability and thus predictive peaks around mid-range

soil moisture conditions, where small differences in antecedent wetness may cause transitions to switch in between matrix and preferential flow (Zehe et al. 2005; Zehe and Blöschl 2004; Zehe and Sivapalan 2008), creating coexisting regimes across sites. In both cases, typical seasonal states are easier to characterize and thus show lower uncertainty or variability, while in the mid-range, where multiple mechanisms might overlap, it produces the strongest spread.

Figure 6d shows how the seasonal cycle of multimodal predictions varies across soil moisture classes. For the low soil moisture class ($0.0\text{--}0.2\text{ m}^3\text{ m}^{-3}$), the frequency of multimodality remains relatively stable throughout the year, with multimodal predictions ranging around 30%–45%. This means uncertainty grows under dry conditions, probably because even small errors in precipitation rates may have a large impact on how well the model captures drying processes. The medium soil moisture class ($0.2\text{--}0.4\text{ m}^3\text{ m}^{-3}$) shows a rise in multimodal predictions during the summer months, when greater variability in precipitation and evapotranspiration also occurs. This seasonal increase likely reflects the complex interplay between rainfall events and drying, which introduces competing modes of soil moisture dynamics and thus greater predictive

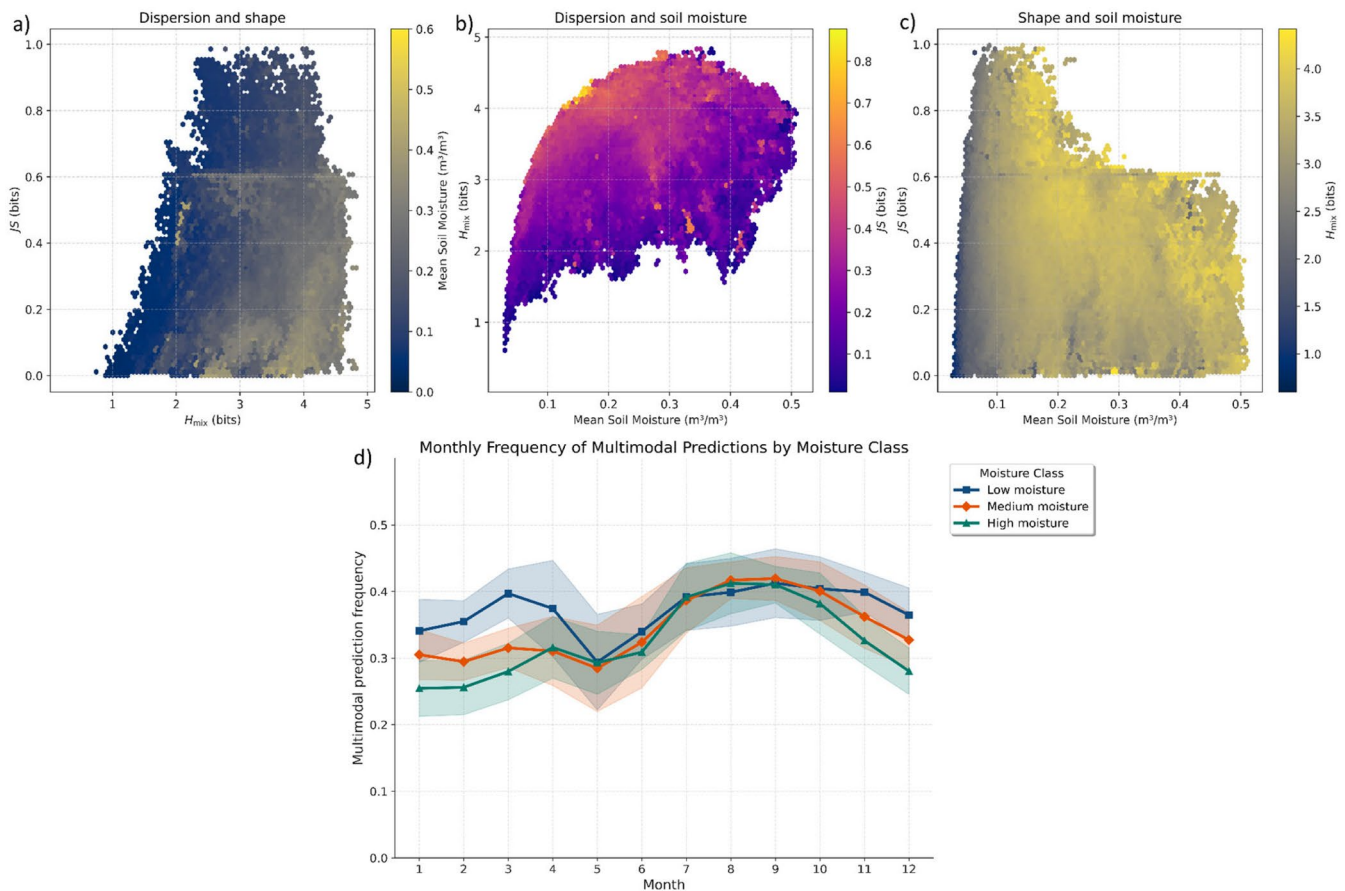


FIGURE 6 | Generic relationships between: (a) Entropy and JS divergence (colour coded by soil moisture); (b) Entropy and predicted soil moisture (colour coded by JS divergence); (c) JS divergence and predicted soil moisture (colour coded by entropy); and (d) the frequency of multimodality occurrence across the months of the year.

ambiguity. For the high soil moisture class ($0.4\text{--}0.6\text{ m}^3\text{ m}^{-3}$), multimodality is lowest in winter and spring, suggesting that wetter and cooler conditions produce more predictable dynamics. However, multimodality rises sharply during late summer, when high soil moisture often results from intense convective storms. Such events generate transient and localized conditions that are harder to capture, leading to diverse soil responses and higher uncertainty. Across all classes, the highest overall frequency of multimodal predictions occurs between July and September, a period characterized by convective rainfall and greater weather variability across much of the CONUS.

The distribution of prediction thresholds (Figure 7) across the four uncertainty classes (simple-confident, simple-uncertain, complex-confident, complex-uncertain) shows soil moisture dependent patterns in the model's uncertainty structure. In the driest conditions ($0.0\text{--}0.1\text{ m}^3\text{ m}^{-3}$), predictions are dominated by simple-confident and complex-confident behaviour, indicating that the model is certain in general (narrow distributions) but can express non-Gaussian structure when variations in forcing lead to multiple possible trajectories. As moisture increases into the $0.1\text{--}0.3\text{ m}^3\text{ m}^{-3}$ range, uncertainty shifts toward the complex-uncertain class, reflecting the intermediate soil moisture states (also shown in Figure 5b) where competing wetting and drying processes generate both broad and structurally ambiguous predictive distributions. Once soil reaches moderately wet conditions ($0.3\text{--}0.4\text{ m}^3\text{ m}^{-3}$), the balance slightly shifts again: the simple-uncertain class becomes

more important, and this dominance grows in the wettest range ($0.4\text{--}0.6\text{ m}^3\text{ m}^{-3}$), where it clearly becomes the leading class. In this domain higher water storage and slower drainage lead to wider, more uniform distributions, which reduces multimodality and structural complexity.

5 | Discussion

5.1 | Accuracy and Bias Problem

The results demonstrate that DL models can reproduce a significant portion of soil moisture dynamics even at ungauged locations across the CONUS. Combining all meteorological products did not yield significant improvements beyond what DayMet alone achieved, suggesting redundancy among the forcings—where additional datasets provided little new information. While Kratzert et al. (2021) demonstrated that multi-forcing setups in rainfall-runoff modelling can exploit complementary information to improve predictive performance, their use of basin-averaged inputs limits direct comparability with the present study. Despite no site-specific calibration, the model reaches performance comparable to other regional to continental soil moisture prediction studies (Xi et al. 2025; Geng et al. 2024; Vergopolan et al. 2021). While the general accuracy is satisfactory, clear bias patterns emerge. The dynamic and uncertainty patterns observed in the results point to a structural limitation that goes beyond statistical calibration: in

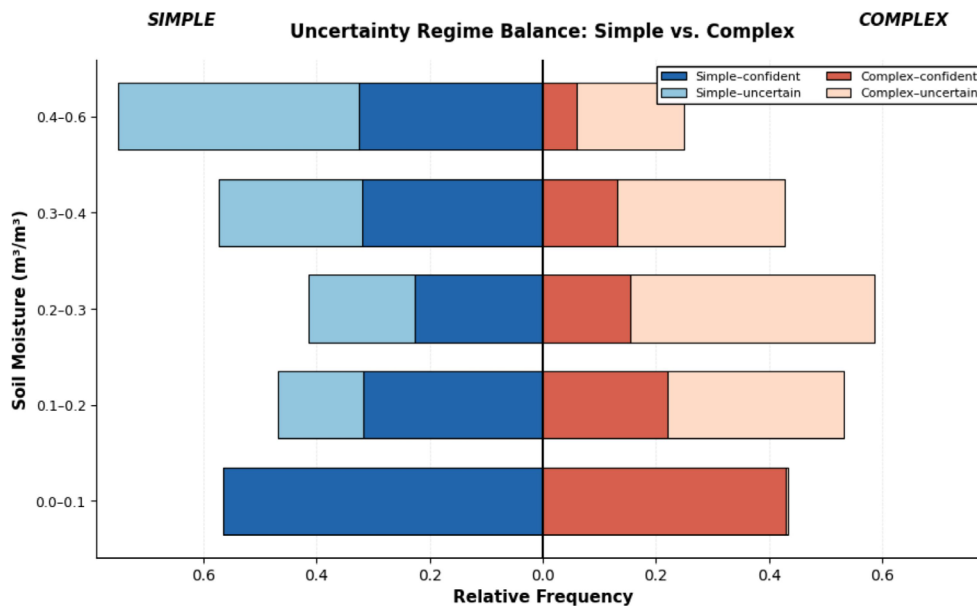


FIGURE 7 | Dominating uncertainty regimes based on soil moisture classes. The diagram summarizes how the balance between simple vs. complex and confident vs. uncertain predictive behaviour varies across soil moisture conditions. Each bar represents a soil moisture interval, with the left side showing “simple” prediction structures (lower JS divergence) and the right side showing “complex” structures (high JS divergence).

ungauged settings, the model is mostly capable of capturing soil moisture changes but not always where the mean should sit. The temporal dynamics appear transferable. What does not transfer as well are the local controls that govern absolute VWC. While soil texture is included in the model (and shows physically realistic behaviour; see Section 4.2), other key factors such as porosity, water content at field capacity and the permanent wilting point or rooting depth are not represented in the input data at all. These properties may vary sharply spatially and directly influence the mean levels, making it difficult for the model to understand site-specific baselines.

One technical reason for such biases might arise from the presence of TDR artefacts. Potential “air spikes” caused by poor probe-soil contact might result in systematically too low VWC values. This is because TDR probes estimate soil moisture by measuring the dielectric permittivity of the material surrounding the rods. Water has a very high dielectric constant (≈ 80), soil has moderate ($\approx 3-7$, depending on texture), and air has very low (≈ 1). When the probe loses good contact with the soil (e.g., due to shrinkage cracks, loose installation, freeze-thaw movements) air fills the space around part of the rods. Because the TDR signal travels through whatever medium is in contact with the rods, even a thin air gap can significantly reduce the measured dielectric permittivity, which the calibration converts into artificially dry VWC values.

The interplay between structural heterogeneity and measurement errors clarifies why the model appears both sharp and miscalibrated (as demonstrated with the used strictly proper-scoring rules): the predictive variance is appropriate for capturing dynamics, yet it lacks the needed information to adjust when the mean shifts. Inflating predictive intervals would not resolve this issue. Wider intervals would improve coverage, but reduce sharpness and weaken the usefulness of the uncertainty signal, especially for drought early-warning where

decisions depend on anomaly detection rather than broad uncertainty bands. In essence, this would treat a systematic mismatch as if it were random noise, rather than addressing the underlying cause.

The patterns show that ungauged soil moisture prediction is, as expected, more reliable when the task is to discover changes and ranks rather than actual magnitudes and states. The model is not failing, it is rather reflecting the limits of what can be learned from forcing data alone. The challenge is therefore less about fixing the uncertainty framework (e.g., through calibration) and more about improving the model’s understanding of “place”. If the model would understand the structural components better, the probabilistic predictions should naturally become both sharper and better calibrated.

5.2 | How Well the Models’ Uncertainty Outputs Connect to Reality?

The joint behaviour of dispersion (H_{mix}) and shape complexity (JS divergence) indicates that predictive uncertainty in the probabilistic soil moisture predictions is structured by physiographic and textural controls rather than representing homogeneous model noise. Classical field studies support this interpretation. Grayson et al. (1997) demonstrated that soil moisture patterns exhibit persistent spatial organization shaped by landscape position, with wetter valley bottoms and drier slopes forming stable spatial templates. Western et al. (2002) showed that this organization arises from the combined effects of topography, soil properties, and drainage pathways, leading to systematic differences in both the mean and variability of soil moisture across terrain units. Vereecken et al. (2014) further highlighted that soil moisture variability is governed by soil hydraulic characteristics, texture, and local storage conditions across a range of spatial scales. Together,

these studies indicate that spatial differences in variability and multimodality should be expected outcomes of physical landscape controls rather than artefacts of modelling. The observed patterns in H_{mix} and JS divergence align with this understanding: higher dispersion reflects differences in storage and drainage capacity, while increased shape complexity marks environments that support multiple hydrological response modes.

Regions with low H_{mix} but elevated JS divergence (such as high-elevation areas) align with known process complexity in such regions, where snow accumulation and variable thaw timing can lead to divergent wetness trajectories under similar meteorological forcing. By contrast, steep and high-elevation terrain exhibits systematically low H_{mix} , reflecting limited storage and rapid drainage pathways that constrain the range of possible soil moisture states. This pattern is in line with studies showing that topographic relief accelerates runoff generation and shortens subsurface residence times, thereby narrowing the soil moisture variability threshold.

Across textures, the uncertainty structure reflects how storage capacity and hydraulic behaviour create the balance between locally controlled variability and more organized responses. Sandy soils show low H_{mix} , consistent with their limited storage, lower moisture content, and rapid drainage: variability mostly remains narrow, governed by local vertical fluxes, analogous to the locally controlled, weakly organized patterns observed under dry conditions by Grayson and Blöschl (2001). Each soil “patch” behaves like its own bucket, getting wet when it rains, drying when it evaporates, and draining downward, mostly independent of what is happening upslope. The moderate JS divergence indicates that threshold-like infiltration (i.e., abrupt shifts in infiltration behaviour) and preferential pathways can still produce divergent trajectories after rainfall. Silt-rich textures ($\approx 40\%$ – 50% silt) show higher H_{mix} and a rise in JS divergence, reflecting intermediate retention and permeability that permit broader variability and multiple plausible hydrological responses. Clayey soils display high H_{mix} but lower JS divergence, producing a more aligned behaviour, decreasing multimodality.

5.3 | Generic Patterns in Soil Moisture Variability

H_{mix} has a clear peak in variability at intermediate soil moisture levels. This pattern has also been observed and modelled in the study of Teuling and Troch (2005) in three basins with different characteristics (Louvain-la-Neuve, Belgium; VCR-LTER, USA; Tarrawarra basin, Australia), in the research of Famiglietti et al. (2008) where they analysed four soil moisture field campaigns to characterize the behaviour of soil moisture variability across scales, in the paper of Lawrence and Hornberger (2007) where they examined the patterns of soil moisture variability across climate zones, and in our study for a small, experimental basin (Attert-catchment, Bischof et al. 2026), where several co-located sensors exhibit their greatest divergence in this transitional moisture range. In that setting, the variability arises from local heterogeneity, such as small differences in infiltration pathways, root-zone structure, or micro-topography (Zehe and Blöschl 2004; Zehe et al. 2005), leading to different soil

moisture responses under the same meteorological forcings. In contrast, the CONUS-scale model has access to only a single measurement at each location, with no representation of sub-grid variability, yet it produces uncertainty patterns that follow a similar moisture-dependent structure. This suggests that the intermediate moisture regime is associated with hydrological characteristics that are persistent across environments that remain detectable even after aggregation to much coarser scales. From a scaling perspective, this similarity between local sensor-derived variability and large-scale learned uncertainty indicates that some aspects of soil moisture dynamics may be more robust across scales than the detailed small-scale mechanisms themselves. Although the model cannot resolve the physical heterogeneity that drives variability at the point scale, it captures a moisture-dependent uncertainty structure that is consistent with expectations and aligns with behaviours observed in catchment studies. In this sense, the result aligns with the perspective outlined by Blöschl and Sivapalan (1995), who noted that some hydrological behaviours attenuate strongly with scale while others retain a recognizable imprint even after upscaling. The similarity between local sensor divergence and large-scale modelled uncertainty does not imply that the underlying fine-scale mechanisms are resolved; rather, it suggests that the intermediate-moisture regime is one of those patterns whose aggregated signal persists. This partial consistency across scales highlights how probabilistic, data-driven models may capture certain robust features of soil-moisture variability despite the absence of explicit information about small-scale processes.

The distribution of regimes in soil moisture intervals revealed a tendency that drier conditions contain a larger share of complex states, especially complex-uncertain. In these cases, the model represents multiple separated mixture components with internal disagreement, indicating that multiple different pathways might be plausible. This pattern is similar to the characteristics of the local-control regime by Grayson et al. (1997), in which low moisture, weak hydrological connectivity, and strong local-scale variability lead to divergent drying and wetting behaviour. As soil moisture increases, the balance shifts toward a dominance of simple-uncertain states. Here, the mixture collapses to a single dominant component with higher entropy but lower JS divergence, implying that the model has less structural ambiguity. Such behaviour aligns with Grayson's non-local control regime, where wetter soils and greater connectivity generate more coherent, landscape-scale moisture responses.

6 | Conclusion

This study demonstrates that probabilistic deep learning can provide a meaningful representation of soil moisture dynamics and variability across continental scales, even when operating under the severe constraints imposed by sparse, heterogeneous in situ observations and the absence of site-specific calibration. The Gaussian Mixture LSTM captures temporal behaviour and relative wetting-drying dynamics at ungauged locations, while exhibiting systematic biases in absolute volumetric water content, particularly at greater depths. These biases reflect missing structural information, such as porosity and rooting characteristics, rather than deficiencies in the probabilistic formulation.

Analysis of predictive uncertainty reveals that uncertainty is spatially structured and linked to physiographic and soil-textural controls. Measures of dispersion and distributional complexity indicate higher variability and multimodality at intermediate soil moisture levels, consistent with findings from plot- and catchment-scale studies. This suggests that some moisture-dependent patterns persist across scales, even when fine-scale heterogeneity is not explicitly resolved.

Overall, the results indicate that probabilistic deep learning can provide physically interpretable uncertainty information and highlight scale-robust features of soil moisture variability, while also clarifying the limits of ungauged prediction imposed by missing local structural controls.

Acknowledgments

Open Access funding enabled and organized by Projekt DEAL.

Funding

This research has been supported by the Deutsche Forschungsgemeinschaft (German Research Foundation - DFG) via the project - Implementation of an InfraStructure for dAta-BasEd Learning in environmental sciences (ISABEL - grant no. 496155047). The article processing charges for this open-access publication were covered by the Karlsruhe Institute of Technology (KIT).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets used in this study are available at <https://doi.org/10.5281/zenodo.18016491>. These data include all processed soil moisture observations, meteorological forcings, and static attributes required to reproduce the analyses. The full model implementation is available at <https://doi.org/10.5281/zenodo.18219621>.

References

- Bárdossy, A., and W. Lehmann. 1998. "Spatial Distribution of Soil Moisture in a Small Catchment. Part 1: Geostatistical Analysis." *Journal of Hydrology* 206: 1–15.
- Bartholomé, E., and A. S. Belward. 2005. "GLC2000: A New Approach to Global Land Cover Mapping From Earth Observation Data." *International Journal of Remote Sensing* 26: 1959–1977. <https://doi.org/10.1080/01431160412331291297>.
- Bischof, B., D. Klotz, H. V. Gupta, E. Zehe, and R. Loritz. 2026. "Exploring Soil Moisture Dynamics and Variability Across Scales and Geological Settings Using Gaussian Mixture-Long Short-Term Memory Networks." *Journal of Hydrology* 664: 134364. <https://doi.org/10.1016/j.jhydrol.2025.134364>.
- Blöschl, G., and M. Sivapalan. 1995. "Scale Issues in Hydrological Modelling: A Review." *Hydrological Processes* 9: 251–290.
- Brocca, L., L. Ciabatta, C. Massari, S. Camici, and A. Tarpanelli. 2017. "Soil Moisture for Hydrological Applications: Open Questions and New Opportunities." *Water* 9: 140. <https://doi.org/10.3390/w9020140>.
- Bronstert, A., B. Creutzfeldt, T. Graeff, et al. 2012. "Potentials and Constraints of Different Types of Soil Moisture Observations for Flood Simulations in Headwater Catchments." *Natural Hazards* 60: 879–894. <https://doi.org/10.1007/s11069-011-9874-9>.
- Calamita, G., L. Brocca, A. Perrone, et al. 2012. "Electrical Resistivity and TDR Methods for Soil Moisture Estimation in Central Italy Test Sites." *Journal of Hydrology* 454–455: 101–112. <https://doi.org/10.1016/j.jhydrol.2012.06.001>.
- CW3E. 2019. AR Summary and Outlook: Landfalling ARs 3–10 January 2019. Center for Western Weather and Water Extremes. Accessed 16 December 2025. https://cw3e.ucsd.edu/wp-content/uploads/2019/01/jan102019_outlook/10jan19_summary_and_outlook.pdf.
- Darscheid, P., A. Guthke, and U. Ehret. 2018. "A Maximum-Entropy Method to Estimate Discrete Distributions From Samples Ensuring Nonzero Probabilities." *Entropy* 20: 601. <https://doi.org/10.3390/e2008060>.
- Dorigo, W., A. Xaver, M. Vreugdenhil, et al. 2011. "The International Soil Moisture Network: A Data Hosting Facility for Global In Situ Soil Moisture Measurements." *Hydrology and Earth System Sciences* 15: 1675–1698. <https://doi.org/10.5194/hess-15-1675-2011>.
- Entekhabi, D., E. G. Njoku, P. E. O'Neill, et al. 2010. "The Soil Moisture Active Passive (SMAP) Mission." *Proceedings of the IEEE* 98: 704–716. <https://doi.org/10.1109/JPROC.2010.2043918>.
- Famiglietti, J. S., D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson. 2008. "Field Observations of Soil Moisture Variability Across Scales." *Water Resources Research* 44: W01423. <https://doi.org/10.1029/2006WR005804>.
- Fang, K., D. Kifer, K. Lawson, and C. Shen. 2020. "Evaluating the Potential and Challenges of an Uncertainty Quantification Method for Long Short-Term Memory Models for Soil Moisture Predictions." *Water Resources Research* 56: e2020WR028095. <https://doi.org/10.1029/2020WR028095>.
- Geng, Q., S. Yan, and Q. Li. 2024. "Enhancing Data-Driven Soil Moisture Modeling With Physically-Guided LSTM Networks." *Frontiers in Forests and Global Change* 7: 1353011. <https://doi.org/10.3389/ffgc.2024.1353011>.
- Grayson, R., and G. Blöschl, eds. 2001. *Spatial Patterns in Catchment Hydrology: Observations and Modelling*. Cambridge University Press.
- Grayson, R. B., A. W. Western, F. H. S. Chiew, and G. Blöschl. 1997. "Preferred States in Spatial Soil Moisture Patterns: Local and Nonlocal Controls." *Water Resources Research* 33: 2897–2908. <https://doi.org/10.1029/97WR02174>.
- Guo, Z., P. A. Dirmeyer, and T. DelSole. 2011. "Land Surface Impacts on Subseasonal and Seasonal Predictability." *Geophysical Research Letters* 38: L24812. <https://doi.org/10.1029/2011GL049945>.
- Han, Q., Y. Zeng, L. Zhang, et al. 2023. "Global Long-Term Daily 1 km Surface Soil Moisture Dataset With Physics-Informed Machine Learning." *Scientific Data* 10: 101. <https://doi.org/10.1038/s41597-023-02011-7>.
- Hartmann, J., and N. Moosdorf. 2012. "The New Global Lithological Map Database GLiM: A Representation of Rock Properties at the Earth Surface." *Geochemistry, Geophysics, Geosystems* 13: Q12004. <https://doi.org/10.1029/2012GC004370>.
- Hauser, M., R. Orth, and S. I. Seneviratne. 2016. "Role of Soil Moisture Versus Recent Climate Change for the 2010 Heat Wave in Russia." *Geophysical Research Letters* 43: 2819–2826. <https://doi.org/10.1002/2016GL068036>.
- Hengl, T., J. M. de Jesus, R. A. MacMillan, et al. 2014. "SoilGrids1 km—Global Soil Information Based on Automated Mapping." *PLoS One* 9: e105992. <https://doi.org/10.1371/journal.pone.0105992>.
- Herbst, M., and B. Diekkrüger. 2003. "Modelling the Spatial Variability of Soil Moisture in a Micro-Scale Catchment and Comparison With Field Data Using Geostatistics." *Physics and Chemistry of the Earth, Parts A/B/C* 28: 239–245. [https://doi.org/10.1016/S1474-7065\(03\)00033-0](https://doi.org/10.1016/S1474-7065(03)00033-0).
- Hillel, D. 1982. *Introduction to Soil Physics*. Academic Press.

- Humphrey, V., A. Berg, P. Ciais, et al. 2021. "Soil Moisture–Atmosphere Feedback Dominates Land Carbon Uptake Variability." *Nature* 592: 65–69. <https://doi.org/10.1038/s41586-021-03325-5>.
- Jackisch, C., C. Angermann, N. Allroggen, et al. 2017. "Form and Function in Hillslope Hydrology: In Situ Imaging and Characterization of Flow-Relevant Structures." *Hydrology and Earth System Sciences* 21: 3749–3775. <https://doi.org/10.5194/hess-21-3749-2017>.
- Klotz, D., F. Kratzert, M. Gauch, et al. 2022. "Uncertainty Estimation With Deep Learning for Rainfall–Runoff Modeling." *Hydrology and Earth System Sciences* 26: 1673–1693. <https://doi.org/10.5194/hess-26-1673-2022>.
- Koster, R. D., P. A. Dirmeyer, Z. Guo, et al. 2004. "Regions of Strong Coupling Between Soil Moisture and Precipitation." *Science* 305: 1138–1140. <https://doi.org/10.1126/science.1100217>.
- Kratzert, F., D. Klotz, S. Hochreiter, and G. S. Nearing. 2021. "A Note on Leveraging Synergy in Multiple Meteorological Data Sets With Deep Learning for Rainfall–Runoff Modeling." *Hydrology and Earth System Sciences* 25: 2685–2703. <https://doi.org/10.5194/hess-25-2685-2021>.
- Lawrence, J. E., and G. M. Hornberger. 2007. "Soil Moisture Variability Across Climate Zones." *Geophysical Research Letters* 34: L20402. <https://doi.org/10.1029/2007GL031382>.
- Li, Y., S. Yan, and J. Gong. 2023. "Quantifying Uncertainty in Soil Moisture Retrieval Using a Bayesian Neural Network Framework." *Computers and Electronics in Agriculture* 215: 108414. <https://doi.org/10.1016/j.compag.2023.108414>.
- Linke, S., B. Lehner, T. Fellmann, et al. 2019. "Global Hydro-Environmental Sub-Basin and River Reach Characteristics at High Spatial Resolution." *Scientific Data* 6: 283. <https://doi.org/10.1038/s41597-019-0300-6>.
- Liu, J., F. Rahmani, K. Lawson, and C. Shen. 2022. "A Multiscale Deep Learning Model for Soil Moisture Integrating Satellite and In Situ Data." *Geophysical Research Letters* 49: e2021GL096847. <https://doi.org/10.1029/2021GL096847>.
- Loritz, R., A. Kleidon, C. Jackisch, et al. 2019. "A Topographic Index Explaining Hydrological Similarity by Accounting for the Joint Controls of Runoff Formation." *Hydrology and Earth System Sciences* 23: 3807–3821. <https://doi.org/10.5194/hess-23-3807-2019>.
- Mälikke, M., S. K. Hassler, T. Blume, M. Weiler, and E. Zehe. 2020. "Soil Moisture: Variable in Space but Redundant in Time." *Hydrology and Earth System Sciences* 24: 2633–2653. <https://doi.org/10.5194/hess-24-2633-2020>.
- Metzger, M. J., R. G. H. Bunce, R. H. G. Jongman, R. Sayre, A. Trabucco, and R. Zomer. 2013. "High-Resolution Bioclimate Map of the World." *Global Ecology and Biogeography* 22: 630–638. <https://doi.org/10.1111/geb.12022>.
- Muñoz-Sabater, J., E. Dutra, A. Agustí-Panareda, et al. 2021. "ERA5-Land: A State-Of-The-Art Global Reanalysis Dataset for Land Applications." *Earth System Science Data* 13: 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>.
- Orth, R. 2021. "Global Soil Moisture Data Derived Through Machine Learning Trained With In Situ Measurements." *Scientific Data* 8: 170. <https://doi.org/10.1038/s41597-021-00964-1>.
- Perry, M. A., and J. D. Niemann. 2008. "Generation of Soil Moisture Patterns at the Catchment Scale by EOF Interpolation." *Hydrology and Earth System Sciences* 12: 39–53. <https://doi.org/10.5194/hess-12-39-2008>.
- Pfister, L., N. Martínez-Carreras, C. Hissler, et al. 2017. "Bedrock Geology Controls on Catchment Storage, Mixing, and Release: A Comparative Analysis of 16 Nested Catchments." *Hydrological Processes* 31: 1828–1845. <https://doi.org/10.1002/hyp.11134>.
- Ramankutty, N., and J. A. Foley. 1999. "Estimating Historical Changes in Global Land Cover: Croplands From 1700 to 1992." *Global Biogeochemical Cycles* 13: 997–1027. <https://doi.org/10.1029/1999GB900046>.
- Reichle, R. H., Q. Liu, R. D. Koster, et al. 2025. *SMAP L4 Global 3-Hourly 9 Km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 8*. NASA National Snow and Ice Data Center Distributed Active Archive Center. <https://doi.org/10.5067/02LGW4DGJYRX>.
- Roberts, T. M., I. Colwell, C. Chew, S. Lowe, and R. Shah. 2022. "A Deep-Learning Approach to Soil Moisture Estimation With GNSS-R." *Remote Sensing* 14: 3299. <https://doi.org/10.3390/rs14143299>.
- Robinson, D. A., C. S. Campbell, J. W. Hopmans, et al. 2008. "Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review." *Vadose Zone Journal* 7: 358–389. <https://doi.org/10.2136/vzj2007.0143>.
- Robinson, N., J. Regetz, and R. P. Guralnick. 2014. "EarthEnv-DEM90: A Nearly-Global, Void-Free, Multi-Scale Smoothed, 90 m Digital Elevation Model From Fused ASTER and SRTM Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 57–67. <https://doi.org/10.1016/j.isprsjprs.2013.11.002>.
- Seneviratne, S., D. Lüthi, M. Litschi, et al. 2006. "Land–Atmosphere Coupling and Climate Change in Europe." *Nature* 443: 205–209. <https://doi.org/10.1038/nature05095>.
- Seneviratne, S. I., T. Corti, E. L. Davin, et al. 2010. "Investigating Soil Moisture–Climate Interactions in a Changing Climate: A Review." *Earth Science Reviews* 99: 125–161. <https://doi.org/10.1016/j.earscirev.2010.02.004>.
- Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27: 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Teuling, A. J., and P. A. Troch. 2005. "Improved Understanding of Soil Moisture Variability Dynamics." *Geophysical Research Letters* 32: L05404. <https://doi.org/10.1029/2004GL021935>.
- Thornton, P. E., M. M. Thornton, B. W. Mayer, et al. 2021. *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4*. ORNL DAAC. <https://doi.org/10.3334/ORNLDAAAC/2129>.
- US Drought Monitor. 2019. U.S. Drought Monitor Map Archive, University of Nebraska–Lincoln, USDA, and NOAA, Accessed 16 December 2025. <https://droughtmonitor.unl.edu>.
- Vereecken, H., J. A. Huisman, Y. Pachepsky, et al. 2014. "On the Spatio-Temporal Dynamics of Soil Moisture at the Field Scale." *Journal of Hydrology* 516: 76–96. <https://doi.org/10.1016/j.jhydrol.2013.11.061>.
- Vergopolan, N., N. W. Chaney, M. Pan, et al. 2021. "SMAP-HydroBlocks, a 30-m Satellite-Based Soil Moisture Dataset for the Conterminous US." *Scientific Data* 8: 264. <https://doi.org/10.1038/s41597-021-01050-2>.
- Western, A. W., R. B. Grayson, and G. Blöschl. 2002. "Scaling of Soil Moisture: A Hydrologic Perspective." *Annual Review of Earth and Planetary Sciences* 30: 149–180. <https://doi.org/10.1146/annurev.earth.30.091201.140434>.
- Xi, X., Q. Zhuang, and X. Liu. 2025. "A Hybrid Physics-Guided Deep Learning Modeling Framework for Predicting Surface Soil Moisture." *Journal of Geophysical Research: Machine Learning and Computation* 2: e2025JH000682. <https://doi.org/10.1029/2025JH000682>.
- Xia, Y., K. Mitchell, M. Ek, et al. 2012. "Continental-Scale Water and Energy Flux Analysis and Validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2): 1. Intercomparison and Application of Model Products." *Journal of Geophysical Research: Atmospheres* 117: D03109. <https://doi.org/10.1029/2011JD016048>.
- Zehe, E., R. Becker, A. Bárdossy, and E. Plate. 2005. "Uncertainty of Simulated Catchment Runoff Response in the Presence of Threshold Processes: Role of Initial Soil Moisture and Precipitation." *Journal of Hydrology* 315: 183–202. <https://doi.org/10.1016/j.jhydrol.2005.03.038>.

- Zehe, E., and G. Blöschl. 2004. "Predictability of Hydrologic Response at the Plot and Catchment Scales: Role of Initial Conditions." *Water Resources Research* 40: W10202. <https://doi.org/10.1029/2003WR002869>.
- Zehe, E., T. Graeff, M. Morgner, A. Bauer, and A. Bronstert. 2010. "Plot and Field Scale Soil Moisture Dynamics and Subsurface Wetness Control on Runoff Generation in a Headwater in the Ore Mountains." *Hydrology and Earth System Sciences* 14: 873–889. <https://doi.org/10.5194/hess-14-873-2010>.
- Zehe, E., and M. Sivapalan. 2008. "Threshold Behaviour in Hydrological Systems and Geo-Ecosystems: Manifestations, Controls and Implications for Predictability." *Hydrology and Earth System Sciences Discussions* 5: 2765–2827.

Appendix A

Detailed Description of Different Input Variables

TABLE A1 | Summary table of the used ERA5-Land dynamic variables including general information and units.

Name	Supporting information	Units
2-m temperature	Air temperature 2m above surface	K
Potential evaporation	Total water evaporation, including transpiration	m (water equivalent)
Surface pressure	Atmospheric pressure at surface level	Pa
Runoff	Sum of surface and subsurface runoff	m
Total precipitation	Accumulated rainfall and snowfall	m
Surface net solar radiation	Solar radiation reaching the Earth's surface	J m^{-2}
Soil temperature	Soil temperature in four different soil layers (0–7 cm, 7–28 cm, 28–100 cm, and 100–289 cm) of the ECMWF Integrated Forecasting System.	K
Forecast albedo	Reflectivity of Earth's surface	Dimensionless (0–1)
Volumetric water content	Volume of water in four different soil layers (0–7 cm, 7–28 cm, 28–100 cm, and 100–289 cm) of the ECMWF Integrated Forecasting System.	$\text{m}^3 \text{m}^{-3}$

TABLE A2 | Summary table of the used NLDAS dynamic variables including general information and units.

Name	Supplementary information	Units
Albedo	Surface albedo, the fraction of incoming shortwave radiation that is reflected by the land surface	Dimensionless (0–1)
Temperature	2-m, near-surface air-temperature	K
Potential evapotranspiration	The amount of water that would evaporate and transpire if sufficient moisture were available. In NLDAS, PET is derived by either using the Penman-Monteith or Priestley-Taylor methods	$\text{kg m}^{-2} \text{s}^{-1}$
Surface pressure	Surface air pressure at the land surface	Pa
Precipitation	The flux of liquid or frozen water reaching the surface	m

TABLE A3 | Summary table of the used DayMet dynamic variables including general information and units.

Name	Supplementary information	Units
Shortwave radiation	Daily average shortwave radiation at the surface, representing incoming solar energy	W/m^2
Max temperature	Daily maximum 2-m air temperature	K
Min temperature	Daily minimum 2-m air temperature	K
Vapour pressure	Daily average vapour pressure near the surface, derived from station humidity records	Pa
Precipitation	Daily total precipitation, including both rainfall and snowfall, interpolated from station observations	mm/day

TABLE A4 | Summary table of the used Hydro ATLAS static attributes including general information and units.

Name	Supplementary information	Units
Climate zones	The Global Environmental Stratification (GENS) is a statistically derived global bioclimate classification, aggregated into 18 environmental zones (Metzger et al. 2013)	class
Elevation	EarthEnv-DEM90 digital elevation model that provides elevation values for a pixel resolution of 3 arc-seconds (approximately 90 m at the equator) (Robinson et al. 2014)	m
Slope degree	EarthEnv-DEM90 digital elevation model that provides elevation values for a pixel resolution of 3 arc-seconds (approximately 90 m at the equator) (Robinson et al. 2014). Slope values were computed at 3 arc-second resolution based on Horn's method, with latitudinal corrections for the distortion in the XY spacing of geographic coordinates by approximating the geodesic distance between cells.	degree
Clay fraction	SoilGrids1km contains spatial predictions for a selection of soil properties (at six standard depths) including sand, silt and clay fractions as well as soil organic carbon stocks (Hengl et al. 2014)	Dimensionless (0–1)
Silt fraction	SoilGrids1km contains spatial predictions for a selection of soil properties (at six standard depths) including sand, silt and clay fractions as well as soil organic carbon stocks (Hengl et al. 2014)	Dimensionless (0–1)
Sand fraction	SoilGrids1km contains spatial predictions for a selection of soil properties (at six standard depths) including sand, silt and clay fractions as well as soil organic carbon stocks (Hengl et al. 2014)	Dimensionless (0–1)
Land cover	GLC2000 (Global Land Cover in the year 2000) database that distinguishes 22 land cover classes (Bartholomé and Belward 2005)	class
Potential natural vegetation	EarthStat database including a global map of natural vegetation classified into 15 vegetation types (Ramankutty and Foley 1999)	class
Lithological class	Global Lithological Map (GLiM) database was assembled from geological maps with a target resolution of 1:1 million and ideally with a national extent or larger, ranging from 1965 to 2012, and translated into lithological information with the help of regional literature (Hartmann and Moosdorf 2012)	class

Appendix B

Selected Hyperparameters

TABLE B1 | Summary of hyperparameters.

Hyperparameter	Value(s)
Sequence length	180
Number of layers	2
Hidden size	128
Batch size	64
Number of epochs	10
Dropout	0.3
Learning rate	0.0005
Number of mixtures	3
Noise	0.05

Appendix C

PIT Diagrams and Skill Retention Ration (CRPS/RMSE)

Probability Integral Transform

The probability integral transform (PIT) evaluates how well the predictive distribution aligns with the observed outcomes by mapping each observation to its prediction-based percentile. For a predictive cumulative distribution function F_t and observation y_t , the PIT value is defined as $u_t = F_t(y_t)$. If the predicted distributions are correctly calibrated, these PIT values follow a uniform distribution on $[0,1]$. A PIT diagram visualizes the empirical distribution of the u_t values and provides an insight: deviations from uniformity reveal systematic biases or misrepresentation of uncertainty, such as underdispersion, overdispersion, or skewness in the predicted distributions.

CRPS—RMSE Ratio and LOWESS Trend

The CRPS—RMSE ratio provides a simple way to relate probabilistic to deterministic performance by comparing the CRPS and RMSE scores. Since CRPS reduces RMSE for a deterministic forecast, their ratio offers an interpretable measure of how much “skill” is retained when moving from point predictions to full predictive distributions. A ratio close to one indicates that the probabilistic forecast offers limited improvement over a deterministic counterpart, while lower values reflect gains achieved through well-calibrated and sharp uncertainty information. The LOWESS curve is a non-parametric method that fits a smooth curve to a scatterplot of data, revealing underlying trends without assuming specific data structure.

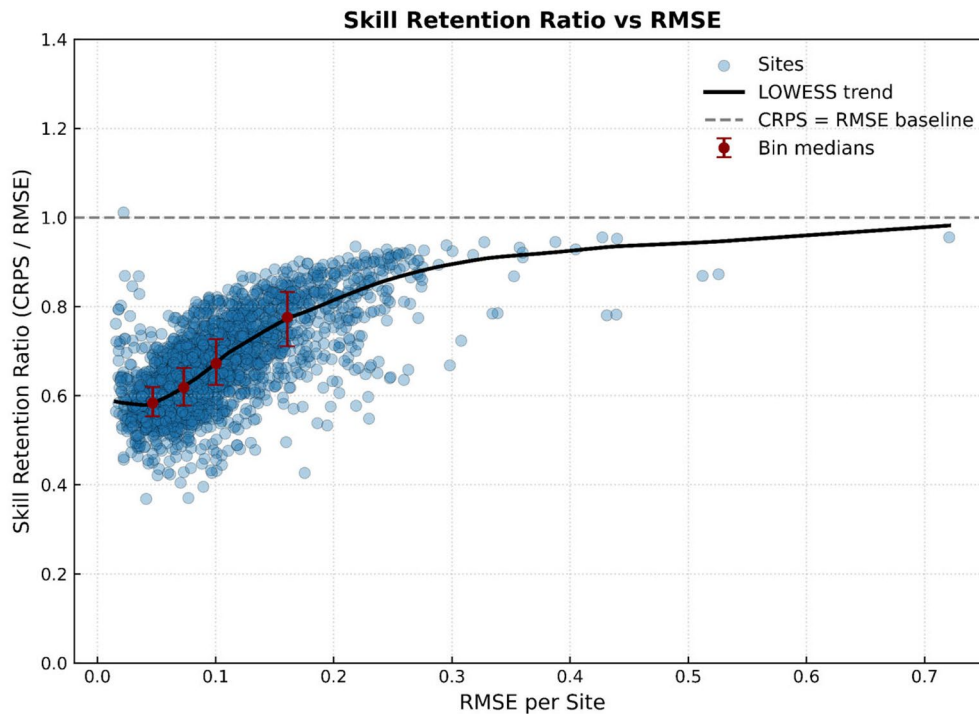


FIGURE C1 | CRPS—RMSE ratio and LOWESS trend for the model predictions across all folds.

Appendix D

Relationships Between Static Attributes and Uncertainties

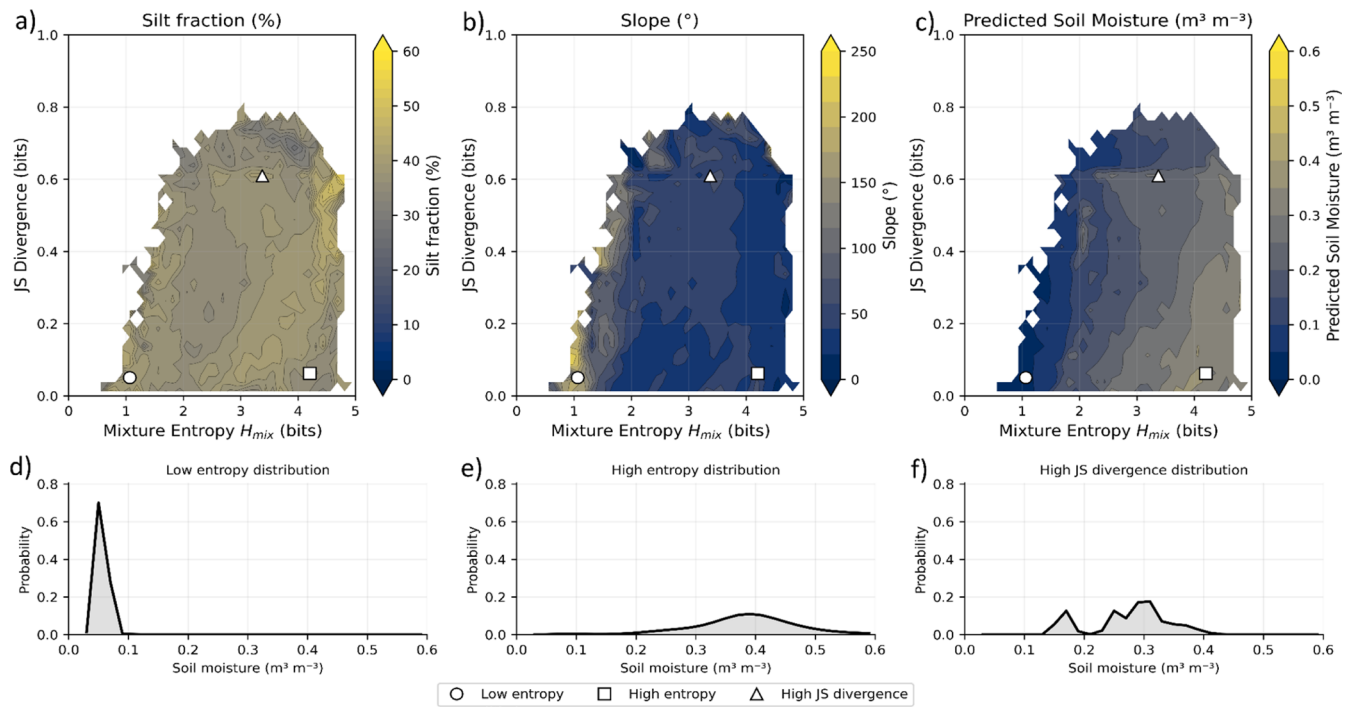


FIGURE D1 | Panels (A–C) show the relationships between mixture dispersion (H_{mix}), distributional complexity (JS divergence), and different environmental attributes. Panels (D–F) depict example distributions for three different parts of the plot (i.e., low entropy, high entropy, and high JS divergence). The figure summarizes predictions for a chosen date (15 January 2019) for all grid cells within the CONUS.