

Geophysical Research Letters[®]



RESEARCH LETTER

10.1029/2025GL118607

Key Points:

- A physics-aware generative model to estimate subsurface soil moisture using only surface observations
- The ensemble size plays a critical role in improving the stability and reliability of predictions, with diminishing gains beyond a certain threshold
- The approach is validated across 24 geographically distributed sites representing various climatic zones and tested on hourly and 10-min temporal resolutions data to assess its spatial and temporal generalizability

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Singh and K. Gaurav,
a.singh4@leeds.ac.uk;
abhilash.iiserb@gmail.com;
kgaurav@iiserb.ac.in

Citation:

Singh, A., Singh, V., & Gaurav, K. (2025). Physics-aware probabilistic modeling of subsurface soil moisture using diffusion processes across different climate settings. *Geophysical Research Letters*, 52, e2025GL118607. <https://doi.org/10.1029/2025GL118607>

Received 31 JUL 2025

Accepted 3 OCT 2025

Author Contributions:

Conceptualization: Abhilash Singh, Kumar Gaurav

Data curation: Vidhi Singh

Formal analysis: Vidhi Singh

Funding acquisition: Kumar Gaurav

Investigation: Abhilash Singh,

Vidhi Singh, Kumar Gaurav

Methodology: Abhilash Singh,

Vidhi Singh, Kumar Gaurav

Software: Abhilash Singh

Supervision: Kumar Gaurav

Validation: Abhilash Singh

© 2025. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Physics-Aware Probabilistic Modeling of Subsurface Soil Moisture Using Diffusion Processes Across Different Climate Settings

Abhilash Singh¹ , Vidhi Singh² , and Kumar Gaurav² 

¹School of Mathematics, Faculty of Engineering and Physical Sciences, University of Leeds, Leeds, UK, ²Fluvial Geomorphology and Remote Sensing Laboratory, Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, India

Abstract We developed a physics-aware denoising diffusion based probabilistic model for estimating subsurface soil moisture from surface observations. Unlike traditional physical-based methods that rely on site-specific soil parameters, our approach leverages a data-driven framework constrained by smoothness and Fickian diffusion principles to ensure physically consistent predictions. The model is trained and evaluated on hourly soil moisture data from 20 globally distributed sites, and further validated on high-resolution 10-min observations from four African stations. The results demonstrate robust performance across depths (10–40 cm), with the model maintaining high accuracy and low bias, even under varying temporal resolutions. We also analyzed the effect of input noise through a structured uncertainty experiment, highlighting the model's stability and reliability. By eliminating the need for explicit physical inputs and enabling uncertainty quantification, this framework offers a scalable solution for operational soil moisture monitoring, particularly in data-sparse or heterogeneous regions.

Plain Language Summary Soil moisture is important for agronomy, weather prediction, and managing water resources. However, it is difficult to measure soil moisture in deeper layers of the soil because sensors are expensive and not available everywhere. In this work, we created a model that can estimate subsurface soil moisture using only the surface soil moisture measurements. The model uses patterns learned from real data and follows how water naturally spreads underground. We tested it using data from 20 different places around the world and also on more detailed data collected every 10 min in Africa. The results show that our model gives accurate estimates, even in places with limited information. It can also tell us how certain or uncertain the estimates are. This tool can help farmers, scientists, and weather services make better decisions where detailed soil data is not available.

1. Introduction

Soil moisture plays an important role in regulating land-atmosphere interactions, influencing processes such as infiltration, evapotranspiration, and surface runoff (Li et al., 2022; McColl et al., 2017; A. Singh et al., 2025a). Accurate soil moisture information, particularly at the subsurface depths, is key for hydrological modeling, agricultural decision-making, and climate risk assessment (Wagner et al., 2024). However, in situ measurements of subsurface soil moisture remain spatially sparse and temporally discontinuous due to instrumentation limitations and logistical constraints, especially in remote or under-resourced regions. As a result, there is an acute need for models that can estimate sub-surface soil moisture information from surface observations, which are comparatively easier to obtain from satellites and ground-based sensors (Ahmad et al., 2022; A. Singh & Gaurav, 2024; A. Singh, Niranjannaik, & Gaurav, 2025).

We can broadly categorize the existing approaches to estimate subsurface soil moisture into empirical, physically-based, and data-driven models. Empirical models use simplified relationships derived from site-specific observations, which can be useful for localized applications, but typically lack transferability (Verma & Nema, 2022). Physical models such as the Richards equation or Green-Ampt infiltration rely on well-established hydrological theory but require detailed knowledge of site-specific parameters such as hydraulic conductivity, porosity, and soil water retention characteristics (Lee et al., 2020; Zeng et al., 2018). These parameters are often unavailable or difficult to obtain. The data-driven models, including statistical regressions and machine learning algorithms (Du

Visualization: Abhilash Singh,
Kumar Gaurav
Writing – original draft: Abhilash Singh,
Vidhi Singh, Kumar Gaurav
Writing – review & editing:
Abhilash Singh, Kumar Gaurav

et al., 2024), offer flexibility in capturing nonlinear relationships. Such models are often limited by their dependence on large training data sets and a lack of physical interpretability (V. Singh et al., 2025).

Recent advances in deep learning have opened up new opportunities for modeling soil moisture dynamics, particularly through the use of generative models and physics-informed architectures. These approaches enable the incorporation of structural constraints that reflect physical processes, while retaining the adaptability of data-driven learning frameworks (Xi et al., 2025). Despite recent progress, many existing models continue to depend on explicit physical inputs or predefined soil parameters, which restricts their applicability across diverse landscapes. In addition, moving from hourly to higher-frequency data introduces further challenges, as models must remain stable under increased temporal variability and potential measurement noise (A. Singh et al., 2025a).

This study proposes a probabilistic, physics-aware denoising diffusion model designed to estimate soil moisture at subsurface depths using only surface moisture measurements. The model integrates smoothness and curvature regularization terms inspired by Fickian diffusion theory as weak physics to guide the learning process, without relying on explicit physical parameters. This approach eliminates the dependency on difficult-to-obtain variables while incorporating essential physical principles, thereby enhancing its practicality and ensuring broader applicability across diverse sites and climatic conditions. We evaluate the model across 20 sites with hourly soil moisture observations, representing diverse hydro-climatic conditions, and test its adaptability using 10-min data from four stations in Zambia. This study aims to answer the following key research questions;

- Can a purely data-driven model, augmented with physics-inspired regularization, generalize across diverse hydro-climatic conditions?
- How does the ensemble size affect the stability and reliability of predictions?
- What is the sensitivity of the model to input uncertainty?
- Can the model maintain predictive performance when applied to high-frequency input data?

2. Data and Methods

2.1. In Situ Observations

We have compiled soil moisture data from the International Soil Moisture Network repository (Dorigo et al., 2011, 2021; Gruber et al., 2013). This includes the data from 20 stations across seven globally distributed networks: AMMA-CATCH (2) (Lebel et al., 2009), CTP-SMTMN (7) (Yang et al., 2013), MAQU (1), NAQU (4), NGARI (4) (Dente et al., 2012; Su et al., 2011), SKKU (1) (Nguyen et al., 2017), and TWENTE (1) (van der Velde et al., 2023). Our database covers six different Köppen–Geiger climate classes (Beck et al., 2023). These stations are located in the regions of Benin, China, South Korea, and Netherlands (Figure S1 in Supporting Information S1).

We have used the soil moisture recorded at depths of 5, 10, 20, and 40 cm. This provides a representative subsurface profile of the soil column. Observations are available at an hourly resolution, with temporal coverage ranging from 7 months to 12 years across stations (Table S1 in Supporting Information S1). The AMMA-CATCH stations employed CS616 sensors (Time Domain Reflectometry), while other networks predominantly used Frequency Domain Reflectometry sensors, including ECH2O EC-TM, EC-TM, and 5TM. Additionally, these stations cover a wide spectrum of soil textures and land cover types, enabling a firm evaluation of the model's generalizability. For further validation, we use soil moisture observations available at 10-min resolution at comparable depths from four stations in Zambia: Bbondo (Cwa), Kapululira (BSh), Kasamanda (Aw), and Margaret (Cwa) are managed by the Zambia Meteorological Department (ZMD). These sites are situated in distinct hydro-climatic settings with monitoring duration extending over 2–5 years.

2.2. Methodology

We formulated the problem as a conditional generative modeling, where our goal is to predict soil moisture at 10, 20, and 40 cm depths, denoted as the response variable $\mathbf{y}_0 = [\text{SM}_{10}, \text{SM}_{20}, \text{SM}_{40}] \in \mathbb{R}^3$, given the surface soil moisture at 5 cm, represented as the conditional input $\mathbf{x}_{\text{cond}} = \text{SM}_5 \in \mathbb{R}^1$. A single multi-output model is trained to jointly predict these three depths from the surface input. This formulation is physically motivated by the fact that water transport in unsaturated soils is primarily governed by capillary forces and matric potential gradients, processes formalized by Richards' equation, which extends Darcy's law to incorporate nonlinear hydraulic

properties and gravity-driven flow. While not derived from Fick's law, Richards' equation can be viewed as akin to a nonlinear generalization of Fickian diffusion, where the diffusivity varies with moisture content and the driving gradient is the matric potential (Wu, 2003). Fick's First Law relates moisture flux to concentration gradients, while Fick's Second Law governs the spatiotemporal evolution of moisture content through its second spatial derivative (Paul et al., 2014). To align with this physical understanding, we employ a denoising diffusion probabilistic model (DDPM), which enables stochastic generation of subsurface profiles conditioned on surface data (Nichol & Dhariwal, 2021). Importantly, we embed domain-specific regularization, namely smoothness and curvature constraints derived from discretized diffusion operators, directly into the learning objective. This ensures that the generated moisture profiles are not only data-consistent but also physically plausible, providing a principled bridge between statistical modeling and hydrological theory. To ensure proper temporal generalization and to avoid any leakage of seasonal patterns, the data set for each station was divided chronologically, with the first 70% of the time series used for training and the remaining 30% reserved for testing.

2.2.1. Forward Diffusion Process

The first component of our model is a forward noising process that perturbs the true subsurface soil moisture profiles $\mathbf{y}_0 \in \mathbb{R}^3$ over a series of T discrete time steps. This process simulates the gradual corruption of the original data by Gaussian noise, thereby transforming the data distribution into an approximately isotropic Gaussian as $t \rightarrow T$. The forward process is defined as a Markov chain, as shown in Equations 1 and 2.

$$q(\mathbf{y}_{1:T}|\mathbf{y}_0) = \prod_{t=1}^T q(\mathbf{y}_t|\mathbf{y}_{t-1}), \quad (1)$$

$$q(\mathbf{y}_t|\mathbf{y}_{t-1}) = \mathcal{N}(\mathbf{y}_t; \sqrt{\alpha_t} \mathbf{y}_{t-1}, (1 - \alpha_t) \mathbf{I}_3), \quad (2)$$

where $\alpha_t = 1 - \beta_t \in (0, 1)$ controls the amount of information retained from the previous step, and β_t is a small variance hyperparameter. Due to the properties of Gaussian distributions, it is possible to sample $\mathbf{y}_t$ directly from $\mathbf{y}_0$ using Equation 3.

$$q(\mathbf{y}_t|\mathbf{y}_0) = \mathcal{N}(\mathbf{y}_t; \sqrt{\bar{\alpha}_t} \mathbf{y}_0, (1 - \bar{\alpha}_t) \mathbf{I}_3). \quad (3)$$

where $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. This closed-form solution significantly simplifies the training procedure, allowing the model to learn from single-step noise corruption instead of simulating the full Markov chain.

2.2.2. Reverse Denoising Process and Conditional Generation

To generate realistic subsurface moisture predictions, we model the reverse diffusion process using a neural network that learns to denoise corrupted representations of the true signal at each timestep. The goal is to approximate the reverse-time conditional probability distribution, as shown in Equation 4.

$$p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}_{\text{cond}}), \quad (4)$$

where $\mathbf{y}_t$ is the noisy version of the clean subsurface profile $\mathbf{y}_0$, and $\mathbf{x}_{\text{cond}} \in \mathbb{R}^1$ denotes the surface soil moisture at 5 cm used as a conditioning variable.

Rather than modeling the full reverse distribution directly, we adopt a DDPM and train a neural network $\hat{\epsilon}_\theta: \mathbb{R}^3 \times \{0, \dots, T-1\} \times \mathbb{R}^1 \rightarrow \mathbb{R}^3$ to predict the additive Gaussian noise at step t from $(\mathbf{y}_t, \mathbf{x}_{\text{cond}})$. As described in Section 2.2.1, the forward process allows us to simulate $\mathbf{y}_t$ in closed form as

$$\mathbf{y}_t = \sqrt{\bar{\alpha}_t} \mathbf{y}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3), \quad (5)$$

which enables efficient training without explicitly simulating the full Markov chain.

The denoising network then takes this noisy input, the timestep t , and the conditional feature $\mathbf{x}_{\text{cond}}$ to predict the original noise using Equation 6.

$$\hat{\epsilon}_\theta = \text{Denoiser}(\mathbf{y}_t, t, \mathbf{x}_{\text{cond}}) \quad (6)$$

With the predicted noise, a reconstruction of the clean data can be obtained using Equation 7.

$$\hat{\mathbf{y}}_0 = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{y}_t - \sqrt{1 - \alpha_t} \hat{\epsilon}_\theta(\mathbf{y}_t, t, \mathbf{x}_{\text{cond}}) \right) \quad (7)$$

To train the model, we minimize the mean squared error between the predicted noise and the true noise used to generate $\mathbf{y}_t$, as shown in Equation 8.

$$\mathcal{L}_{\text{data}} = \mathbb{E}_{\mathbf{y}_0, t, \epsilon} \left[\left\| \hat{\epsilon}_\theta(\mathbf{y}_t, t, \mathbf{x}_{\text{cond}}) - \epsilon \right\|_2^2 \right] \quad (8)$$

Substituting the expression for $\mathbf{y}_t$ from Equation 5, we arrive at the explicit training objective, as shown in Equation 9.

$$\mathcal{L}_{\text{data}} = \mathbb{E}_{\mathbf{y}_0, t, \epsilon} \left[\left\| \hat{\epsilon}_\theta \left(\sqrt{\alpha_t} \mathbf{y}_0 + \sqrt{1 - \alpha_t} \epsilon, t, \mathbf{x}_{\text{cond}} \right) - \epsilon \right\|_2^2 \right] \quad (9)$$

This training formulation simplifies the learning process, as it circumvents the need to approximate the full reverse transition distribution and instead learns to directly denoise a noisy input at each timestep.

2.2.3. Physics-Informed Regularization: Smoothness and Curvature

To integrate physical constraints based on Fickian diffusion theory, we introduce two regularization terms into the training objective. The first is a smoothness penalty, as shown in Equation 10.

$$\mathcal{L}_{\text{smooth}} = \sum_{i=1}^{n-1} (\hat{\epsilon}_{i+1} - \hat{\epsilon}_i)^2, \quad (10)$$

which discourages large gradients between adjacent depths and encourages physically coherent, smoothly varying moisture profiles. The second term is a curvature regularization that penalizes high second-order derivatives of the predicted noise profile, as shown in Equation 11.

$$\mathcal{L}_{\text{fick}} = \sum_{i=2}^{n-1} (\hat{\epsilon}_{i+1} - 2\hat{\epsilon}_i + \hat{\epsilon}_{i-1})^2. \quad (11)$$

This term is inspired by Fick's Second Law of diffusion, which states that the temporal change in concentration (or moisture) is proportional to the second spatial derivative of concentration (Paul et al., 2014). In a vertical soil profile with uniform diffusivity D , the law takes the form shown in Equation 12.

$$\frac{\partial \theta(z, t)}{\partial t} = D \cdot \frac{\partial^2 \theta(z, t)}{\partial z^2}. \quad (12)$$

Since the model operates over depth layers, we discretize the second derivative and impose it as a penalty on the predicted noise. The final objective is a weighted sum of all three components, as shown in Equation 13.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda_{\text{smooth}} \cdot \mathcal{L}_{\text{smooth}} + \lambda_{\text{fick}} \cdot \mathcal{L}_{\text{fick}}, \quad (13)$$

where λ_{smooth} and λ_{fick} are hyperparameters. In practice, both λ_{smooth} and λ_{fick} were fixed globally to 0.1 across all sites and depths. This choice was guided by preliminary sensitivity tests, where values in the range $\{0.01, 0.1, 1.0\}$

were evaluated. A weight of 0.1 consistently provided a balance between discouraging unrealistic oscillations and preserving predictive accuracy, whereas larger values led to over-smoothing and smaller values made the constraints negligible. Fixing the weights globally avoids site-specific tuning and maintains the domain-agnostic character of the framework. While the computational form of the smoothness and curvature terms resembles classical penalization schemes used in statistics (e.g., difference penalties in spline smoothing) (Eilers & Marx, 1996), here these operators are explicitly grounded in the physics of subsurface water transport. The first-order difference penalty discourages unrealistic vertical gradients in moisture, aligning with capillary-driven redistribution, whereas the second-order difference penalty corresponds to the discretized form of Fick's Second Law. In this sense, the penalties act as weak, domain-agnostic physical constraints rather than generic regularizers, enabling physically interpretable predictions without requiring site-specific soil parameters. Further details regarding the optimization strategy, including the gradient formulation and Adam update rules (Dubey et al., 2019; Kingma & Ba, 2014; Xie et al., 2024), are provided in the Section S1 in Supporting Information S1. The sampling procedure used to generate ensemble forecasts and quantify uncertainty is also described therein (Section S2 in Supporting Information S1). Additionally, the full set of evaluation metrics and their formulations are outlined in Section S3 in Supporting Information S1 for completeness and reproducibility. The complete pseudocode outlining the proposed algorithm is presented in Algorithm 1 in Supporting Information S1.

3. Results and Discussion

3.1. Model Performance

To evaluate the model performance, we plotted scatter plots between observed and predicted subsurface soil moisture values, along with normalized error histograms representing the distribution of prediction errors (Figures 1a and 1b). To maintain clarity in the visual representation and avoid over-crowding, one representative station was selected from each of the six distinct Köppen–Geiger climate classifications present in the full set of 20 stations for 10 cm depth. The selected sites span diverse climatic regimes: AMMA-CATCH_Belefoungou-Top (Aw)—tropical savanna, CTP-SMTMN_L03 (ET)—tundra, MAQU_NST-24 (Dwc)—cold continental with dry winters, NGARI_SQ14 (BWk)—cold desert, SKKU_SKKU-Jinwicheon-1 (Cwa)—humid subtropical with dry winters, and Twente_ITCSM-11c (Cfb)—temperate oceanic. This selection facilitates the assessment of model performance across varying hydro-climatic conditions.

Across these stations, the model achieved high accuracy, with coefficient of determination (R^2) values ranging from 0.91 to 0.99. The best performance was observed at MAQU_NST-24 (Dwc), where the model yields $R^2 = 0.991$, normalized Root Mean Square Error (nRMSE) = 0.099, and $\sigma = 0.00126$, indicating minimal error and high stability in capturing snow-influenced seasonal soil moisture dynamics. AMMA-CATCH_Belefoungou-Top (Aw) also exhibits excellent results ($R^2 = 0.989$, nRMSE = 0.112), likely due to the pronounced wet and dry seasons that define soil moisture variability in tropical savanna climates.

The Twente_ITCSM-11c (Cfb) site exhibits an R^2 of 0.951 and an nRMSE of 0.36, reflecting the influence of a temperate oceanic climate characterized by evenly distributed precipitation and mild temperature variability. Similarly, CTP-SMTMN_L03 (ET) shows relatively higher uncertainty. Despite a strong $R^2 = 0.913$, the station had an nRMSE (0.320) and error dispersion ($\sigma = 0.0226$), attributed to complex freeze-thaw processes and shallow active layers common in tundra environments.

NGARI_SQ14 (BWk) demonstrates robust performance under arid conditions ($R^2 = 0.933$, nRMSE = 0.266), with nearly zero bias, suggesting the model's capability to predict soil moisture even under conditions of low vegetation and infrequent precipitation. SKKU_SKKU-Jinwicheon-1 (Cwa), located in a monsoon-driven region, shows $R^2 = 0.987$ with nRMSE = 0.155. The moderate increase in prediction error may be linked to rapid transitions between dry winters and wet summers, introducing nonlinearity in soil moisture dynamics.

In summary, our model demonstrates strong generalization across all climatic zones. Stations located in regions with well-defined and predictable seasonal patterns (Aw, Dwc, and Cfb) exhibited lower error metrics and reduced uncertainty. In contrast, regions influenced by freeze-thaw transitions (ET) or monsoonal rainfall variability (Cwa) presented greater challenges, as reflected in increased nRMSE and error dispersion. These results highlight the importance of accounting for regional climate characteristics when evaluating model performance for sub-surface soil moisture prediction. The results for all depths and all stations are summarized in Table S3 in

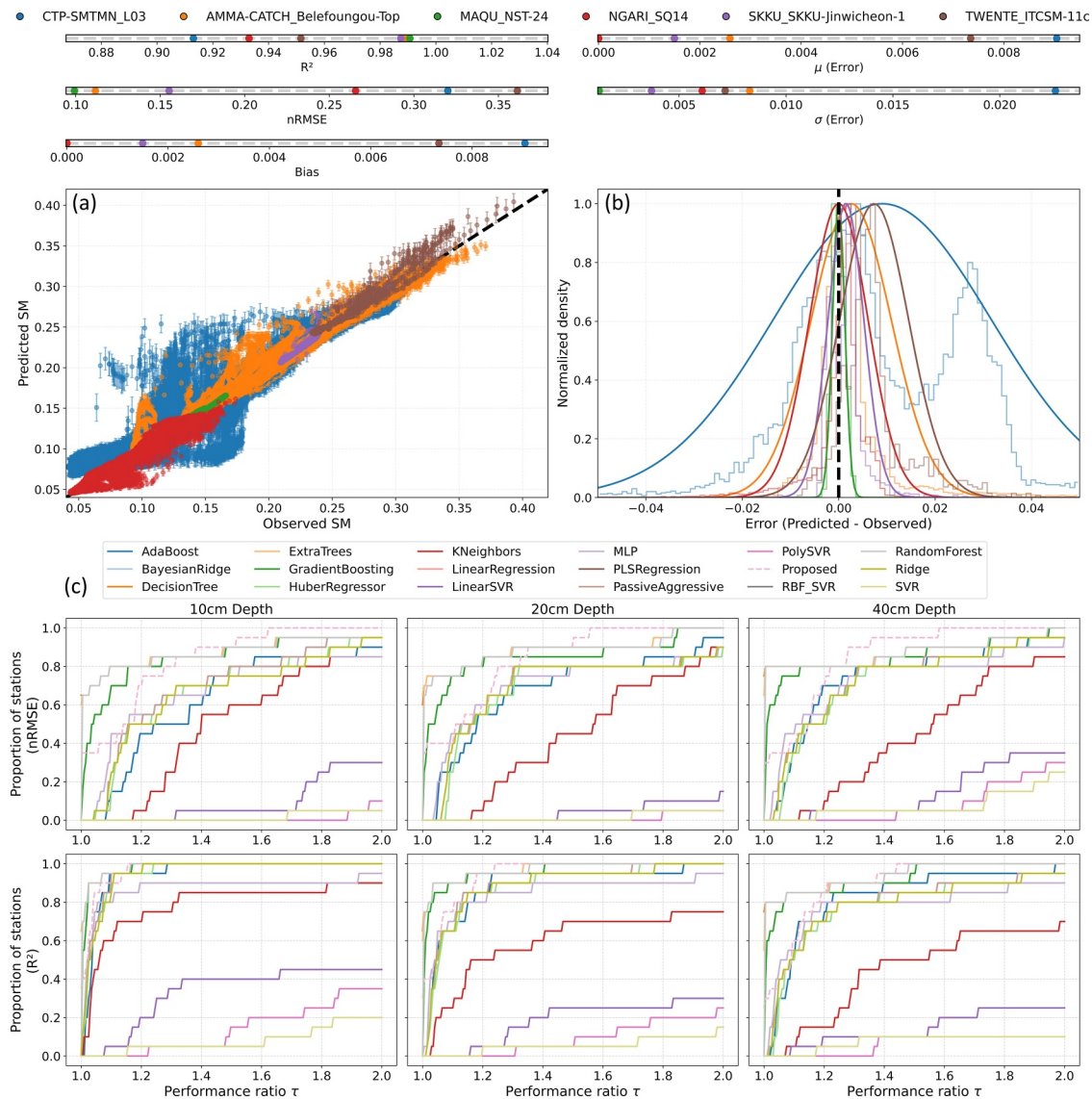


Figure 1. (a) Scatter plot of observed versus predicted sub-surface soil moisture values, where each point represents a prediction instance for a given station. Error bars indicate the ensemble standard deviation (model uncertainty) around the predicted mean. The dashed black line denotes the 1:1 reference line. (b) Normalized error histograms (prediction error = predicted – observed) overlaid with Gaussian fits for each site. Histogram densities and fitted probability density functions are normalized to unity to enable consistent comparison across sites. A vertical dashed line at zero error highlights the bias direction. Inset panels display site-wise summary metrics: R^2 , normalized Root Mean Square Error (nRMSE), and bias for the observed versus predicted comparison (left); and fitted μ and σ of prediction error distributions (right). Each marker corresponds to a station, color-coded consistently across the figure. (c) Dolan-Moré performance profiles for nRMSE (top row) and coefficient of determination (R^2 , bottom row) across all the soil moisture monitoring stations at three different depths (10, 20, and 40 cm).

Supporting Information S1, where a similar performance pattern is observed among stations sharing the same climate classification, further supporting the influence of regional hydro-climatic conditions on model behavior.

3.2. Comparison With Benchmark

For a comprehensive evaluation, the proposed model is compared with a diverse set of 17 other regression algorithms, encompassing ensemble methods, linear regression techniques, support vector machines, and neural networks. The models considered in this study include: AdaBoost, BayesianRidge, DecisionTree, ExtraTrees, GradientBoosting, HuberRegressor, KNeighbors, LinearRegression, LinearSVR, MLP, PLSRegression, PassiveAggressive, PolySVR, Proposed, RBF_SVR, RandomForest, Ridge, and SVR. This wide-ranging comparison framework allows for a systematic assessment of predictive capabilities under varying model structures.

To quantify the generalization ability of each model across stations and soil depths, we constructed the Dolan–Moré performance profiles (Figure 1c). This technique involves computing a performance ratio τ for each model m and station s , normalized relative to the best-performing model at that station. Specifically, for nRMSE and coefficient of determination (R^2), the performance ratios are defined as:

$$\tau_{m,s}^{\text{nRMSE}} = \frac{\text{nRMSE}_{m,s}}{\min_{m'} \text{nRMSE}_{m',s}}, \quad \tau_{m,s}^{R^2} = \frac{\max_{m'} R_{m',s}^2}{R_{m,s}^2} \quad (14)$$

Lower values of τ correspond to better relative performance. The cumulative performance profile $\rho_m(\tau)$ is then computed as the fraction of stations for which a model's performance ratio is less than or equal to a given threshold:

$$\rho_m(\tau) = \frac{1}{|S|} \sum_{s \in S} \mathbb{I}(\tau_{m,s} \leq \tau) \quad (15)$$

where $|S|$ denotes the total number of stations and $\mathbb{I}$ is the indicator function. These profiles are plotted over the interval $\tau \in [1, 2]$. Models that achieve higher $\rho_m(\tau)$ values at lower τ thresholds are considered more robust and generalizable across different test locations. The profiles are interpreted as follows: a model that reaches a higher proportion of stations (closer to 1.0 on the y-axis) at lower values of τ demonstrates superior consistency in performance.

The proposed model exhibits strong performance consistency across all tested depths and evaluation metrics. It was the only model whose profile curve reached a value of 1.0 on the y-axis at the lowest τ , signifying that it either matched or exceeded the best-performing model at every station with the least performance loss. This result highlights the robustness and adaptability of the proposed approach, particularly in handling spatial heterogeneity, and supports its suitability for deployment in diverse environmental conditions.

To evaluate the statistical significance of model performance differences, we conducted a pairwise comparison of absolute errors using Tukey's Honest Significant Difference test across 20 monitoring sites at the 10 cm depth. The resulting heatmaps offer a comprehensive visual summary, where each subplot corresponds to a specific site and highlights model pairs with statistically significant differences ($p < 0.05$), marked by black “X” (Figure S2 in Supporting Information S1). The proposed diffusion-based model stands out by consistently showing significant differences from baseline and benchmark models across a majority of sites. This suggests that its predictive performance is not only superior in magnitude but also statistically distinguishable. The proposed model exhibits the highest number of statistically significant pairwise differences at 12 sites (out of 20), with each registering 17 significant comparisons. These sites include Site 2 to Site 10 and Site 15, 16, and 18. This consistently strong performance across multiple locations reflects the model's broad applicability and effectiveness. Notably, Sites 1, 13, and 19 also demonstrated substantial improvements, each with 14 significant differences, while Site 12 showed 12. Even at Site 11, where overall distinctions among models were less prominent, the proposed model achieved five significant comparisons, further indicating its robustness. These findings highlight the model's robustness and its consistent advantage in estimating subsurface soil moisture under diverse environmental conditions. Results for the 20 and 40 cm depths follow a similar trend and are provided in the Section S5 in Supporting Information S1.

3.3. Stochastic Robustness Analysis

We conducted a stochastic robustness analysis by training the model across 30 independent runs using different random seeds. This approach evaluates the consistency and reliability of model performance under varying initial conditions, a critical consideration in deep learning-based models prone to converging at different local minima. For each run, we computed key performance metrics (R^2 and nRMSE), across all stations and depths (10, 20, and 40 cm). To visualize the results, we used violin plots, which provide a comprehensive representation of the metric distributions, including central tendency, spread, and the presence of outliers (Figure 2a). This evaluation framework enables a robust statistical assessment of the model behavior, extending beyond single-seed reporting practices.

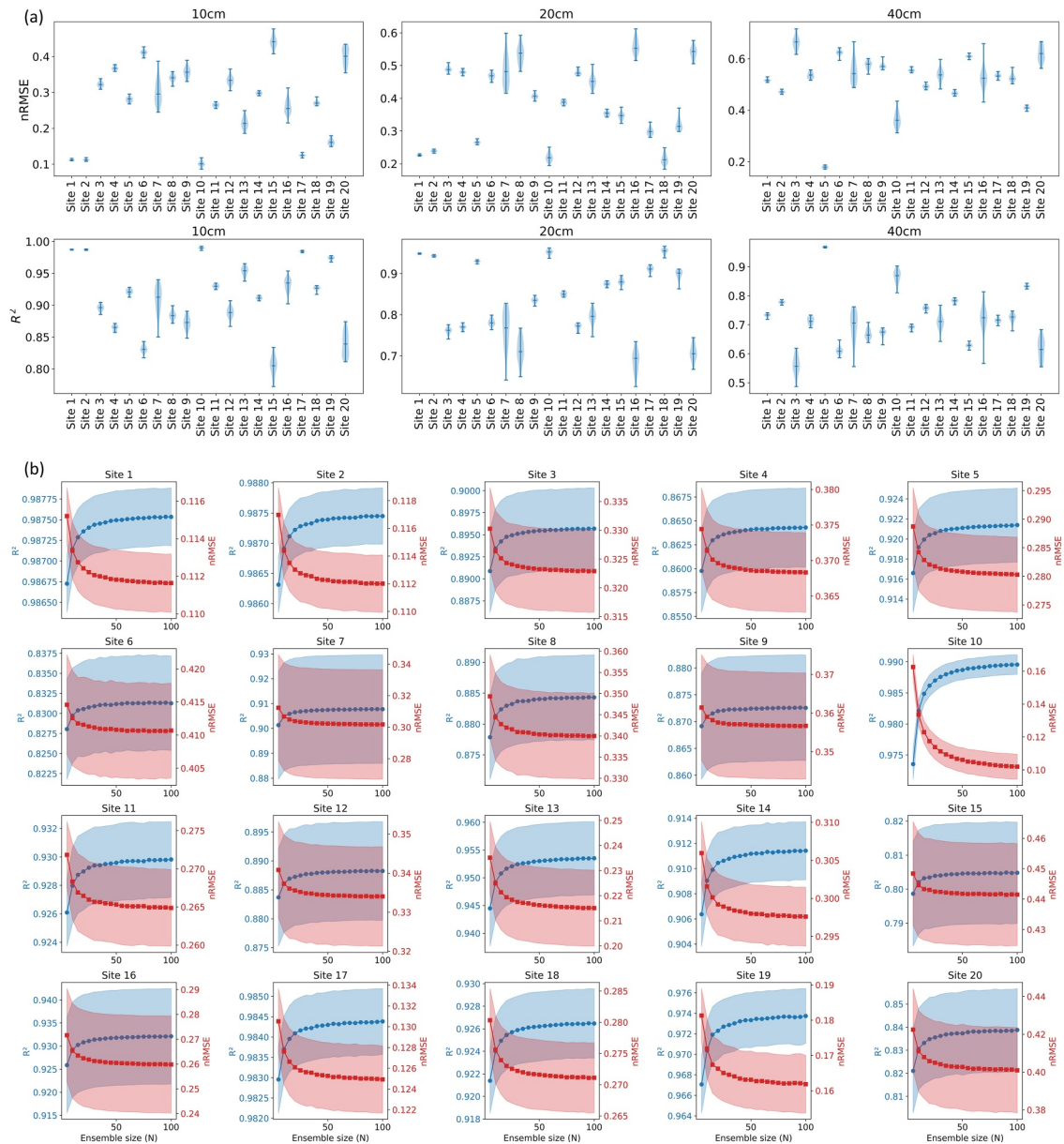


Figure 2. (a) Violin plots showing the distribution of evaluation metrics (normalized Root Mean Square Error (nRMSE) and R^2) for subsurface soil moisture predictions at three depths (10, 20, and 40 cm) across 20 different sites. Each violin represents the variability in performance over 30 independent training seeds for a given site, identified here as Site 1–Site 20 (sorted alphabetically). The top row illustrates the nRMSE while the bottom row displays R^2 . Medians are shown as central white bars. (b) Station-wise performance of the ensemble diffusion model in predicting 10 cm soil moisture, evaluated using the coefficient of determination and nRMSE. Each subplot represents 1 of 20 selected stations. The x-axis shows the ensemble size (N), varying from 5 to 100 in steps of 5. For each N , the solid line depicts the mean value across all random seeds, while the shaded area represents the standard deviation, capturing uncertainty due to seed-based variability.

The violin plots and statistical summaries demonstrate that the model maintains high stability and consistency across different random seeds, particularly at the 10 cm depth. For instance, the average standard deviation of R^2 at 10 cm across all sites is approximately 0.0066, with 75% of sites exhibiting a range below 0.01, indicating minimal sensitivity to initialization. In contrast, at 40 cm, the standard deviation of R^2 increases by an average of 158% compared to 10 cm, and the performance range exceeds 0.02 in over 30% of the sites, suggesting greater variability in deeper predictions. A similar trend is observed in nRMSE values, where the median range increases from approximately 0.0075 at 10 cm to 0.0098 at 40 cm, highlighting how subsurface predictions are more prone to uncertainty propagation from input perturbations. Despite this, approximately 60% of the stations maintain

stable R^2 values above 0.60 even at 40 cm, confirming the model's ability to generalize effectively under stochastic perturbations while capturing depth-dependent soil moisture dynamics.

3.4. Performance at Different N

To assess prediction stability and performance across ensemble configurations, we conducted a stochastic robustness analysis by running the model with 30 random seeds for ensemble sizes ranging from $N = 5$ to $N = 100$ (Figure 2b). At the 10 cm depth, we observed a steady but saturating improvement in model accuracy with increasing ensemble size. The mean R^2 increased from 0.9084 at $N = 5$ to 0.9143 at $N = 100$, reflecting a relative improvement of approximately 0.65%. In parallel, the mean normalized RMSE (nRMSE) decreased from 0.2857 to 0.2739, a 4.14% reduction. Standard deviations across seeds remained low (averaging approximately 0.0066 for R^2 and 0.0104 for nRMSE), indicating high stability of predictions. Beyond $N = 50$, the performance metrics began to saturate, suggesting diminishing returns with further increases in ensemble size. These results highlight that while ensembling contributes to robustness and marginal accuracy gains, a moderately sized ensemble may be sufficient for effective sub-surface soil moisture estimation. A similar trend of saturating improvements is observed at 20 and 40 cm depths, and these results are presented and described in detail in the Section S5 in Supporting Information S1.

3.5. Uncertainty Analysis

To evaluate how uncertainties in input measurements propagate through the model, we conducted a systematic input uncertainty analysis. In doing so, the model was trained to predict subsurface soil moisture at 10, 20, and 40 cm depths using surface soil moisture at 5 cm as input. After training on 70% of each station's time series and validating on the remaining 30%, we generated 100 ensemble predictions to capture natural stochasticity in model outputs for reference. To quantify the sensitivity of the model to input noise, we introduced artificial perturbations to 5 cm input using three representative types of uncertainty: White Noise (WN), Autocorrelated Noise (AN), and Structured Noise (SN). WN simulates high-frequency sensor noise, while AN emulates low-frequency or persistent sensor drift over time, and SN represents systematic, cyclic patterns such as those induced by diurnal or seasonal biases. For each noise type, we applied perturbations at $\pm 1\%$, $\pm 5\%$, and $\pm 10\%$ levels to 50% of the test samples, reflecting realistic ranges of sensor or preprocessing errors. These perturbed inputs were passed through the trained model, and the resulting predictions are compared to the outputs from the unperturbed baseline. The uncertainty effect is quantified as the mean absolute percentage deviation in predicted soil moisture at each depth. The results highlight distinct patterns in sensitivity with respect to both depth and the nature of the input perturbation (Figure 3).

WN induced moderate output deviations across depths. At 10 cm depth, the average effect ranged from approximately 6.76% for $\pm 1\%$ perturbations to 7.42% for $\pm 10\%$, with site-specific variation between 2.1% and 12.5%. The 20 cm layer showed the greatest sensitivity, with average changes increasing from 10.28% ($\pm 1\%$) to 10.78% ($\pm 10\%$), and deviations reaching up to 24.2% at Site 14. In contrast, the 40 cm predictions were comparatively stable, with average effects ranging from 7.82% to 8.18%, and maximum deviations staying below 18.3%, indicating a dampening of WN impacts with depth. On the other hand, AN shows the most significant influence among all noise types. At 10 cm depth, its impact is comparable to WN, with average deviations ranging from 6.78% ($\pm 1\%$) to 9.06% ($\pm 10\%$). However, the effects became notably more pronounced at deeper layers. The model's sensitivity peaked at the 20 cm depth, where the average deviation under $\pm 10\%$ input noise reached 12.14%, with Site 14 exhibiting the highest observed change of 24.77%. Even at 40 cm, the average deviation remained substantial, rising from 7.84% to 9.30% across noise levels, highlighting the pronounced impact of persistent input biases on deeper soil moisture predictions. Finally, the SN also led to consistent performance degradation, particularly at the intermediate depth. At 20 cm, the average impact reached up to 10.51%, indicating notable vulnerability in this layer. In comparison, the perturbation effects at 10 and 40 cm remained lower, with average deviations not exceeding 7.18% and 8.00%, respectively. These results suggest that while structured surface anomalies are less severe than autocorrelated drifts, they still systematically influence soil moisture dynamics, especially in the intermediate soil zone. Overall, 20 cm depth consistently exhibits the highest susceptibility to input noise, with an average deviation of 10.65% across all noise types, indicating that this intermediate layer is most sensitive to uncertainties originating from surface perturbations. The 10 cm layer showed moderate sensitivity (7.26%), while the 40 cm depth is relatively more robust (8.11%). Among the noise types,

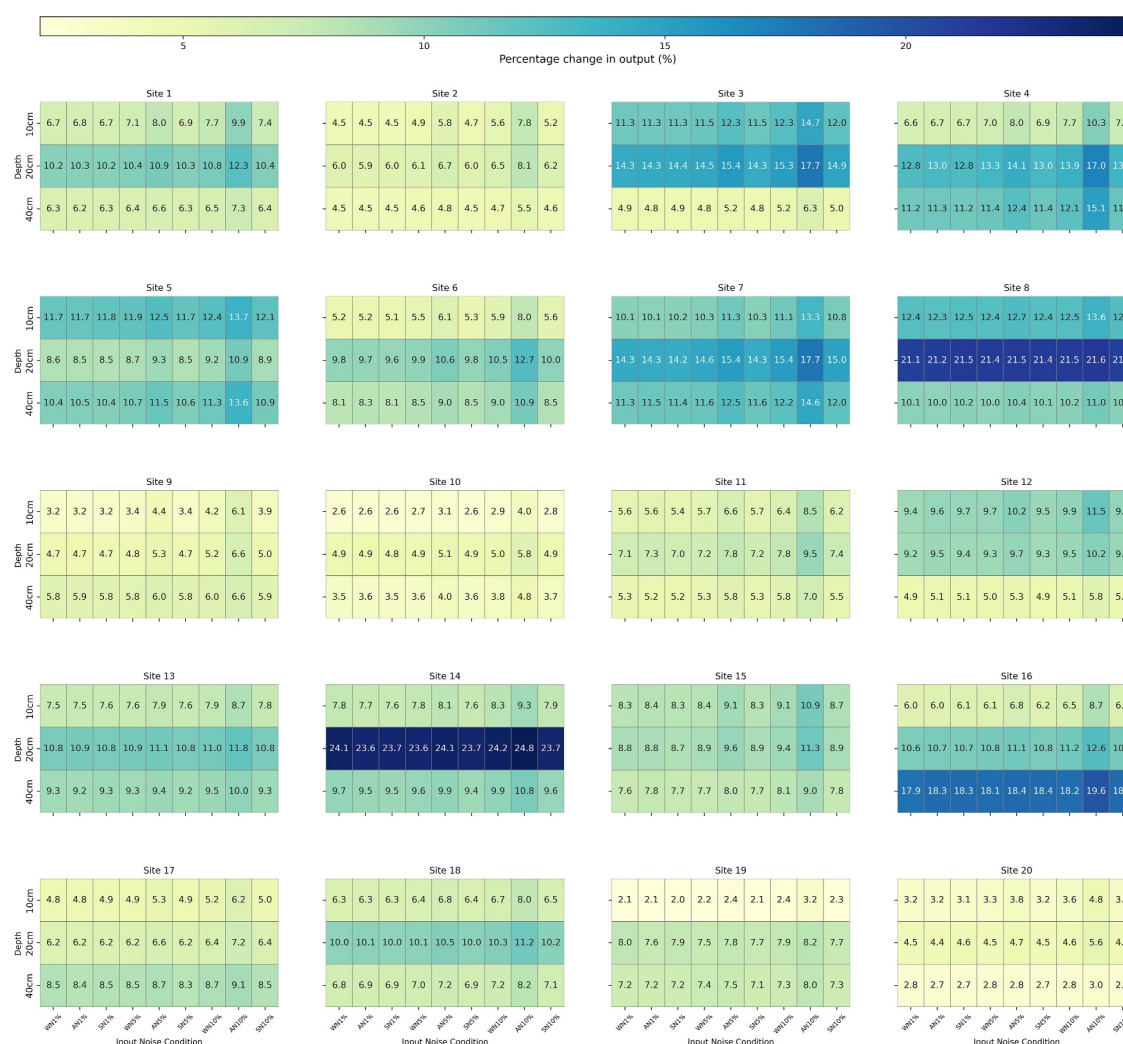


Figure 3. Impact of input uncertainty on model output predictions across all 20 soil monitoring stations. Each subplot represents a distinct site (Site 1–Site 20), ordered alphabetically by station name. The heatmaps display the percentage change in model outputs in response to three types of input perturbations, White Noise, Autocorrelated Noise, and Structured Noise, applied at three uncertainty levels: $\pm 1\%$, $\pm 5\%$, and $\pm 10\%$. Columns represent specific combinations of noise type and uncertainty level (e.g., WN1%, AN5%, SN10%), while rows indicate soil depths (10, 20, and 40 cm). Annotated values in each cell indicate the magnitude of output deviation (in %), and the color intensity reflects the scale of impact.

AN had the strongest impact (up to 11.10% at 20 cm), followed by structured and WN, which showed similar effects. Variability across stations further emphasized site-specific responses, with Site 8 consistently displaying the highest average sensitivity (14.73%), while Site 20 exhibits the most stable behavior (3.66%).

3.6. Performance at High Temporal Resolution Data Sets

Finally, beyond evaluating the model on hourly soil moisture data across 20 diverse stations, we also assessed its performance using higher temporal resolution data from four sites in Zambia, where observations are available at 10-min intervals. This shift from hourly to 10-min data introduces additional challenges, including increased temporal noise, sharper variability in soil moisture signals, and heightened sensitivity to missing or inconsistent measurements. These conditions require the model to demonstrate strong temporal generalization capabilities, particularly in regions where supporting metadata may be sparse. Despite these challenges, the model maintained strong predictive performance at most sites (Figure S7 in Supporting Information S1). Three stations (Bbondon, Kapululira, and Margaret) exhibited high R^2 values ranging from 0.95 to 0.96, with normalized RMSE (nRMSE) values of 0.27 (Kapululira), 0.32 (Margaret), and 0.61 (Bbondon). Kasamanda showed relatively lower accuracy ($R^2 = 0.82$, nRMSE = 0.49), potentially due to increased sub-hourly variability or site-specific noise. Across all

sites, prediction errors remained centered around zero, with low standard deviations, indicating unbiased and consistent performance. These findings demonstrate the model's suitability for high-frequency soil moisture forecasting and its potential for supporting early warning systems and short-term water resource management in data-limited environments.

4. Conclusion

This study presents a physics-aware, probabilistic framework for subsurface soil moisture estimation using a diffusion-based generative model. By formulating the task as a conditional generation problem, the model learns to infer deeper soil moisture profiles solely from surface measurements without any ancillary physical parameters or soil-specific hydraulic constants. This makes the approach inherently scalable and adaptable across heterogeneous landscapes and climate zones. Unlike conventional physics-based models such as Richards' equation or the Green-Ampt infiltration method which require site-specific calibration of parameters like saturated hydraulic conductivity, porosity, or soil water retention curves, our formulation relies entirely on data-driven learning, guided by domain-inspired regularizations. The smoothness and curvature terms derived from Fickian diffusion theory function as weak physical constraints, improving the representation of system behavior while avoiding dependence on hard-to-measure variables and remaining free of parametric assumptions. Extensive evaluation across 20 global sites with hourly observations and four high-resolution African stations with 10-min data confirms the model's ability to produce accurate, robust, and physically consistent predictions. Its generalization across spatial, temporal, and climatic variability highlights the potential of diffusion-based models as a lightweight and transferable alternative for hydrological forecasting in data-scarce or operationally constrained environments.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

In situ data can be downloaded from Zenodo (A. Singh et al., 2025b). The code developed for this study can be downloaded from Zenodo (A. Singh et al., 2025c). The same is also available for download at <https://abhi-lashsingh.net/codes.html>.

Acknowledgments

We gratefully acknowledge the Geohydrology Observatory Facility at IISER Bhopal, where the experiments were conducted, which played a key role in conceptualizing this study. A. S. would like to thank the Zambia Meteorological Department (ZMD) for providing the gauge station data sets used in this study (accessed through University of Leeds). K. G. would like to acknowledge the funding from the Indo-French Center for the Promotion of Advanced Research: IFCPAR/CEFIPRA through a research Grant 6707-1, and an institutional grant received from IISERB.

References

- Ahmad, J. A., Forman, B. A., & Kumar, S. V. (2022). Soil moisture estimation in South Asia via assimilation of SMAP retrievals. *Hydrology and Earth System Sciences*, 26(8), 2221–2243. <https://doi.org/10.5194/hess-26-2221-2022>
- Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., et al. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. *Scientific Data*, 10(1), 724. <https://doi.org/10.1038/s41597-023-02549-6>
- Dente, L., Su, Z., & Wen, J. (2012). Validation of SMOS soil moisture products over the Maqu and Twente regions. *Sensors*, 12(8), 9965–9986. <https://doi.org/10.3390/s120809965>
- Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., et al. (2021). The international soil moisture network: Serving Earth system science for over a decade. *Hydrology and Earth System Sciences Discussions*, 2021(11), 1–83. <https://doi.org/10.5194/hess-25-5749-2021>
- Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., 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(5), 1675–1698. <https://doi.org/10.5194/hess-15-1675-2011>
- Du, J., Kimball, J., Jencso, K., Hoylman, Z., Brust, C., Ketchum, D., et al. (2024). Machine-learning based multi-layer soil moisture forecasts—An application case study of the Montana 2017 flash drought. *Water Resources Research*, 60(10), e2023WR036973. <https://doi.org/10.1029/2023wr036973>
- Dubey, S. R., Chakraborty, S., Roy, S. K., Mukherjee, S., Singh, S. K., & Chaudhuri, B. B. (2019). diffgrad: An optimization method for convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(11), 4500–4511. <https://doi.org/10.1109/tnnls.2019.2955777>
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2), 89–121. <https://doi.org/10.1214/ss/1038425655>
- Gruber, A., Dorigo, W. A., Zwieback, S., Xaver, A., & Wagner, W. (2013). Characterizing coarse-scale representativeness of in situ soil moisture measurements from the international soil moisture network. *Vadose Zone Journal*, 12(2), vzj2012–vzj2170. <https://doi.org/10.2136/vzj2012.0170>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lebel, T., Cappelaere, B., Galle, S., Hanan, N., Kergoat, L., Levis, S., et al. (2009). Amma-catch studies in the Sahelian region of West-Africa: An overview. *Journal of Hydrology*, 375(1–2), 3–13. <https://doi.org/10.1016/j.jhydrol.2009.03.020>
- Lee, S., Chu, M. L., & Schmidt, A. R. (2020). Effective Green-AMPT parameters for two-layered soils. *Journal of Hydrologic Engineering*, 25(4), 04020004. [https://doi.org/10.1061/\(asce\)jhe.1943-5584.0001897](https://doi.org/10.1061/(asce)jhe.1943-5584.0001897)

- Li, W., Migliavacca, M., Forkel, M., Denissen, J. M., Reichstein, M., Yang, H., et al. (2022). Widespread increasing vegetation sensitivity to soil moisture. *Nature Communications*, 13(1), 3959. <https://doi.org/10.1038/s41467-022-31667-9>
- McColl, K. A., Alemohammad, S. H., Akbar, R., Konings, A. G., Yueh, S., & Entekhabi, D. (2017). The global distribution and dynamics of surface soil moisture. *Nature Geoscience*, 10(2), 100–104. <https://doi.org/10.1038/ngeo2868>
- Nguyen, H. H., Kim, H., & Choi, M. (2017). Evaluation of the soil water content using cosmic-ray neutron probe in a heterogeneous monsoon climate-dominated region. *Advances in Water Resources*, 108, 125–138. <https://doi.org/10.1016/j.advwatres.2017.07.020>
- Nichol, A. Q., & Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. In *International conference on machine learning* (pp. 8162–8171).
- Paul, A., Laurila, T., Vuorinen, V., & Divinski, S. V. (2014). Fick's laws of diffusion. In *Thermodynamics, diffusion and the kirkendall effect in solids* (pp. 115–139). Springer.
- Singh, A., & Gaurav, K. (2024). PIML-SM: Physics-informed machine learning to estimate surface soil moisture from multi-sensor satellite images by leveraging swarm intelligence. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–13. <https://doi.org/10.1109/tgrs.2024.3502618>
- Singh, A., Niranjan, M., & Gaurav, K. (2025). Overcoming data scarcity: A transfer learning framework with fine-tuned neural networks and multi-sensor satellite image fusion for soil moisture estimation. *Engineering Applications of Artificial Intelligence*, 159, 111636. <https://doi.org/10.1016/j.engappai.2025.111636>
- Singh, A., Singh, V., & Gaurav, K. (2025a). Leveraging neural operator and sliding window technique for enhanced subsurface soil moisture imputation under diverse precipitation scenarios. *Journal of Geophysical Research: Machine Learning and Computation*, 2(3), e2025JH000730. <https://doi.org/10.1029/2025jh000730>
- Singh, A., Singh, V., & Gaurav, K. (2025b). Physics-aware probabilistic modeling of subsurface soil moisture using diffusion processes across different climate settings [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.16639110>
- Singh, A., Singh, V., & Gaurav, K. (2025c). Physics-aware probabilistic modeling of subsurface soil moisture using diffusion processes across different climate settings [Software]. *Zenodo*. <https://doi.org/10.5281/zenodo.16639888>
- Singh, V., Verma, D., Singh, A., & Gaurav, K. (2025). Multi-depth soil moisture dynamics to rainfall events: An automated machine learning approach. *Hydrology and Earth System Sciences Discussions*, 2025, 1–40. <https://doi.org/10.5194/egusphere-2025-961>
- Su, Z., Wen, J., Dente, L., Van Der Velde, R., Wang, L., Ma, Y., et al. (2011). The Tibetan Plateau observatory of Plateau scale soil moisture and soil temperature (Tibet-OBS) for quantifying uncertainties in coarse resolution satellite and model products. *Hydrology and Earth System Sciences*, 15(7), 2303–2316. <https://doi.org/10.5194/hess-15-2303-2011>
- van der Velde, R., Benninga, H.-J. F., Retsios, B., Vermunt, P. C., & Salama, M. S. (2023). Twelve years of profile soil moisture and temperature measurements in Twente, the Netherlands. *Earth System Science Data*, 15(4), 1889–1910. <https://doi.org/10.5194/essd-15-1889-2023>
- Verma, S., & Nema, M. K. (2022). Development of an empirical model for sub-surface soil moisture estimation and variability assessment in a lesser Himalayan watershed. *Modeling Earth Systems and Environment*, 8(3), 3487–3505. <https://doi.org/10.1007/s40808-021-01316-z>
- Wagner, W., Lindorfer, R., Hahn, S., Kim, H., Vreugdenhil, M., Gruber, A., et al. (2024). Global scale mapping of subsurface scattering signals impacting ASCAT soil moisture retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–20. <https://doi.org/10.1109/tgrs.2024.3429550>
- Wu, L. (2003). Soil water diffusion. *Encyclopedia of Water Science*, 865–867.
- Xi, X., Zhuang, Q., & Liu, X. (2025). A hybrid physics-guided deep learning modeling framework for predicting surface soil moisture. *Journal of Geophysical Research: Machine Learning and Computation*, 2(3), e2025JH000682. <https://doi.org/10.1029/2025jh000682>
- Xie, X., Zhou, P., Li, H., Lin, Z., & Yan, S. (2024). Adan: Adaptive nesterov momentum algorithm for faster optimizing deep models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12), 9508–9520. <https://doi.org/10.1109/tpami.2024.3423382>
- Yang, K., Qin, J., Zhao, L., Chen, Y., Tang, W., Han, M., et al. (2013). A multiscale soil moisture and freeze–thaw monitoring network on the third pole. *Bulletin of the American Meteorological Society*, 94(12), 1907–1916. <https://doi.org/10.1175/bams-d-12-00203.1>
- Zeng, J., Zha, Y., & Yang, J. (2018). Switching the Richards' equation for modeling soil water movement under unfavorable conditions. *Journal of Hydrology*, 563, 942–949. <https://doi.org/10.1016/j.jhydrol.2018.06.069>