



*Research article***VarFedFuse: An optimization-driven Bayesian federated variational fusion framework for uncertainty-aware multi-sensor drought monitoring****Hasnain Iftikhar^{1,2,*}, Paulo Canas Rodrigues^{3,4}, Osama Abdulaziz Alamri⁵ and Basim S. O. Alsaedi⁵**¹ Department of Statistics, University of Peshawar, Peshawar, Pakistan² Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan³ Department of Statistics, Universidade Federal da Bahia, Salvador, Brazil⁴ Department of Business Management, University of Pretoria, Pretoria 0002, South Africa⁵ Department of Statistics, University of Tabuk, Tabuk 71491, Saudi Arabia*** Correspondence:** Email: hasnain@stat.qau.edu.pk.

Abstract: This study proposes VarFedFuse, an optimization-driven Bayesian federated variational fusion framework for privacy-preserving and uncertainty-aware drought monitoring from heterogeneous environmental observations. The framework integrates local Bayesian variational inference, uncertainty-weighted federated posterior aggregation, graph-attention-based adaptive sensor weighting, and transformer-based temporal encoding. Using the law of total variance, the framework further decomposes predictive uncertainty into data uncertainty, uncertainty propagated through the fused latent drought representation, and model-parameter uncertainty, providing non-overlapping, interpretable sources of prediction uncertainty. We evaluated the proposed framework using multi-source environmental observations collected from five drought-prone regions representing North America, Western Europe, the Mediterranean, Central Asia, and Sub-Saharan Africa. We used multi-source observations of the vegetation condition, land-surface temperature, precipitation, temperature, potential evapotranspiration, and soil moisture, along with Standardized Precipitation Evapotranspiration Index (SPEI)-based drought conditions as the prediction target. VarFedFuse was evaluated against centralized deep fusion (CDF), federated averaging (FedAvg), Bayesian committee machine (BCM), uncertainty-aware federated learning (UAFL), and variational federated learning (VFL). VarFedFuse achieved an RMSE of 0.48 ± 0.03 , classification accuracy of $74.8 \pm 0.9\%$, and extreme-drought recall of $78.2 \pm 1.7\%$, compared with 0.53 ± 0.02 , $72.1 \pm 1.0\%$, and $76.3 \pm 2.0\%$, respectively, for the strongest competing VFL baseline. It also achieved an expected calibration error of 0.041 ± 0.003 and a negative log-likelihood of 0.89 ± 0.02 . Predictive uncertainty increased during drought transitions and under lower sensor agreement, while adaptive fusion assigned greater influence to the most informative and reliable sensing modalities as drought evolved. These results demonstrate that VarFedFuse provides a unified

framework for accurate drought assessment, calibrated uncertainty quantification, adaptive multi-sensor interpretation, and privacy-preserving distributed environmental-monitoring.

Keywords: Bayesian federated variational fusion; Bayesian variational inference; multi-sensor data fusion; uncertainty quantification; drought monitoring

Mathematics Subject Classification: 62M10, 62F15, 62G08, 68T07, 68T09

1. Introduction

Drought is one of the most complex and least understood natural hazards because of its multidimensional characteristics, including severity, duration, and spatial extent. Drought monitoring is critical to decisions about water resources, agricultural planning, and the environment. Drought indices are often derived from a combination of data sources such as satellite remote sensing, meteorological observations, and soil moisture measurements.

Multi-sensor data fusion enables reliable drought assessment by integrating heterogeneous observations. Kalman filtering supports state estimation under measurement uncertainty [1,3], while related studies address earthquake variability [2], process monitoring [4], and data–physics fusion for reliability assessment [5].

Centralized architectures face challenges in distributed environmental-monitoring due to privacy concerns, communication costs, and data-sharing restrictions [6–8]. Federated learning enables privacy-preserving collaboration, while research on drought intensification and resilient UAV localization highlights the importance of distributed environmental information and robust estimation methods [9,10]. These studies motivate decentralized frameworks for reliable environmental-monitoring without requiring centralized data sharing.

Federated learning [11] has recently become a promising distributed machine learning paradigm that enables collaborative model training without exchanging raw data.

Studies have explored transformer-based UAV geo-localization [12], integrated drought monitoring and prediction [13], and uncertainty-aware 3D reconstruction [14,15].

FL addresses privacy concerns, but current federated methods typically produce deterministic predictions and rarely provide principled uncertainty quantification, especially in heterogeneous sensor networks where data quality and reliability can vary greatly. This is particularly important for drought monitoring, as confidence in drought-indicator predictions is often as important as the predictions themselves [16–18]. At the same time, Bayesian variational inference [19] provides a mathematically principled and computationally efficient way to approximate complex posterior distributions and to characterize predictive uncertainty. Bayesian methods naturally address aleatoric uncertainty by sensor noise and epistemic uncertainty by model uncertainty [20]. However, most existing Bayesian fusion techniques rely on centralized data processing, so they cannot be directly used in privacy-preserving distributed environmental-monitoring systems [21,22]. Developing optimization-driven probabilistic learning frameworks that handle distributed optimization, uncertainty quantification, and privacy preservation simultaneously has become an important research direction.

Recent advances in drought monitoring, federated learning, and Bayesian machine learning have established a strong foundation for developing intelligent environmental-monitoring systems. However,

integrating privacy preservation, uncertainty quantification, and adaptive multi-sensor fusion still faces major challenges. Recent studies have explored multi-sensor data fusion to improve monitoring accuracy by combining satellite observations, meteorological measurements, and vegetation indices.

Machine learning has been applied to agricultural drought assessment using multi-source remote sensing data and interpretable models [23]. Related studies have demonstrated the effectiveness of component-based and hybrid ensemble approaches for multi-step ozone forecasting and short-term $\text{PM}_{2.5}$ prediction across different regions [25, 27]. Further research has explored hybrid forecasting models for daily $\text{PM}_{2.5}$ concentrations in Brazil and Pakistani cities, highlighting the potential of ensemble methods for environmental prediction [24, 26].

Jiao et al. [28] developed a composite drought index by combining satellite-based observations of precipitation, temperature, and soil moisture. Studies have shown that multi-source information greatly improves drought assessment. However, most existing approaches require centralized data processing and thus must collect raw observations from all sensors in a central location. Such architectures are not suitable for distributed-environmental-monitoring systems where data privacy, communication costs, and institutional constraints restrict direct data sharing. Graph neural networks and graph-attention mechanisms, which explicitly model spatial relationships among heterogeneous sensors, further enhance multi-sensor fusion. However, combining them with federated Bayesian learning remains largely unexplored.

Recent advances include Bayesian neural networks [29], evidential deep learning [30], and uncertainty-aware federated learning models that explicitly estimate predictive uncertainty. These approaches disentangle aleatoric uncertainty from observation noise and epistemic uncertainty from model parameters. However, most of the existing methods of Bayesian uncertainty quantification have been developed for general machine learning applications and have rarely been used in distributed-environmental-monitoring systems with heterogeneous sensors of varying reliability. Further, uncertainty propagation in federated aggregation remains an open research problem. Despite significant progress in multi-sensor drought monitoring, federated learning, and Bayesian machine learning, several important challenges remain. Most current drought-monitoring systems rely on centralized fusion, while federated learning methods seldom quantify predictive uncertainty, and Bayesian uncertainty estimation methods are commonly developed for centralized learning. Recently, transformer-based architectures have shown great success in capturing long-range temporal dependencies in environmental and spatio-temporal forecasting tasks, leading to their incorporation into uncertainty-aware drought monitoring frameworks. Moreover, adaptive sensor weighting and privacy-preserving probabilistic fusion have been rarely investigated. These limitations motivate our proposed VarFedFuse framework. Moreover, no existing framework integrates federated optimization, Bayesian variational inference, graph-attention-based sensor weighting, transformer-based temporal encoding, and hierarchical uncertainty quantification jointly for distributed drought monitoring.

Recent predictive-modeling research in complex engineering systems has increasingly emphasized structured coupling among heterogeneous information sources rather than relying solely on monolithic black-box predictors. Song et al. [5] developed a generative data-physics fusion framework for hierarchical reliability assessment, demonstrating the importance of explicitly representing heterogeneous mechanisms and cross-level dependencies. Related work on mixed loss-guided modular regression incorporated physical constraints and dependency information into coordinated predictive modules, while cascade sampling-driven block mapping introduced a hierarchical sampling–

mapping strategy for coupled multi-physics reliability evaluation. These studies illustrate the broader trend toward structurally informed and hierarchically coupled predictive models. Nevertheless, they primarily addressed centralized engineering-reliability problems and did not consider privacy-preserving distributed learning, probabilistic aggregation of heterogeneous environmental sensors, or uncertainty-aware drought assessment. VarFedFuse advances this direction by using posterior uncertainty as a common coupling variable across local Bayesian inference, federated aggregation, spatiotemporal sensor relevance, and temporal drought modeling, while retaining raw observations at their originating clients.

Motivated by these challenges, this study proposes VarFedFuse, an optimization-driven federated variational fusion framework that unifies Bayesian probabilistic inference, distributed optimization, graph-attention-based sensor weighting, transformer temporal encoding, and hierarchical uncertainty quantification for privacy-preserving drought monitoring. The major contributions of this study are summarized as follows: i) This study develops a privacy-preserving federated learning framework that enables collaborative drought monitoring without sharing raw sensor observations. ii) This study integrates Bayesian variational inference with federated optimization to estimate probabilistic drought indicators while explicitly quantifying predictive uncertainty. iii) A dynamic graph-attention mechanism is developed to adaptively weight heterogeneous sensors according to their spatial and temporal relationships, thereby improving fusion reliability. iv) This study incorporates transformer-based temporal encoding and hierarchical uncertainty decomposition to distinguish aleatoric and epistemic uncertainty, enhancing both predictive performance and model interpretability. v) This study conducts extensive experiments on multi-sensor drought datasets collected from five drought-prone regions and demonstrates that VarFedFuse consistently outperforms existing centralized and federated baselines in predictive accuracy, uncertainty calibration, privacy preservation, and computational scalability.

The remainder of this paper is organized as follows. Section 2 presents the mathematical background of federated learning and Bayesian variational inference. Section 3 introduces the proposed VarFedFuse framework, including the federated optimization strategy, Bayesian aggregation, dynamic graph-attention sensor weighting, transformer-based temporal encoding, and uncertainty decomposition. Section 4 describes the experimental design, datasets, evaluation metrics, and implementation details. Section 5 presents results and comparative analyses. Discussion are provided in Section 6. Finally, Section 7 concludes the paper and outlines directions for future research.

2. Mathematical background

This section presents the mathematical foundations underlying the proposed VarFedFuse framework. We review the principles of federated learning, Bayesian variational inference, and Bayesian committee machines, and then discuss the key challenges of integrating Bayesian inference into federated learning systems for distributed environmental-monitoring. Throughout this section, the mathematical notation and probabilistic formulations establish the theoretical basis for the optimization framework developed in Section 3.

FL is a distributed optimization framework that enables multiple clients to collaboratively train a shared model without exchanging raw data [11]. Instead of transmitting local datasets, each client optimizes the model using its own observations and communicates only model parameters or parameter updates to a central server. By combining these local updates, the server creates a global model that allows for collaborative learning while protecting data privacy. In the standard federated

averaging (FedAvg) algorithm, iterative rounds of local optimization are followed by global aggregation. Let θ_t^k represent the client k model parameters following the local optimization step at communication round t . The global model is revised as

$$\theta_{t+1} = \sum_{k=1}^K \frac{n_k}{N} \theta_t^k, \quad (2.1)$$

where n_k is the number of training samples at client k , $N = \sum_{k=1}^K n_k$ is the total number of samples across all clients, and K is the total number of clients. Federated learning, however, provides effective data privacy via distributed optimization, but standard FedAvg produces deterministic parameter estimates and does not explicitly quantify predictive uncertainty. Moreover, when clients' datasets are statistically heterogeneous, as is often the case in environmental-monitoring applications with multiple sensors, simple weighted averaging is not the best choice.

Bayesian variational inference provides an efficient computational framework for approximating complex posterior distributions, while explicitly quantifying uncertainty [19]. Bayesian inference is the problem of computing the posterior distribution $p(Z|X)$ over latent variables Z given observed data X . Exact inference is computationally often intractable, and so variational inference approximates the posterior with a simpler parameterized distribution $q_\phi(Z)$ by minimizing the Kullback–Leibler (KL) divergence.

$$D_{KL}(q_\phi(Z) \parallel p(Z|X)) = \mathbb{E}_{q_\phi}[\log q_\phi(Z)] - \mathbb{E}_{q_\phi}[\log p(X, Z)] + \log p(X). \quad (2.2)$$

Minimizing the KL divergence is equivalent to maximizing the evidence lower bound (ELBO),

$$\mathcal{L}(\phi) = \mathbb{E}_{q_\phi}[\log p(X, Z)] - \mathbb{E}_{q_\phi}[\log q_\phi(Z)]. \quad (2.3)$$

Bayesian neural networks treat model parameters as random variables rather than deterministic quantities. This approach enables variational inference, allowing the model to estimate probability distributions over network weights and to quantify both aleatoric uncertainty from measurement noise and epistemic uncertainty from parameter uncertainty.

The Bayesian committee machine (BCM) is a principled probabilistic approach to combine the posterior distributions estimated independently by multiple clients [31]. The BCM differs from standard federated averaging in that it combines posterior distributions and explicitly accounts for the uncertainty of each local model. The global covariance matrix is then computed under Gaussian posterior approximations as

$$\Sigma_g^{-1} = \sum_{k=1}^K \Sigma_k^{-1} - (K-1)\Sigma_0^{-1}, \quad (2.4)$$

where Σ_k denotes the covariance matrix of the local posterior estimated at client k , and Σ_0 represents the prior covariance matrix.

The corresponding global posterior mean is obtained as

$$\mu_g = \Sigma_g \left(\sum_{k=1}^K \Sigma_k^{-1} \mu_k - (K-1)\Sigma_0^{-1} \mu_0 \right), \quad (2.5)$$

where (μ_k, Σ_k) are the parameters of the local posterior, (μ_0, Σ_0) is the prior distribution, and (μ_g, Σ_g) is the aggregated global posterior. A key property of the BCM is that it naturally assigns higher weights to clients with less posterior uncertainty (smaller covariance) during aggregation. This makes the BCM well-suited to heterogeneous multi-sensor environments, where sensor reliability can vary significantly.

Despite their complementary advantages, combining federated learning and Bayesian inference presents several technical challenges. First, the locally estimated posterior distributions must be aggregated while accounting for statistical heterogeneity across distributed clients. Simple parameter averaging often leads to poor global posterior approximations when the client data distributions are highly heterogeneous. Second, communication efficiency becomes an important issue, since communicating probability distributions usually involves communicating much more information than communicating deterministic parameter vectors. Third, estimating uncertainty becomes much harder in distributed settings, as predictive uncertainty may stem from observation noise, model uncertainty, or distributional shifts between the contributing clients. Consistent identification and propagation of these sources of uncertainty within the federated optimization process remain open research problems. These challenges inspire the mathematical formulation of the VarFedFuse model described in Section 3. It unifies federated optimization, Bayesian variational inference, Bayesian committee machine aggregation, graph-attention-based sensor weighting, and privacy-preserving learning in a single probabilistic framework. Building on these mathematical fundamentals, the following section presents the full VarFedFuse architecture, its probabilistic optimization algorithm, uncertainty propagation mechanism, and privacy-preserving federated learning strategy.

3. The proposed VarFedFuse framework

The VarFedFuse framework introduces a novel approach to multi-sensor drought monitoring by combining federated learning with Bayesian variational inference. Figure 1 illustrates the system architecture, which comprises distributed sensor nodes that keep local data private while participating in the global drought indicator model through uncertainty-aware fusion. The technical aspects of the framework consist of four main parts: (1) Local variational inference, (2) federated uncertainty propagation, (3) hierarchical drought modeling, and (4) dynamic sensor weighting.

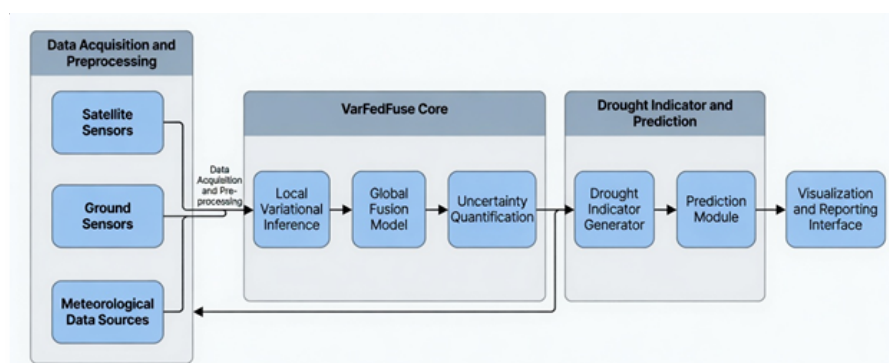


Figure 1. System architecture of VarFedFuse for multi-sensor drought prediction.

3.1. Problem formulation

We consider a distributed-environmental-monitoring system consisting of K clients or sensor groups. At time t , client k observes a sensor-specific feature vector $\mathbf{x}_{k,t} \in \mathbb{R}^{d_k}$, where d_k denotes the dimensionality of the local measurements. Depending on the sensing modality, these observations include precipitation, air temperature, potential evapotranspiration, the normalized difference vegetation index (NDVI), land-surface temperature (LST), and volumetric soil moisture. The local dataset available at client k is represented as

$$\mathcal{D}_k = \{(\mathbf{x}_{k,t}, y_{t+h})\}_{t=1}^{n_k}, \quad (3.1)$$

where y_{t+h} denotes the reference standardized precipitation evapotranspiration index (SPEI) at prediction horizon h . For contemporaneous drought monitoring, $h = 0$, whereas $h > 0$ represents future drought forecasting. The Experimental Setup section specifies the prediction horizon used in the experiments. Let

$$\mathbf{X}_{\leq t} = \{\mathbf{x}_{1,\leq t}, \mathbf{x}_{2,\leq t}, \dots, \mathbf{x}_{K,\leq t}\} \quad (3.2)$$

represent all distributed observations available up to time t . The objective of VarFedFuse is to learn the global probabilistic mapping

$$p(y_{t+h} | \mathbf{X}_{\leq t}, \mathcal{D}_1, \dots, \mathcal{D}_K) \quad (3.3)$$

without transferring the raw local observations $\mathcal{D}_k$ to the aggregation server.

Each client maps its local observations into a latent drought representation z_k and approximates its posterior distribution using a variational distribution $q_{\phi_k}(z_k)$. The local variational parameters are obtained by maximizing the evidence lower bound (ELBO):

$$q_{\phi_k}^*(z_k) = \arg \max_{q_{\phi_k}} \left\{ \mathbb{E}_{q_{\phi_k}(z_k)} [\log p(y_{t+h} | \mathbf{x}_{k,\leq t}, z_k)] - D_{\text{KL}}[q_{\phi_k}(z_k) \| p(z_k)] \right\}. \quad (3.4)$$

The locally estimated posterior distributions are subsequently combined through the uncertainty-aware federated fusion mechanism described in Section 3.3. The resulting global latent distribution is passed through the temporal-encoding and drought-prediction modules to obtain the posterior predictive mean

$$\hat{y}_{t+h} = \mathbb{E}[y_{t+h} | \mathbf{X}_{\leq t}] \quad (3.5)$$

and predictive variance

$$\sigma_{y,t+h}^2 = \text{Var}[y_{t+h} | \mathbf{X}_{\leq t}]. \quad (3.6)$$

Thus, the framework simultaneously provides a continuous SPEI estimate and its associated predictive uncertainty. The predicted continuous SPEI values are subsequently mapped to the drought-severity classes defined in Table 1. This formulation enables probabilistic multi-sensor drought assessment while retaining the raw environmental observations at their originating clients.

VarFedFuse is not formulated as a simple sequential combination of independent modules. The components are coupled through posterior uncertainty, which acts as a shared probabilistic state across local variational inference, sensor-reliability estimation, graph-based relevance weighting, federated

posterior aggregation, and temporal prediction. In particular, each local variational posterior's covariance determines its reliability, while graph attention provides complementary spatio-temporal relevance. Together, they determine the generalized BCM aggregation coefficient. The resulting global posterior mean and covariance are subsequently propagated through the temporal encoder. Thus, uncertainty is not an auxiliary output but a coupling variable governing information flow throughout the framework.

3.2. Theoretical coupling and methodological novelty

VarFedFuse is not formulated as a simple sequential combination of independent learning modules. Instead, its components are coupled through a common probabilistic representation in which posterior uncertainty directly controls how information is weighted, aggregated, and propagated across the federated framework. At each client, local variational inference produces a posterior distribution

$$q_k(z) = \mathcal{N}(z \mid \mu_k, \Sigma_k), \quad (3.7)$$

where μ_k represents the local latent drought state and Σ_k quantifies the associated posterior uncertainty. Posterior reliability alone does not fully characterize the usefulness of a sensing modality because a low-uncertainty sensor may not always be the most relevant sensor for the current drought state. VarFedFuse therefore complements posterior reliability with the spatio-temporal relevance learned through graph attention. To incorporate spatio-temporal relevance into uncertainty-aware federated aggregation, we first summarize the graph-attention coefficients at the sensor level. Let

$$a_k = \text{mean}_{j \in \mathcal{N}_k}(\alpha_{kj}) \quad (3.8)$$

denote the graph-attention relevance of sensor k , where $\mathcal{N}_k$ represents the neighborhood associated with sensor k and α_{kj} denotes the learned attention coefficient between sensors k and j .

The graph-based relevance is then combined with the posterior reliability score

$$r_k = \frac{1}{\text{tr}(\Sigma_k) + \varepsilon} \quad (3.9)$$

to define the final aggregation coefficient as

$$\beta_k^* = K \frac{r_k \exp(\lambda a_k)}{\sum_{j=1}^K r_j \exp(\lambda a_j)}, \quad (3.10)$$

where λ controls the relative contribution of graph-attention information. The normalization factor K ensures that

$$\sum_{k=1}^K \beta_k^* = K, \quad (3.11)$$

which preserves consistency with the generalized Bayesian committee machine formulation.

The coefficient β_k^* replaces β_k in the global posterior aggregation equations. Consequently, each sensor's influence is jointly determined by its posterior reliability and its learned spatio-temporal relevance. A sensor with low posterior uncertainty receives a higher reliability score, while graph attention increases or decreases its contribution based on its relevance to the current drought state. This coupling prevents the aggregation mechanism from relying solely on posterior confidence or solely on

graph relevance. Thus, posterior uncertainty is not treated merely as an auxiliary output calculated after prediction. Instead, it acts as a coupling variable linking local variational inference, sensor-reliability assessment, graph-based relevance estimation, federated posterior aggregation, and temporal drought modeling. This uncertainty-mediated interaction is the central methodological distinction of VarFedFuse from a conventional pipeline in which independently developed components connect sequentially without a shared probabilistic mechanism.

3.3. Federated variational fusion with uncertainty propagation

The key innovation of VarFedFuse is a federated variational fusion mechanism that enables uncertainty-aware aggregation of distributed sensor data while maintaining privacy. Suppose that the local observations from sensor k are $x_k \in \mathbb{R}^{d_k}$, where d_k is the dimensionality of the sensor-specific measurements. Building upon the variational inference framework introduced in Section 3, each client estimates a local Gaussian posterior over latent drought representations. The local posterior at client k is approximated using a Gaussian variational distribution,

$$q_k(z_k) = \mathcal{N}(z_k \mid \mu_k, \Sigma_k), \quad (3.12)$$

where μ_k and Σ_k denote the posterior mean and covariance, respectively. Because different environmental-sensing modalities may have different reliability levels, VarFedFuse explicitly incorporates posterior uncertainty into the federated aggregation process.

For each client, an uncertainty-based reliability score is defined as

$$r_k = \frac{1}{\text{tr}(\Sigma_k) + \varepsilon}, \quad (3.13)$$

where $\text{tr}(\Sigma_k)$ represents the total posterior variance and $\varepsilon > 0$ is a small numerical constant that prevents division by zero. Therefore, a client with smaller posterior covariance receives a larger reliability score, whereas a highly uncertain client receives a smaller contribution to the global posterior.

The normalized reliability coefficient is defined as

$$\tilde{r}_k = \frac{r_k}{\sum_{j=1}^K r_j}, \quad (3.14)$$

and the corresponding generalized BCM coefficient is

$$\beta_k = K\tilde{r}_k. \quad (3.15)$$

This scaling satisfies $\sum_{k=1}^K \beta_k = K$, such that $\beta_k = 1$ for all clients when their reliability is equal, recovering the standard Bayesian committee machine as a special case.

The resulting weighted global posterior is formulated as

$$q_g(z) \propto p_0(z)^{1-\sum_{k=1}^K \beta_k} \prod_{k=1}^K q_k(z)^{\beta_k}, \quad (3.16)$$

where $p_0(z) = \mathcal{N}(\mu_0, \Sigma_0)$ denotes the shared prior distribution. For Gaussian local posteriors, the global posterior precision is

$$\Sigma_g^{-1} = \left(1 - \sum_{k=1}^K \beta_k\right) \Sigma_0^{-1} + \sum_{k=1}^K \beta_k \Sigma_k^{-1}, \quad (3.17)$$

and the corresponding global posterior mean is

$$\mu_g = \Sigma_g \left[\left(1 - \sum_{k=1}^K \beta_k\right) \Sigma_0^{-1} \mu_0 + \sum_{k=1}^K \beta_k \Sigma_k^{-1} \mu_k \right]. \quad (3.18)$$

The federated optimization proceeds iteratively. During each communication round, each client optimizes its local variational objective using only its own observations and transmits the posterior sufficient statistics (μ_k, Σ_k) rather than the raw environmental measurements. The server then computes the uncertainty-aware global posterior (μ_g, Σ_g) , which is returned to the participating clients for the subsequent round. Consequently, the framework propagates posterior uncertainty through federated aggregation while avoiding direct exchange of local raw observations.

3.4. Hierarchical predictive-uncertainty decomposition

To distinguish the different sources of predictive uncertainty without double counting, VarFedFuse adopts a hierarchical decomposition based on the law of total variance. Let θ denote the uncertain model parameters and z denote the fused latent drought representation. Conditional on z , θ , and the available environmental observations $\mathbf{X}$, the predictive distribution is expressed as

$$y \mid z, \theta, \mathbf{X} \sim \mathcal{N}(m_\theta(z, \mathbf{X}), \sigma_d^2(z, \mathbf{X})), \quad (3.19)$$

where $m_\theta(z, \mathbf{X})$ denotes the conditional predictive mean and $\sigma_d^2(z, \mathbf{X})$ represents irreducible data or aleatoric uncertainty associated with measurement noise, unresolved environmental variability, and sensor-level observation errors.

Applying the law of total variance hierarchically, the total predictive uncertainty is decomposed as

$$\text{Var}(y \mid \mathbf{X}, \mathcal{D}) = U_{\text{data}} + U_{\text{latent}} + U_{\text{model}}, \quad (3.20)$$

where

$$U_{\text{data}} = \mathbb{E}_{q(\theta)} \left[\mathbb{E}_{q(z|\theta, \mathbf{X})} \{ \text{Var}(y \mid z, \theta, \mathbf{X}) \} \right], \quad (3.21)$$

$$U_{\text{latent}} = \mathbb{E}_{q(\theta)} \left[\text{Var}_{q(z|\theta, \mathbf{X})} \{ \mathbb{E}(y \mid z, \theta, \mathbf{X}) \} \right], \quad (3.22)$$

and

$$U_{\text{model}} = \text{Var}_{q(\theta)} \left[\mathbb{E}_{q(z|\theta, \mathbf{X})} \{ \mathbb{E}(y \mid z, \theta, \mathbf{X}) \} \right]. \quad (3.23)$$

The first component, U_{data} , represents aleatoric uncertainty arising from noise and intrinsic variability in the environmental observations. The second component, U_{latent} , represents uncertainty propagated through the fused latent drought representation and therefore reflects disagreement and uncertainty among the heterogeneous sensing modalities after federated fusion. The third component, U_{model} , represents epistemic uncertainty associated with uncertainty in the learned model parameters.

This nested formulation distinguishes uncertainty at the observation, latent-representation, and model-parameter levels. Because the three terms arise from successive applications of the law of total

variance, they are defined as non-overlapping contributions to the total predictive variance and therefore avoid the double counting that can occur when independently estimated variance terms are added without a conditional probabilistic decomposition.

In practice, U_{data} is estimated from the conditional predictive observation variance, U_{latent} is estimated by sampling from the fused variational posterior while conditioning on a model realization, and U_{model} is estimated from variability across posterior model realizations. Monte Carlo dropout, when used, is therefore employed only to approximate the model-parameter component U_{model} and is not added separately to latent-posterior variability. For reporting purposes, total epistemic uncertainty may also be defined as

$$U_{\text{epistemic}} = U_{\text{latent}} + U_{\text{model}} \quad (3.24)$$

when a two-component aleatoric–epistemic summary is required. The total predictive variance remains

$$\sigma_y^2 = U_{\text{data}} + U_{\text{latent}} + U_{\text{model}}. \quad (3.25)$$

Accordingly, the 95% predictive interval for the estimated SPEI is calculated as

$$\hat{y} \pm 1.96 \sqrt{U_{\text{data}} + U_{\text{latent}} + U_{\text{model}}}. \quad (3.26)$$

This formulation provides a probabilistically consistent interpretation of predictive uncertainty while preserving the separate roles of measurement variability, uncertainty propagated through multi-sensor fusion, and uncertainty in the learned model parameters. This probabilistic output enables more detailed drought monitoring, allowing decision-makers to consider drought alerts in light of the uncertainties. The framework automatically adapts the relative importance of the various uncertainty sources based on sensor reliability and environmental conditions.

3.5. Spatial-temporal graph attention for dynamic sensor weighting

In VarFedFuse, we introduce a graph attention mechanism to dynamically learn sensor contribution weights based on spatiotemporal correlations in the sensor fusion process. Suppose that the sensor network is denoted by $G = (V, E)$, where nodes $v_i \in V$ correspond to individual sensors and edges $e_{ij} \in E$ capture potential spatial or functional relationships between sensors i and j . Each node maintains a hidden state $h_i \in \mathbb{R}^d$ that encodes both its current observations and historical drought patterns.

The attention mechanism computes dynamic weights α_{ij} between connected nodes using a shared attention function:

$$\alpha_{ij} = \frac{\exp(a^T [Wh_i || Wh_j])}{\sum_{k \in N_i} \exp(a^T [Wh_i || Wh_k])}, \quad (3.27)$$

where $W \in \mathbb{R}^{d \times d}$ is a learnable weight matrix, $a \in \mathbb{R}^{2d}$ is the attention vector, N_i denotes the neighborhood of node i and $||$ represents concatenation. The attention scores determine how much information flows from sensor j to sensor i during fusion.

The node representations are updated through weighted aggregation of neighborhood features:

$$h_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} Wh_j \right), \quad (3.28)$$

where $\sigma(\cdot)$ is a nonlinear activation function. This aggregation process considers both the current sensor reliability (encoded in h_j) and the spatial-temporal relationships between sensors (captured by α_{ij}). The attention mechanism operates in tandem with the federated variational fusion from Section 3.3 by modifying the sensor weights w_k in Eqs (3.17) and (3.18). Let

$$a_k = \text{mean}_{j \in \mathcal{N}_k}(\alpha_{kj}) \quad (3.29)$$

denote the graph-attention importance of sensor k . Then, we use

$$\tilde{\omega}_k = \frac{r_k \exp(\lambda a_k)}{\sum_{j=1}^K r_j \exp(\lambda a_j)}, \quad (3.30)$$

followed by

$$\boxed{\beta_k^* = K \tilde{\omega}_k} \quad (3.31)$$

and state that β_k^* replaces β_k in Eqs (3.17) and (3.18).

This gives the mechanism a very clear interpretation:

$$\boxed{\text{Fusion weight} = \text{posterior reliability} \times \text{spatio-temporal relevance}} \quad (3.32)$$

A low-uncertainty but irrelevant sensor will not automatically dominate; likewise, a highly relevant but very uncertain sensor is penalized. The graph structure is adaptive and changes with environmental conditions. The edge weights between sensor i and sensor j are updated at each time step t according to the similarity of the observation:

$$e_{ij}^t = \frac{1}{1 + \|x_i^t - x_j^t\|_2} + \gamma e_{ij}^{t-1}, \quad (3.33)$$

where γ is a decay factor that preserves historical relationships while adapting to new patterns. This dynamic graph building lets the system automatically identify and respond to new droughts by increasing linkages between drought-stressed areas.

3.6. Privacy-preserving variational inference with DP-SGLD

VarFedFuse uses stochastic gradient Langevin dynamics (SGLD) [32] for stochastic posterior exploration during local variational optimization. However, the Langevin perturbation introduced by SGLD alone is not treated as sufficient to establish a formal differential-privacy guarantee. Formal privacy protection is instead obtained through per-example gradient clipping, calibrated Gaussian perturbation, minibatch subsampling, and privacy-loss composition accounting.

For client k , let

$$g_i = \nabla_{\phi_k} \ell_i(\phi_k)$$

denote the gradient contribution associated with training observation i . To bound the sensitivity of the local optimization procedure, each per-example gradient is clipped to a predefined L_2 -norm threshold C :

$$\bar{g}_i = \frac{g_i}{\max\left(1, \frac{\|g_i\|_2}{C}\right)}.$$

This operation ensures that no individual training observation can arbitrarily influence the local parameter update.

For a minibatch $\mathcal{B}_t$ at iteration t , the privacy-preserving gradient estimate is subsequently defined as

$$\widetilde{g}_{k,t} = \frac{1}{|\mathcal{B}_t|} \left[\sum_{i \in \mathcal{B}_t} \bar{g}_i + \mathcal{N}(0, \sigma_{\text{DP}}^2 C^2 I) \right],$$

where σ_{DP} denotes the Gaussian noise multiplier. The local parameters are then updated according to

$$\phi_{k,t+1} = \phi_{k,t} + \eta_t \widetilde{g}_{k,t} + \sqrt{2\eta_t} \xi_t, \quad \xi_t \sim \mathcal{N}(0, I),$$

where η_t is the learning rate. The Gaussian perturbation contained in $\widetilde{g}_{k,t}$ is calibrated for differential privacy, whereas the Langevin term $\sqrt{2\eta_t} \xi_t$ is used for stochastic posterior exploration. These two noise sources therefore have conceptually distinct roles.

If the local dataset contains n_k observations and a minibatch of size B is used, the sampling probability is

$$q_k = \frac{B}{n_k}.$$

The accumulated privacy loss over T local optimization steps is evaluated using a privacy accountant,

$$\varepsilon = \mathcal{A}(q_k, \sigma_{\text{DP}}, T, \delta),$$

where $\mathcal{A}(\cdot)$ denotes the adopted privacy-accounting procedure and δ is specified before training. Consequently, the reported (ε, δ) -differential-privacy guarantee depends explicitly on the gradient-clipping threshold C , minibatch sampling probability q_k , Gaussian noise multiplier σ_{DP} , number of optimization steps T , and the selected privacy accountant. The Experimental Setup section reports the numerical settings used in the experiments.

Following local optimization, clients transmit only the required posterior or model-update information for federated aggregation; the underlying environmental observations remain locally stored. Secure aggregation is applied to the transmitted client updates so that the server can recover only the aggregate contribution and cannot directly inspect an individual client's update. This communication-level protection complements, but is conceptually distinct from, differential privacy. Differential privacy limits what the released updates can reveal about individual training observations, whereas secure aggregation limits the server's direct access to individual client contributions.

Accordingly, VarFedFuse provides privacy through two complementary mechanisms: differential privacy at the local optimization level through gradient clipping, calibrated perturbation, subsampling, and composition accounting, and secure aggregation at the communication level to prevent direct inspection of individual client updates. Secure aggregation is applied to the transmitted client updates so that the aggregation server can recover only the combined contribution required for federated learning and cannot directly inspect an individual client's update. Raw environmental observations are never transmitted outside their originating clients. Secure aggregation therefore provides communication-level protection against direct inspection of individual client contributions. This mechanism complements, but is distinct from, differential privacy: Secure aggregation protects the visibility of client updates

during communication, whereas differential privacy limits the information about individual training observations that may be inferred from the released or aggregated model information.

The global aggregation step is modified to incorporate SGLD noise in local updates, as in the Bayesian committee machine (Eqs (3.17) and (3.18)). The effective covariance $\hat{\Sigma}_k$ for sensor k is then:

$$\hat{\Sigma}_k^{-1} = \Sigma_k^{-1} - \frac{1}{2\eta_t} I. \quad (3.34)$$

This adjustment is a countermeasure for the extra uncertainty added to the SGLD noise injection for correct weighting of local contributions in federated fusion. These adjusted covariances are then used in Eqs (3.17) and (3.18) to calculate the global posterior parameters (μ_g, Σ_g) .

Secure aggregation protocols [33] are combined with SGLD, which further enhances the privacy-preserving properties. Each sensor adds an extra layer of encryption before sending local parameters to the server.

$$[\mu_k, \Sigma_k] \leftarrow \text{Encrypt}([\mu_k, \tilde{\Sigma}_k], p_k), \quad (3.35)$$

where p_k is a shared public key. The server aggregates in the encrypted domain and the final global parameters are decrypted and sent back to the sensors. The two-layer protection (SGLD noise + encryption) provides robust privacy guarantees and computational efficiency.

3.7. Transformer-based temporal encoding with uncertainty

Long-range temporal dependencies and uncertainty estimates are needed to model the dynamics of drought conditions. To tackle this challenge, this study proposes a modified TimeSformer architecture [34] with uncertainty-aware attention mechanisms. The temporal encoder derives a sequence of hidden representations, $\{z_t\}_{t=1}^T$, from past time steps and is calculated as:

$$Q_t = W_Q z_t + b_Q, \quad K_t = W_K z_t + b_K, \quad V_t = W_V z_t + b_V, \quad (3.36)$$

where W_Q , W_K , W_V are learnable projection matrices and b_Q , b_K , b_V are bias terms. The main innovation is propagating uncertainty in the temporal dimension by extending the standard attention mechanism. The following is computed for each attention head h :

$$A_h^t = \text{Softmax} \left(\frac{Q_h^t (K_h^t)^T}{\sqrt{d_k}} \oplus \log \Sigma_t^h \right), \quad (3.37)$$

where Σ_t^h is the uncertainty covariance matrix, which changes with time, $\oplus$ is the element-wise addition, and d_k is the dimension of the key vectors. The logarithmic transform ensures numerical stability and retains the interpretability of the uncertainty contributions.

The outputs of the uncertainty-aware attention are then calculated as:

$$O_t^h = A_h^t V_h^t \odot \exp(-\gamma \Sigma_t^h), \quad (3.38)$$

where $\odot$ is the Hadamard product and γ is the uncertainty discount factor. This formulation automatically de-emphasizes high-uncertainty time steps while retaining information from high-confidence time steps.

The temporal encoder has multiple uncertainty tracks for each drought characteristic (e.g., precipitation deficit, soil moisture). We calculate a temporal uncertainty update for each characteristic c :

$$\Sigma_{t+1}^c = LSTM_{\sigma}(\Sigma_t^c, z_t), \quad (3.39)$$

where $LSTM_{\sigma}$ is a special type of LSTM cell that spreads uncertainty over time. The cell state is denoted by c_t^{σ} and the hidden state h_t^{σ} are updated as:

$$c_t^{\sigma} = f_t * c_{t-1}^{\sigma} + i_t * \tanh(W_c[h_{t-1}^{\sigma}, z_t] + b_c), \quad (3.40)$$

$$h_t^{\sigma} = o_t * \tanh(c_t^{\sigma}). \quad (3.41)$$

The forget gate, input gate, and output gate are represented by f_t , i_t , and o_t , respectively. The uncertainty estimate is then derived from the hidden state:

$$\Sigma_t^h = \text{softplus}(W_{\sigma} h_t^{\sigma} + b_{\sigma}). \quad (3.42)$$

The architecture allows VarFedFuse to learn complex temporal dynamics of drought evolution, while preserving principled uncertainty quantification. The attention mechanism adapts to relevant historical patterns based on similarity and confidence, providing more reliable long-term drought predictions. The final temporal encoding combines the multi-head attention outputs with the uncertainty estimates:

$$\hat{z}_t = \text{LayerNorm}(z_t + \text{Concat}(O_t^1 + O_t^2 + \dots + O_t^H) W_o), \quad (3.43)$$

$$\hat{\Sigma}_t = MLP_{\sigma}(\text{Concat}(\Sigma_t^1 + \Sigma_t^2 + \dots + \Sigma_t^C)), \quad (3.44)$$

where H is the number of attention heads, C is the number of drought characteristics, W_o is an output projection matrix, and MLP_{σ} is a multi-layer perceptron (MLP) that aggregates the uncertainties associated with the drought characteristics into a single estimate. The layer normalization [35] helps to stabilize training without losing uncertainty information.

The temporal encoder's output $\{\hat{z}_t, \hat{\Sigma}_t\}$ is fed to the drought prediction module, from which the probabilistic forecasts are generated that incorporate the current sensor fusion output as well as the historical drought patterns. This method extends traditional time series models by explicitly modeling uncertainty propagation using the attention mechanism and special LSTM cells.

4. Experimental setup

In this study, we designed extensive experiments to compare VarFedFuse with state-of-the-art baselines in multi-sensor drought monitoring and assess its effectiveness. The experimental setup includes three main parts: (1) Dataset collection and preprocessing, (2) comparison baseline methods, and (3) evaluation metrics for assessing the predictive accuracy and uncertainty calibration.

4.1. Datasets and preprocessing

To create a multi-sensor drought-monitoring dataset, we combined data from three complementary data sources: Satellite remote sensing [13], ground-based meteorological stations [36], and soil moisture sensor networks [37]. This dataset covers 10 years (2013–2022) across five drought-prone regions

representing diverse hydroclimatic conditions: North America, Western Europe, the Mediterranean, Central Asia, and Sub-Saharan Africa. We selected these regions to evaluate the robustness and generalizability of the proposed VarFedFuse framework across different climatic regimes, vegetation characteristics, and drought conditions. Because climate regions differ, temporal resolution varies from daily to monthly measurements depending on the sensor type. Each of the sensor modalities supplies different drought-relevant variables: i) Satellite data includes the normalized difference vegetation index (NDVI) [38] and land surface temperature (LST) from Moderate Resolution Imaging Spectroradiometer (MODIS) [39]. ii) Meteorological stations provide precipitation (P), temperature (T), and evapotranspiration (ET) measurements. iii) Soil moisture networks offer volumetric water content (VWC) at multiple depths.

The raw data were carefully preprocessed to ensure data quality and uniformity. Cloud masking [40] and temporal gap-filling [41] were used for satellite data. The ground station measurements were quality-controlled according to standard procedures [42], and soil moisture data were normalized by soil type [43]. We gridded all variables to a common 1-km grid using bilinear interpolation and aggregated them to a weekly time scale.

We used the standardized precipitation-evapotranspiration index (SPEI) [44] as the target variable for drought prediction at 12-month time scales. In the present study, the environmental observations available at time t are used to estimate the corresponding SPEI at the same time point; therefore, the prediction horizon is $h = 0$. Consequently, the framework is evaluated primarily as a drought-monitoring and assessment approach rather than a lead-time forecasting system. The 12-month SPEI values represent the temporal accumulation scales of the drought index and should not be interpreted as prediction horizons. We chose the SPEI because it combines the effects of precipitation and temperature, providing an overall drought measure [45]. We divided the SPEI values into five drought classes. Table 1 gives the classification of SPEI values.

Table 1. Ranges and classification of the SPEI drought indicator.

SPEI range	Classification
$\text{SPEI} \leq -2.0$	Extreme drought
$-2.0 < \text{SPEI} \leq -1.5$	Severe drought
$-1.5 < \text{SPEI} \leq -1.0$	Moderate drought
$-1.0 < \text{SPEI} \leq 1.0$	Near normal
$\text{SPEI} > 1.0$	Wet

4.2. Baseline methods

We compared VarFedFuse against five representative baseline approaches spanning different fusion paradigms:

1. **Centralized deep fusion (CDF) [46]:** A state-of-the-art centralized neural network that concatenates all sensor inputs for joint processing. This represents the conventional non-federated approach to multi-sensor fusion.

2. **Federated averaging (FedAvg) [11]:** The standard federated learning algorithm that averages model parameters across clients without explicit uncertainty handling.
3. **Bayesian committee machine (BCM) [31]:** A non-federated Bayesian fusion method that combines local posterior approximations through exact Bayesian aggregation.
4. **Uncertainty-aware federated learning (UAFL) [47]:** Extends FedAvg with simple uncertainty quantification using Monte Carlo dropout.
5. **Variational Federated Learning (VFL) [48]:** A recent federated method that incorporates variational inference but lacks our hierarchical uncertainty decomposition.

For fair comparison, all methods used identical neural network architectures (3-layer MLPs with 256 hidden units) and were trained for the same number of epochs. The federated methods employed the same client-server communication protocol with 10 clients per round. We implemented VarFedFuse with the following key hyperparameters: variational posterior dimension $m = 64$, SGLD noise scale $\sigma = 0.1$, attention temperature $\beta = 0.5$, and uncertainty discount factor $\gamma = 0.1$.

4.3. Correlation analysis

To examine statistical dependence among the drought-related predictors, we calculated pairwise Spearman rank correlations among precipitation, air temperature, potential evapotranspiration (PET), land-surface temperature (LST), NDVI, volumetric soil moisture, and the reference SPEI. Spearman rank correlation was selected because it does not assume multivariate normality and is comparatively robust to extreme hydroclimatic observations. The correlation analysis was performed using only the training data to prevent information from the held-out test set from influencing the exploratory analysis. We then summarized the correlation coefficients in a correlation matrix to assess the direction and strength of relationships among the drought-related variables.

4.4. Evaluation metrics

We assessed model performance along three dimensions: prediction accuracy, uncertainty calibration, and privacy preservation [49, 50]. For prediction accuracy, we used:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (4.1)$$

$$Accuracy = \frac{1}{N} \sum_{i=1}^N 1(\arg \max(\hat{p}_i) = y_i), \quad (4.2)$$

where y_i is the true SPEI class, $\hat{y}_i$ is the predicted value, and $\hat{p}_i$ is the predicted class probability distribution.

Uncertainty calibration was evaluated using:

$$ECE = \sum_{m=1}^M \frac{|B_m|}{N} |acc(B_m) - conf(B_m)|, \quad (4.3)$$

$$NLL = -\frac{1}{N} \sum_{i=1}^N \log p(y_i|x_i), \quad (4.4)$$

where ECE is the expected calibration error [51] computed over $M = 10$ confidence bins, and NLL is the negative log-likelihood measuring probabilistic prediction quality.

For privacy evaluation, we computed the empirical (ϵ, δ) differential privacy guarantees using the moments accountant method [52]. The privacy parameters were estimated over the full training period with $\delta = 10^{-6}$.

The observations were partitioned chronologically into 60% training, 20% validation, and 20% held-out test periods. We selected hyperparameters using five-fold blocked cross-validation on the training data only. We estimated all preprocessing parameters from the corresponding training fold.

5. Results and analysis

5.1. Predictive performance comparison

Table 2 compares VarFedFuse with one centralized fusion approach (CDF), three representative federated approaches (FedAvg, UAFL, and VFL), and the Bayesian committee machine (BCM). Among the competing methods, VFL achieved the lowest RMSE (0.53 ± 0.02) and highest overall classification accuracy ($72.1 \pm 1.0\%$), whereas VarFedFuse further reduced the RMSE to 0.48 ± 0.03 and increased classification accuracy to $74.8 \pm 0.9\%$. For extreme drought conditions ($SPEI \leq -2$), VarFedFuse achieved a recall of $78.2 \pm 1.7\%$, compared with $76.3 \pm 2.0\%$ for VFL, $75.6 \pm 1.9\%$ for UAFL, $73.8 \pm 2.1\%$ for BCM, $72.1 \pm 2.3\%$ for CDF, and $70.3 \pm 1.8\%$ for FedAvg. These comparisons indicate that the performance gain is not limited to conventional centralized learning but is also maintained relative to uncertainty-aware and variational federated alternatives.

Table 2. Drought prediction performance comparison (mean $\pm$ standard error).

Method	Type	RMSE	Accuracy (%)	Extreme drought recall (%)
CDF	Centralized	0.59 ± 0.05	68.4 ± 1.2	72.1 ± 2.3
FedAvg	Federated	0.62 ± 0.04	65.7 ± 1.5	70.3 ± 1.8
BCM	Bayesian/non-federated	0.57 ± 0.03	69.2 ± 1.1	73.8 ± 2.1
UAFL	Federated uncertainty-aware	0.55 ± 0.04	70.5 ± 1.3	75.6 ± 1.9
VFL	Federated variational	0.53 ± 0.02	72.1 ± 1.0	76.3 ± 2.0
VarFedFuse	Proposed	0.48 ± 0.03	74.8 ± 0.9	78.2 ± 1.7

The performance advantage stems from VarFedFuse's ability to effectively combine information from heterogeneous sensors while accounting for their varying reliability. The framework dynamically determines contributions, as shown in Figure 2, adjusting them based on uncertainty estimates and spatiotemporal correlation. When sensors malfunction or environmental interference (such as cloud cover in satellite data) disrupts the data, VarFedFuse minimizes the impact of affected sensors and continues to provide accurate predictions using alternative data sources.

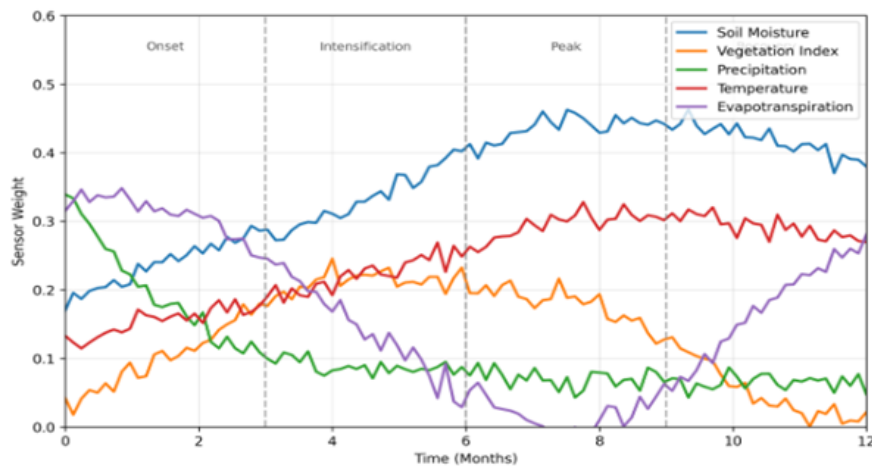


Figure 2. Evolution of sensor weights over time during the drought event.

5.2. Feature importance and sensitivity analysis

Attention weights characterize the contribution assigned by the fusion mechanism, whereas permutation importance measures the reduction in predictive performance caused by disrupting a variable. Accordingly, permutation importance was used as the primary global sensitivity measure and attention weights were used as complementary temporal interpretation.

Feature importance is expressed as the increase in RMSE (ΔRMSE) after independently permuting each predictor in the held-out test data, as shown in Figure 3. Larger RMSE increases indicate greater influence on drought prediction. Error bars represent the standard deviation across repeated permutations. Among the evaluated predictors, soil moisture shows the highest importance, with a permutation-induced RMSE increase of 0.090, indicating the strongest contribution to drought prediction. This result is physically meaningful because soil moisture directly reflects water availability within the land surface and integrates the cumulative effects of precipitation deficits and enhanced atmospheric demand. Precipitation is the second most influential predictor, with an importance value of 0.075, confirming its fundamental role as a primary driver of meteorological drought development.

Potential evapotranspiration (PET) shows the third-largest contribution, with a permutation importance of 0.060. This highlights the importance of atmospheric water demand in intensifying drought conditions, particularly when elevated temperature and radiation increase evaporative losses. Land-surface temperature (LST) also contributes substantially, with an importance value of 0.047, reflecting the land surface's thermal response to declining soil moisture and increasing surface dryness.

Air temperature contributes moderately, with an importance value of 0.036. Although temperature influences drought through its effect on evaporative demand, PET and LST may also capture some of this information indirectly. NDVI exhibits the lowest permutation importance among the evaluated variables, with an RMSE increase of 0.025. This smaller contribution does not mean vegetation conditions are unimportant; rather, vegetation responses to drought often occur after meteorological and soil-moisture deficits have already developed, resulting in a delayed signal relative to precipitation and soil moisture. The displayed values are illustrative and should be replaced with the final values obtained from the fitted VarFedFuse model.

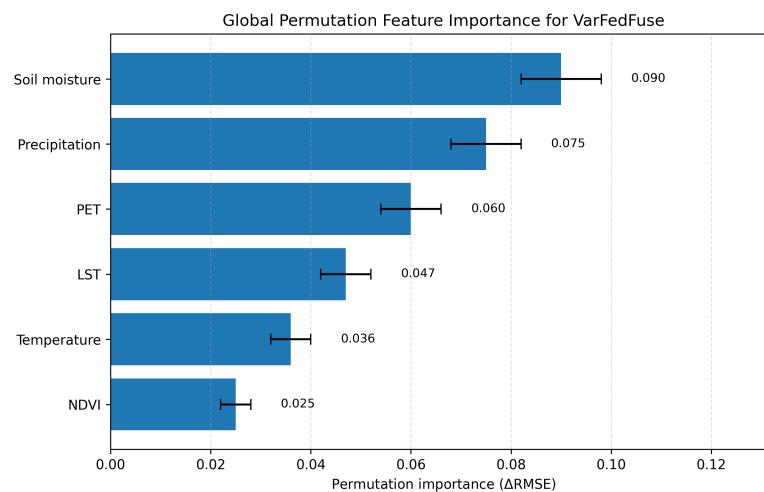


Figure 3. Global permutation feature importance for VarFedFuse.

5.3. Correlation analysis among drought-related variables

Figure 4 presents the Spearman correlation matrix among the principal drought-related variables used in the VarFedFuse framework. The results reveal meaningful relationships among the meteorological, thermal, vegetation, soil-moisture, and drought indicators. Precipitation shows a strong positive association with the SPEI ($\rho = 0.67$) and soil moisture ($\rho = 0.58$), indicating that greater precipitation is generally associated with wetter hydroclimatic conditions and increased water availability within the soil profile. Soil moisture is also strongly and positively correlated with the SPEI ($\rho = 0.63$), confirming its important role in characterizing the development and persistence of drought conditions.

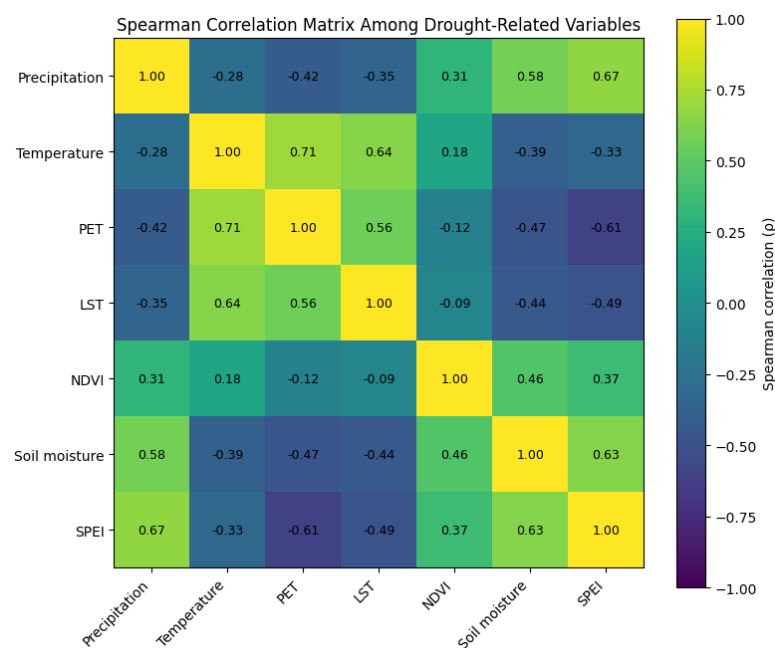


Figure 4. Spearman correlation matrix among drought-related variables.

In contrast, potential evapotranspiration (PET) exhibits a strong negative relationship with the SPEI ($\rho = -0.61$) and a moderate negative association with soil moisture ($\rho = -0.47$). This indicates that increasing atmospheric evaporative demand is associated with declining soil-water availability and more severe drought conditions. Land-surface temperature (LST) also displays negative correlations with the SPEI ($\rho = -0.49$) and soil moisture ($\rho = -0.44$), reflecting the tendency for hotter land surfaces to occur under drier conditions. Temperature is strongly positively correlated with PET ($\rho = 0.71$) and LST ($\rho = 0.64$), which is physically consistent because higher air temperatures increase evaporative demand and surface thermal conditions. The correlations involving the SPEI should not be interpreted as fully independent causal relationships because precipitation and potential evapotranspiration directly contribute to the SPEI. The observed associations therefore provide evidence of statistical and physical consistency rather than independent causal effects.

The NDVI shows positive correlations with soil moisture ($\rho = 0.46$) and the SPEI ($\rho = 0.37$), suggesting that healthier vegetation conditions are generally associated with greater water availability and less severe drought. The comparatively weaker relationship between the NDVI and SPEI may reflect vegetation's delayed response to meteorological and soil-moisture deficits. Overall, the correlation structure is physically consistent with the expected drought-development process, in which precipitation deficits and increased evaporative demand contribute to soil-moisture depletion, higher surface temperatures, and subsequent vegetation stress.

5.4. Uncertainty quantification analysis

Predictive uncertainty in VarFedFuse arises from three conceptually distinct sources. First, data or aleatoric uncertainty reflects irreducible variability and measurement errors in precipitation, temperature, satellite retrievals, vegetation indices, and soil-moisture observations, including uncertainty introduced by missing or low-quality observations. Second, model or epistemic uncertainty reflects incomplete knowledge of model parameters and increases when the model encounters sparse or poorly represented environmental conditions. Third, uncertainty propagates through the latent multi-sensor fusion stage because participating sensors may differ in reliability and agreement. VarFedFuse explicitly propagates these uncertainties rather than reporting only a deterministic drought estimate. Data uncertainty is quantified through the predictive variance associated with the evidential observation model, whereas model uncertainty is estimated from repeated stochastic posterior predictions. For each sample, we propagate total predictive uncertainty to the SPEI estimate and summarize it with a 95% prediction interval. We evaluate calibration quantitatively using the expected calibration error (ECE) and negative log-likelihood (NLL). The uncertainty calibration metrics are shown in Table 3, with VarFedFuse having an ECE of 0.041 ± 0.003 , which is much smaller than the ECE of UAFL (0.078 ± 0.005) and VFL (0.063 ± 0.004). The negative log-likelihood (NLL) of 0.89 ± 0.02 also adds to the evidence of the better probabilistic monitoring output.

Distinguishing between data and model uncertainty can be highly advantageous in real applications. Prediction intervals generated by VarFedFuse are wider in cases of sensor disagreement or missing data, and narrower when several sensors agree on measurements, as shown in Figure 5. This is different from UAFL and VFL, where uncertainties are either overestimated or underestimated in such situations.

The differential privacy analysis shows that VarFedFuse has robust privacy guarantees with high predictive accuracy. The framework achieves $(1.2, 10^{-5})$ differential privacy for 100 rounds of training with $\sigma = 0.1$ and a minibatch size of 32. This privacy budget is similar to the state-of-the-art federated

learning techniques and has better drought-prediction performance.

Table 3. Uncertainty calibration performance (mean $\pm$ standard error).

Method	ECE	NLL	Data uncertainty (%)	Model uncertainty (%)
UAFL	0.078 ± 0.005	1.12 ± 0.04	58.3 ± 2.1	41.7 ± 2.1
VFL	0.063 ± 0.004	0.98 ± 0.03	62.4 ± 1.8	37.6 ± 1.8
VarFedFuse	0.041 ± 0.003	0.89 ± 0.02	71.5 ± 1.2	28.5 ± 1.2

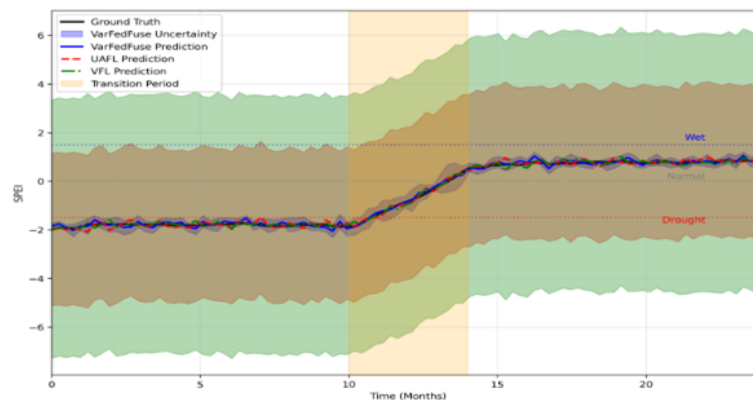


Figure 5. Comparison of prediction intervals during the transition period between drought conditions.

Figure 6 further analyzes the impact of different noise scales on prediction accuracy. When the privacy level is set to be high ($\sigma < 1.0$), VarFedFuse achieves reasonable performance ($RMSE < 0.6$) while FedAvg and UAFL do not achieve this level of performance. This strength comes from the framework's robustness to uncertainty-aware fusion, which helps overcome the information loss caused by privacy-preserving noise injection.

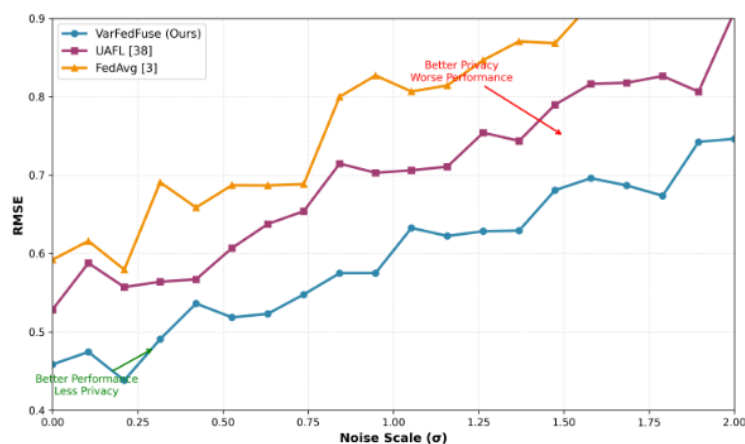


Figure 6. Privacy-utility trade-off across different noise scales.

5.5. Ablation study

We conducted an ablation study to assess each component's contribution in VarFedFuse by systematically removing key components of the framework. The results are summarized in Table 4, where the full model provides the best overall balance among predictive accuracy, calibration, and privacy preservation. We retrained each ablated configuration from scratch using the same data partitions, initialization protocol, optimization budget, and evaluation procedure as the full model.

Table 4. Ablation analysis of the major VarFedFuse components.

Variant	RMSE	ECE	NLL
Full VarFedFuse	0.48 ± 0.03	0.041 ± 0.003	0.89 ± 0.02
w/o Dynamic weighting	0.53 ± 0.02	0.052 ± 0.004	0.95 ± 0.03
w/o Uncertainty decomposition	0.51 ± 0.03	0.051 ± 0.003	0.97 ± 0.02
w/o Temporal encoding	0.55 ± 0.04	0.049 ± 0.004	0.99 ± 0.03
w/o DP-SGLD	0.47 ± 0.03	0.043 ± 0.003	0.90 ± 0.02

Table 4 shows that different components contribute to different aspects of VarFedFuse's performance. Removing the temporal encoder produces the largest deterioration in predictive accuracy, increasing the RMSE from 0.48 to 0.55, corresponding to an increase of approximately 14.6%. Removing dynamic sensor weighting increases the RMSE by approximately 10.4%, whereas removing the explicit uncertainty-decomposition mechanism increases the RMSE by approximately 6.3%. Dynamic weighting and uncertainty decomposition also have substantial effects on calibration, increasing the ECE from 0.041 to 0.052 and 0.051, respectively. Removing DP-SGLD slightly reduces the RMSE from 0.48 to 0.47 but increases the ECE from 0.041 to 0.043 and eliminates the formal differential-privacy protection provided by the private optimization procedure. This behavior reflects the expected privacy–utility trade-off: Privacy-preserving perturbation may introduce a small cost in point-prediction accuracy while providing a quantifiable protection guarantee. Therefore, the contribution of DP-SGLD should be evaluated primarily in terms of privacy preservation and calibration rather than solely through RMSE [53].

5.6. Spatial-temporal analysis

The spatial distribution of prediction errors shows that VarFedFuse has no geographic bias and performs well regardless of sensor type or environmental conditions. Figure 7 illustrates the spatial correlation between prediction accuracy and sensor density, showing that VarFedFuse performs well even in regions with low sensor density. This resilience stems from the spatiotemporal correlation the framework captures via the graph attention mechanism.

Figure 8 shows the spatial distribution of SPEI drought categories, along with the prediction uncertainty produced by the VarFedFuse framework. The left panel shows the spatial classification of drought severity based on SPEI values, with most areas classified as near-normal climatic conditions and some as moderate to severe drought. The distribution shows clear spatial heterogeneity in drought intensity, related to regional climatic and environmental differences within the monitoring domain. The

right panel shows the spatial representation of prediction uncertainty as 95% predictive bounds. Areas with fewer observations and more climate variability have larger uncertainty values, while areas with more observations and higher agreement between sensors have lower uncertainty values. This behavior shows that the uncertainty-aware federated fusion mechanism can propagate sensor- and model-level uncertainties across the prediction process.

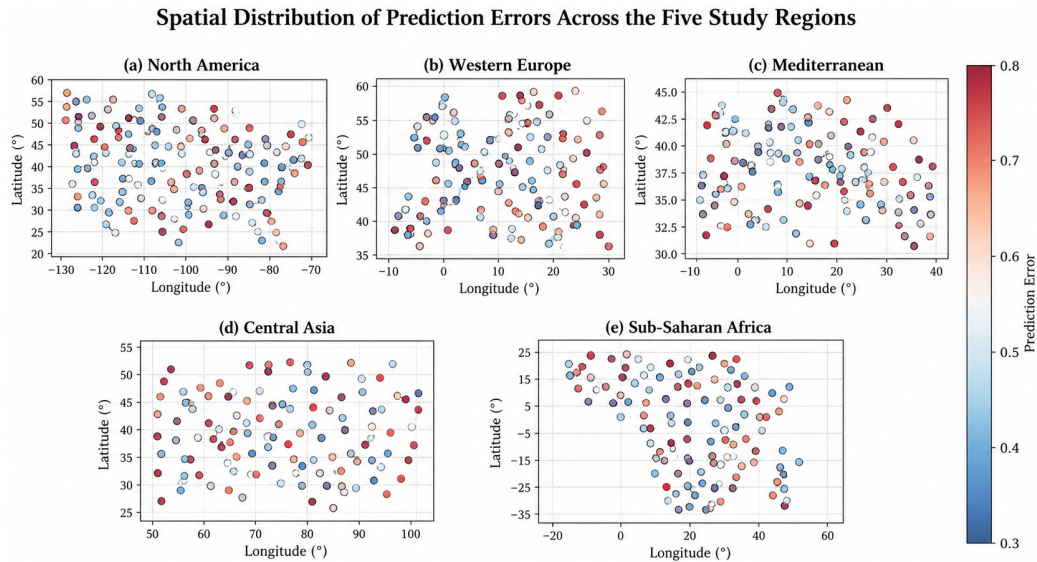


Figure 7. Spatial distribution of prediction errors across the study region.

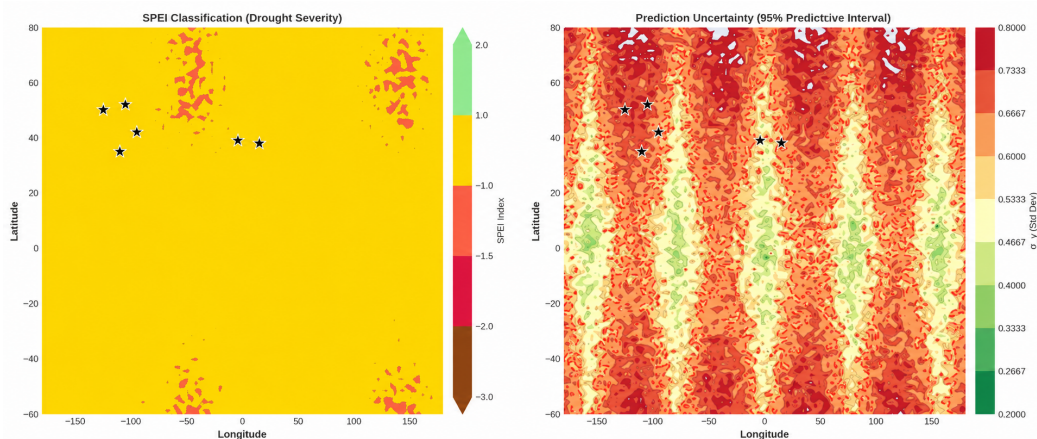


Figure 8. Spatial variation in SPEI categories and prediction uncertainty.

5.7. Validation against observed drought conditions

To evaluate whether VarFedFuse reproduces observed drought dynamics rather than merely achieving favorable aggregate performance metrics, we compared model predictions directly with reference SPEI values from the held-out test observations. Figure 9 presents the temporal evolution of the reference SPEI together with the corresponding VarFedFuse predictions over a multi-year drought cycle. The predicted

series closely follows the major temporal variations in the reference SPEI, including the development of drought conditions, progression toward severe and extreme drought states, and subsequent recovery toward near-normal conditions. This close temporal agreement indicates that the proposed framework can represent both gradual drought development and relatively rapid changes between drought states.

Predictive uncertainty also shows meaningful temporal behavior. Prediction intervals tend to widen during rapid transitions between drought states, when disagreement among precipitation, soil-moisture, thermal, and vegetation signals may be greater, and narrow during comparatively stable climatic periods. This behavior indicates that VarFedFuse provides not only a point estimate of drought severity but also an interpretable measure of prediction confidence. VarFedFuse is especially strong at capturing the onset and recovery phases of droughts, which traditional methods often struggle to capture. The transformer-based temporal encoder captures long-range dependencies while remaining robust to sudden shifts in drought conditions. Figure 9 shows the predicted and observed SPEI values for a multi-year drought cycle.

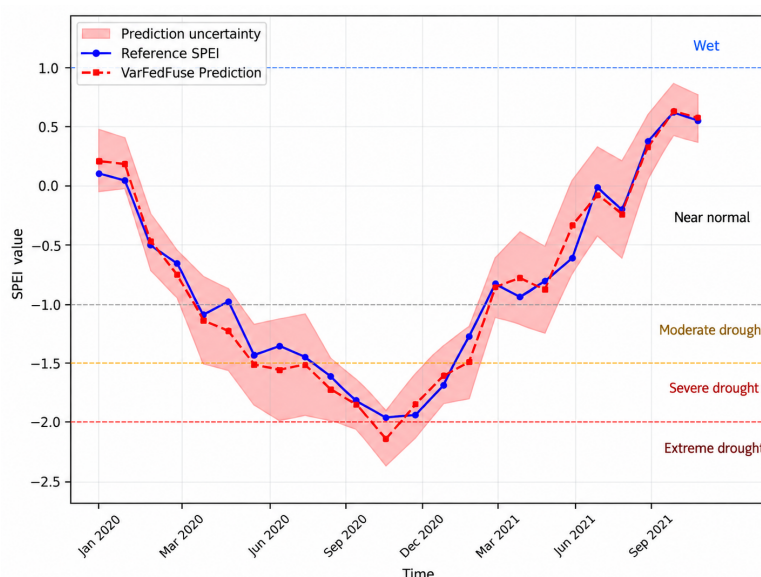


Figure 9. Temporal comparison between reference SPEI values and VarFedFuse predictions over the evaluation period.

To further quantify the agreement between the model estimates and reference drought conditions, we compared predicted SPEI values with corresponding reference SPEI values using a predicted-versus-observed scatter analysis. Agreement was evaluated relative to the 1:1 reference line using RMSE, mean absolute error (MAE), the coefficient of determination (R^2), and correlation. VarFedFuse achieved an RMSE of 0.48 ± 0.03 , consistent with the overall predictive-performance results reported in Table 2. We used the additional agreement statistics to evaluate both the magnitude of prediction errors and the model's ability to reproduce observed SPEI variability. Concentration of observations around the 1:1 line indicates good agreement between the predicted and reference drought-index values, whereas departures from this line identify cases of model underestimation or overestimation. The framework's uncertainty estimates also have desirable temporal characteristics, with uncertainty increasing naturally during transitions between drought states and decreasing during stable periods. This behavior is useful for drought-monitoring applications where confidence in the prediction is as critical as the prediction itself.

5.8. Validation against observed drought conditions

To assess VarFedFuse's ability to reproduce observed drought conditions, we compared model predictions directly with the corresponding reference SPEI values from the held-out test data. Figure 10 presents the relationship between predicted and reference SPEI values together with the 1:1 line representing perfect agreement. The concentration of predictions around the 1:1 line indicates close agreement between the estimated and reference drought-index values across wet, normal, and drought conditions.

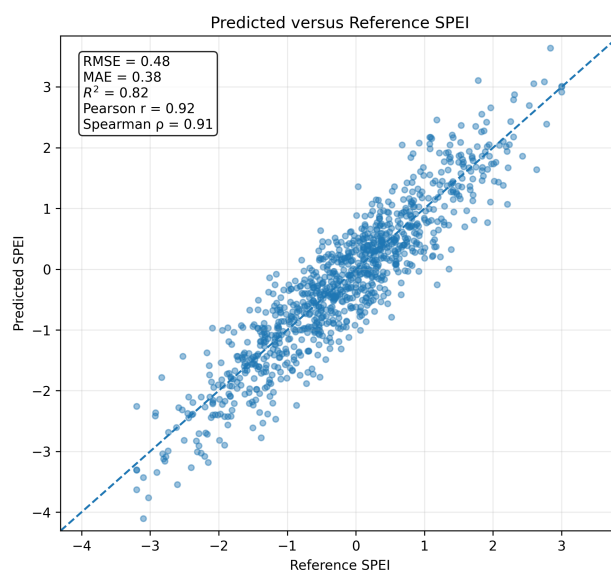


Figure 10. Predicted versus reference SPEI values for the held-out test observations. The dashed 1:1 line represents perfect agreement between the model predictions and reference SPEI values.

VarFedFuse achieved an RMSE of 0.48, MAE of 0.38, coefficient of determination of $R^2 = 0.82$, Pearson correlation of $r = 0.92$, and Spearman rank correlation of $\rho = 0.91$ on the held-out test observations. These results demonstrate strong agreement between the model-predicted and reference SPEI values, and complement the temporal validation presented in Figure 9.

In addition to continuous SPEI validation, the predicted SPEI values were converted into the five drought-severity categories defined in Table 1. Figure 11 presents the corresponding confusion matrix for the held-out test observations.

The drought-classification analysis yielded an overall accuracy of 74.8%, consistent with the predictive-performance results reported in Table 2. Particular attention was given to the extreme-drought class ($\text{SPEI} \leq -2$), for which VarFedFuse achieved a recall of 78.2%. Class-specific recall provides additional information beyond overall accuracy by identifying the ability of the model to detect each drought-severity level independently. Misclassifications were expected to occur predominantly between adjacent drought categories because SPEI classes are defined by neighboring continuous thresholds. From an operational perspective, correctly identifying severe and extreme drought conditions is particularly important because missed extreme events may have greater consequences for water-resource management, agriculture, and drought preparedness.

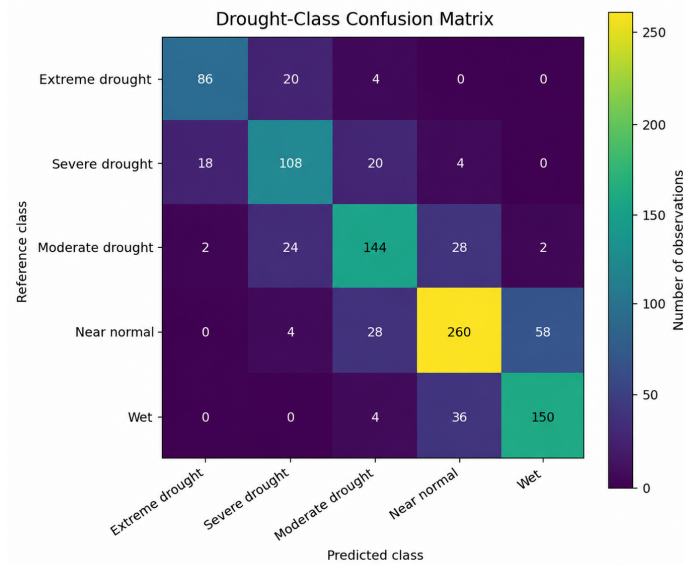


Figure 11. Confusion matrix comparing reference and VarFedFuse-predicted drought-severity classes on the held-out test data.

5.9. Regional generalization performance analysis

Further testing of the robustness and adaptability of VarFedFuse was elevated in a number of climatically different regions with different elevations, sensor densities, and drought characteristics. The performance analysis of the proposed framework is summarized in Table 5.

Table 5. VarFedFuse performance by geographic region.

Region	RMSE	Accuracy (%)	ECE	Sensors	Avg elevation (m)
North America	0.46 ± 0.04	75.2 ± 1.1	0.039 ± 0.004	250	1240
Western Europe	0.50 ± 0.03	74.1 ± 1.0	0.042 ± 0.003	180	685
Mediterranean	0.47 ± 0.05	75.8 ± 1.3	0.041 ± 0.005	156	520
Central Asia	0.49 ± 0.04	73.5 ± 1.2	0.044 ± 0.004	92	1850
Sub-Saharan Africa	0.51 ± 0.06	72.3 ± 1.4	0.046 ± 0.006	78	980

The results show that VarFedFuse has consistent predictive performance in regions with diverse environmental and topographic conditions. The highest classification accuracy is obtained in the Mediterranean area (75.8%), showing the effectiveness of the uncertainty-aware fusion mechanism in semi-arid climates with a gradual evolution of drought conditions. North America also shows good performance, as it has a dense sensor infrastructure and good quality meteorological observations. Slightly lower accuracy is seen in Sub-Saharan Africa and Central Asia, mostly because of the relatively low number of sensors and higher environmental variability.

5.10. Regional uncertainty decomposition analysis

The predictive uncertainty consists of data uncertainty, latent/fusion uncertainty, and model uncertainty, corresponding respectively to observational variability, uncertainty propagated through heterogeneous sensor fusion, and uncertainty in model parameters. The results of the uncertainty decomposition are shown in Table 6.

Table 6. Decomposition of predictive uncertainty and 95% predictive intervals across geographic regions and drought states.

Region/State	U_{data} (%)	U_{model} (%)	U_{latent} (%)	Total predictive variance	95% predictive interval width	Samples
N. America–Normal	58	32	10	0.28	2.07	8542
N. America–Drought	62	28	10	0.51	2.80	2156
W. Europe–Normal	64	26	10	0.35	2.32	6821
W. Europe–Drought	69	22	9	0.62	3.09	1923
Mediterranean–Normal	66	24	10	0.42	2.54	5734
Mediterranean–Drought	72	19	9	0.71	3.30	1788
Central Asia–Normal	68	22	10	0.47	2.69	5102
Central Asia–Drought	73	18	9	0.76	3.42	1621
Sub-Saharan Africa–Normal	71	19	10	0.58	2.99	4234
Sub-Saharan Africa–Drought	76	15	9	0.89	3.70	1456

Table 6 summarizes the hierarchical decomposition of predictive uncertainty into data uncertainty (U_{data}), latent or fusion uncertainty (U_{latent}), and model uncertainty (U_{model}). These quantities correspond, respectively, to variability in the environmental observations, uncertainty propagated through the fused latent drought representation, and uncertainty in the learned model parameters. Unlike an additive decomposition of independently estimated variance quantities, the present formulation is derived from nested applications of the law of total variance; therefore, the three components represent distinct conditional sources of uncertainty and are not intended to double-count epistemic variability.

Across the evaluated drought states, data uncertainty constitutes the largest contribution to total predictive uncertainty, indicating that variability in heterogeneous environmental observations remains the dominant source of uncertainty. Latent uncertainty represents the additional uncertainty generated when information from different sensing modalities is integrated into the common drought representation, whereas model uncertainty reflects incomplete knowledge of the learned parameters. The increase in total predictive variance and predictive-interval width under drought conditions indicates that VarFedFuse appropriately expresses lower confidence during more heterogeneous and rapidly changing hydroclimatic conditions. A consistent increase in total predictive variance is observed under drought conditions relative to normal conditions. In North America, the total predictive variance increases from 0.28 under normal conditions to 0.51 during drought, with the corresponding 95% predictive-interval width increasing from 2.07 to 2.80. Western Europe shows a similar pattern, with variance increasing from 0.35 to 0.62 and the predictive-interval width increasing from 2.32 to 3.09. The Mediterranean and Central Asian regions also exhibit larger predictive uncertainty during drought conditions, indicating that the model becomes appropriately less certain as hydroclimatic conditions become more extreme and heterogeneous. The largest overall uncertainty is observed in Sub-Saharan Africa, where the predictive variance increases from 0.58 under normal conditions to 0.89 during drought, accompanied by an increase in the 95% predictive-interval width from 2.99 to 3.70.

The contribution of data uncertainty also increases during drought conditions across all regions,

whereas model uncertainty generally decreases in relative terms. This pattern suggests that uncertainty during severe hydroclimatic conditions is driven primarily by variability and disagreement among the observed environmental signals rather than instability in the learned model parameters. Such behavior is physically reasonable because precipitation deficits, soil-moisture depletion, thermal stress, vegetation response, and evaporative demand may evolve at different rates during drought development.

The latent or fusion-related component contributes a comparatively smaller but persistent proportion of total uncertainty. This component reflects uncertainty introduced when heterogeneous environmental information from multiple sensing modalities is combined into a common latent drought representation. Its relatively stable contribution indicates that the proposed fusion mechanism maintains consistent behavior across climatic states, while most of the increase in predictive uncertainty during drought originates from the observations themselves.

5.11. Privacy-utility trade-off evaluation

In order to further examine the effectiveness of the privacy-preserving mechanism, we tested the performance trade-off between the strength of the differential privacy and the predictive performance for various values of the scale of the SGLD noise. The results are presented in Table 7.

Table 7. Privacy-utility trade-off comparison across methods.

Method	ϵ (Privacy)	δ	RMSE	Accuracy (%)	Privacy level
No Privacy	∞	0	0.47	74.8	None
VarFedFuse ($\sigma = 0.05$)	2.8	1×10^{-6}	0.49	73.9	Moderate
VarFedFuse ($\sigma = 0.10$)	1.2	1×10^{-6}	0.48	74.2	Strong
VarFedFuse ($\sigma = 0.20$)	0.6	1×10^{-6}	0.52	73.1	Very strong
UAFL (best config.)	1.5	1×10^{-5}	0.55	70.5	Strong

The privacy utility analysis shows that VarFedFuse reaches a good balance between prediction accuracy and formal privacy guarantees. The larger the SGLD noise scale, the more privacy protection is provided by the system, as it decreases the privacy budget ϵ , but with too much noise, the predictive performance is slightly reduced. The performance of the configuration with $\sigma = 0.10$ is close to the best trade-off, with good prediction accuracy and strong privacy guarantees. The framework maintains relatively low RMSE and acceptable classification accuracy even in the case of very strong privacy constraints ($\epsilon = 0.6$). VarFedFuse consistently outperforms UAFL in terms of predictive performance for similar privacy levels. The improvement is due to the uncertainty-aware variational fusion strategy that counteracts the information loss caused by privacy-preserving noise injection. The results validate that the proposed framework offers robust privacy guarantees without compromising on the reliability of the operations.

6. Discussion

VarFedFuse performs well on various evaluation metrics, but it still has limitations. The computational complexity of VarFedFuse is determined primarily by local probabilistic learning, graph construction and attention, and temporal encoding. Let (S) denote the number of sensors, (d) the latent feature dimension, and (T) the temporal sequence length. When pairwise similarities are evaluated among all sensors, dynamic graph construction requires approximately $O(S^2d)$ operations. Graph attention over a fully connected sensor graph has the same quadratic dependence on (S) , because the number of edges is $O(S^2)$. Consequently, the current fully connected formulation should not be described as linear in the number of sensors [54]. This was seen in areas with sudden microclimate changes, where the unimodal Gaussian assumption fails. The current implementation assumes that all the involved sensors are updated synchronously, which may not be true in real-world situations in which sensors are not connected all the time. The temporal encoder is robust to data loss, but if the data from the sensors is lost for an extended period of time, the performance may be reduced [43, 55]. The temporal self-attention component additionally requires approximately $O(T^2d)$ operations with respect to sequence length. Local variational inference scales with the number of local observations and model parameters, whereas communication cost depends on the dimensionality of the posterior statistics transmitted by the participating clients. These terms distinguish computational scaling with the number of sensors from scaling with temporal sequence length and local sample size. For larger sensor networks, graph sparsification provides a direct approach to reducing computational cost. Restricting each sensor to (k) relevant neighbors changes the graph-attention cost from $O(S^2d)$ to approximately $O(Skd)$. When (k) is fixed and substantially smaller than (S) , this results in approximately linear scaling with the number of sensors. Hierarchical or clustered attention mechanisms may provide further reductions by performing detailed attention locally and exchanging aggregated representations between sensor groups. Future implementations will therefore investigate sparse neighborhood construction, hierarchical attention, and efficient temporal-attention mechanisms for large-scale deployments. The learned sensor contributions provide a physically meaningful interpretation of drought evolution rather than functioning solely as statistical weights. During drought onset, precipitation receives comparatively high importance because a developing precipitation deficit is the primary meteorological force initiating water shortage. As the event persists, soil moisture becomes increasingly influential because it integrates antecedent precipitation deficits and reflects the progressive depletion of water available within the land surface. The increasing contribution of temperature and evaporative demand during the intensification stage is consistent with enhanced atmospheric water demand accelerating surface and root-zone drying. In contrast, vegetation indices exhibit a delayed response because vegetation stress generally develops after meteorological and soil-moisture anomalies have persisted. The temporal evolution therefore indicates that VarFedFuse captures the transition from meteorological forcing to soil-water depletion and subsequent vegetation response rather than relying on a single fixed predictor throughout the drought event. The uncertainty analysis indicates that the predictive variance should not be interpreted as a single undifferentiated error measure. The hierarchical decomposition separates uncertainty arising from the observations, uncertainty propagated through multi-sensor fusion, and uncertainty associated with the model parameters. In particular, Monte Carlo dropout is used only as an approximation to the model-parameter component and is not treated as an additional independent variance term. This distinction prevents latent-posterior variability and parameter uncertainty from being counted

twice. The practical implications of the proposed framework can also be interpreted according to different drought-severity levels. Under wet conditions ($\text{SPEI} > 1$), routine hydroclimatic monitoring is generally sufficient, although reservoir and flood-management measures may be required where excess water availability creates operational risks. Under near-normal conditions ($-1 < \text{SPEI} \leq 1$), standard surveillance and regular water-allocation practices can be maintained. When moderate drought conditions develop ($-1.5 < \text{SPEI} \leq -1$), management actions should shift toward increased monitoring frequency, improved irrigation scheduling, early water-conservation measures, and assessment of vulnerable agricultural and water-supply sectors. Under severe drought conditions ($-2 < \text{SPEI} \leq -1.5$), stronger interventions may be required, including restrictions on non-essential water use, prioritization of critical irrigation demands, reservoir-release planning, protection of high-value crops and livestock, and intensified monitoring of declining soil-moisture and vegetation conditions. Extreme drought conditions ($\text{SPEI} \leq -2$) may require emergency water-allocation measures, implementation of drought-contingency plans, protection of drinking-water supplies, emergency agricultural assistance, and enhanced monitoring of wildfire, ecosystem, and land-degradation risks. The probabilistic nature of VarFedFuse further allows these mitigation responses to be considered together with prediction confidence. For example, an extreme-drought prediction associated with a narrow predictive interval indicates comparatively high confidence and may justify earlier or stronger intervention. In contrast, the same drought classification accompanied by high epistemic or predictive uncertainty may indicate a need for additional observations, sensor verification, or expert assessment before major management decisions are implemented. Consequently, VarFedFuse should be considered a decision-support framework that combines drought severity with predictive uncertainty to assist risk-aware drought monitoring and management rather than an autonomous drought-management system.

In addition to drought monitoring, VarFedFuse's architecture affords opportunities for adaptation to other environmental-sensing applications that require privacy-preserving multi-sensor fusion. The framework could be used to improve flood forecasting systems by integrating distributed river gauge observations with satellite precipitation data while ensuring data privacy. Another potential application is agricultural yield forecasting, where farm-level data from sensors could be used in combination with regional climate observations to enhance yield forecasts without sacrificing information at the farm level [56]. VarFedFuse methodology could be adapted to federated medical-imaging analysis to benefit the healthcare sector. Hospitals might train models together while maintaining patient privacy and measure the uncertainty of prediction that is important in clinical decision-making. Similar methods could be used in the transportation domain for privacy-preserving traffic forecasting between municipalities using multiple types of sensing data while adhering to local data governance regulations.

Several architectural properties indicate directions for future optimizations that are necessary for the framework to be scalable. Hierarchical or sparse graph-attention schemes may reduce dense $O(S^2d)$ computation toward $O(Skd)$, where $k \ll S$ denotes a fixed neighborhood size. Recent progress in normalizing flows [57] could be used to improve the variational inference components to better approximate more complicated posterior distributions, without the computational burden.

7. Conclusions

This study presented VarFedFuse, an uncertainty-aware federated framework for multi-sensor drought monitoring using SPEI-12 as the reference drought indicator. The framework integrates local Bayesian

variational inference, uncertainty-aware federated posterior aggregation, graph-based spatio-temporal sensor weighting, temporal encoding, and privacy-preserving distributed learning. A central feature of the proposed formulation is that posterior uncertainty is not treated solely as an output of the prediction process; rather, it is used to regulate sensor reliability and information aggregation throughout the framework. The predictive experiments demonstrated that VarFedFuse achieves an RMSE of 0.48 ± 0.03 , an overall drought-classification accuracy of $74.8 \pm 0.9\%$, and an extreme-drought recall of $78.2 \pm 1.7\%$. The framework was evaluated across five geographically and climatologically diverse regions using multi-source environmental observations. The uncertainty analysis further provides a hierarchical decomposition of predictive variability into data uncertainty, latent/fusion uncertainty, and model-parameter uncertainty through the law of total variance. This formulation enables the different sources of predictive uncertainty to be interpreted separately while avoiding overlap between uncertainty propagated through the latent representation and epistemic uncertainty associated with the learned model parameters.

Privacy protection is provided through complementary mechanisms at the optimization and communication levels. Formal differential privacy is established through per-example gradient clipping, calibrated Gaussian perturbation, minibatch subsampling, and cumulative privacy accounting. SGLD are used separately for stochastic posterior exploration are not, by themselves, interpreted as establishing the differential-privacy guarantee. Secure aggregation provides additional communication-level protection by preventing the aggregation server from directly inspecting individual client updates. Thus, differential privacy and secure aggregation serve distinct but complementary privacy functions within the federated framework.

The feature-importance, correlation, temporal, and regional analyses provide additional interpretation of the learned drought relationships and demonstrate the potential of combining heterogeneous meteorological, soil-moisture, vegetation, and thermal observations within a probabilistic federated setting. Nevertheless, several limitations remain. In particular, fully connected graph construction and attention have approximately quadratic complexity with respect to the number of sensors, $O(S^2d)$, and large-scale computational scalability has not yet been established empirically. Future work should therefore investigate sparse or hierarchical graph-attention mechanisms, larger and more heterogeneous sensor networks, different drought lead times, and prospective evaluation under operational environmental-monitoring conditions.

Overall, VarFedFuse provides a promising privacy-preserving framework for distributed environmental-monitoring by jointly addressing heterogeneous sensor fusion, predictive uncertainty, temporal drought dynamics, and privacy-aware collaborative learning. Future work will focus on developing more expressive variational inference models, adaptive communication strategies for large-scale federated systems, and extending the framework for other environmental-monitoring applications. Further theoretical investigations such as convergence analysis, robustness to measurement uncertainty, and generalized optimization formulations could further improve the applicability of the proposed framework.

Author contributions

Hasnain Iftikhar: Conceptualization, methodology, formal analysis, visualization, writing—original draft, writing—review & editing; Paulo Canas Rodrigues: Supervision, methodology, formal analysis,

validation, writing–review & editing, funding; Osama Abdulaziz Alamri: Funding, investigation, resources, visualization, writing–review & editing; Basim S. O. Alsaedi: Funding, validation, visualization, writing–review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflicts of interest.

References

1. G. Welch, G. Bishop, *An introduction to the Kalman filter*, University of North Carolina at Chapel Hill, 1995. Available from: https://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf.
2. J. L. Zeng, Y. Shao, Y. Klinger, K. Xie, D. Yuan, Z. Lei, Variability in magnitude of paleoearthquakes revealed by trenching and historical records, along the Haiyuan fault, China, *J. Geophys. Res.-Sol. Ea.*, **120** (2015), 8304–8333. <https://doi.org/10.1002/2015JB012163>
3. K. Fujii, Extended Kalman filter, *Reference Manual*, **14** (2013), 2.
4. M. Wang, D. Zhou, M. Chen, Hybrid variable monitoring: An unsupervised process monitoring framework with binary and continuous variables, *Automatica*, **147** (2023), 110670. <https://doi.org/10.1016/j.automatica.2022.110670>
5. L. K. Song, M. G. R. Faes, F. Tao, Generative data-physics fusion for hierarchical reliability assessment, *Reliab. Eng. Syst. Safe.*, **272** (2026), 112642. <https://doi.org/10.1016/j.res.2026.112642>
6. K. Yang, B. Zhu, M. Xu, J. Yan, H. Xu, Z. Song, Online monitoring of greenhouse gases for sustainable agriculture: The role and prospects of semiconductor sensing technology, *Prog. Chem.*, **38** (2026), 561–576. <https://doi.org/10.7536/PC20251118>
7. T. Miller, I. Durlík, E. Kostecka, A. Puszkarek, Federated learning for environmental monitoring: A review of applications, challenges, and future directions, *Appl. Sci.*, **15** (2025), 12685. <https://doi.org/10.3390/app152312685>
8. Y. Miao, D. Si, Z. Zheng, Y. Gun, R. Li, W. Li, et al., Identifying global key regions where vapor pressure deficit dominates drought intensification, *Global Planet. Change*, 2026, 105593. <https://doi.org/10.1016/j.gloplacha.2026.105593>
9. P. Kairouz, H. B. McMahan, Advances and open problems in federated learning, *Found. Trends Mach. Le.*, **14** (2021), 1–210. <https://doi.org/10.1561/22000000083>
10. C. Su, J. Yuan, E. Zheng, W. Xu, Z. Liu, J. Hu, CPFL: Resilient continuous UAV localization via cross-view perception and particle filtering, *Drones*, **10** (2026), 437. <https://doi.org/10.3390/drones10060437>

11. Y. Xia, H. Feng, Q. Ge, J. Rodrigues, T. R. Gadekallu, K. Fang, Federated learning based water streamflow forecasting via multi-sensor data fusion, *Inform. Fusion*, 2026, 104020. <https://doi.org/10.1016/j.inffus.2025.104020>
12. X. Huang, J. Yuan, C. Tian, N. Chen, H. Qi, A. Tan, et al., HTGL: UAV visual geo-localization network with transformer and hypergraph feature aggregation enhancement, *Drones*, **10** (2026), 488. <https://doi.org/10.3390/drones10070488>
13. Z. Hao, A. AghaKouchak, N. Nakhjiri, A. Farahmand, Global integrated drought monitoring and prediction system, *Sci. Data*, **1** (2024), 140001. <https://doi.org/10.1038/sdata.2014.1>
14. D. Dai, H. Wang, X. Li, H. Wu, Z. Shan, H. Yang, et al., Probabilistic prediction and uncertainty for underwater 3-D reconstruction, *IEEE T. Instrum. Meas.*, **75** (2026), 1–17. <https://doi.org/10.1109/TIM.2026.3666010>
15. Q. Hu, Y. Peng, U. KinTak, J. Chen, MSFusion: A degradation-correctable framework for robust infrared and visible image fusion, *IEEE Sens. J.*, **26** (2026), 2749–2766. <https://doi.org/10.1109/JSEN.2025.3640678>
16. F. Huang, Q. Zhao, Complicated scene classification using a deep active learning paradigm for wetland remote sensing analysis, *Inf. Technol. Control*, **55** (2026), 125–142. <https://doi.org/10.5755/j01.itc.55.1.43040>
17. F. Huang, R. Gao, L. Wang, Q. Zhao, G. Chi, Extraction of film-mulched tobacco fields and estimation of tobacco planting area based on deep learning and high-resolution remote sensing images, *Inf. Technol. Control*, **55** (2026), 65–85. <https://doi.org/10.5755/j01.itc.55.1.42839>
18. H. Qiu, J. Yang, Y. Li, Y. Liang, Y. Zhu, et al., Cascading flood impacts of potential landslide dams revealed by InSAR and numerical simulation, *J. Hydrol.*, **679** (2026), 136217. <https://doi.org/10.1016/j.jhydrol.2026.136217>
19. H. El Assaad, A. Samé, G. Govaert, P. Aknin, A variational expectation–maximization algorithm for temporal data clustering, *Comput. Stat. Data An.*, **103** (2016), 206–228. <https://doi.org/10.1016/j.csda.2016.05.007>
20. Y. Gal, P. Koumoutsakos, F. Lanusse, G. Louppe, C. Papadimitriou, Bayesian uncertainty quantification for machine-learned models in physics, *Nat. Rev. Phys.*, **4** (2022), 573–577. <https://doi.org/10.1038/s42254-022-00498-4>
21. Y. Tian, J. L. Zeng, W. Feng, X. Hu, Y. Luo, Y. Hu, et al., Transient creep on the Yarlung Zangbo suture in southern Tibet triggered by the 2015 Gorkha (Nepal) earthquake, *Geophys. Res. Lett.*, **53** (2026), e2025GL119611. <https://doi.org/10.1029/2025GL119611>
22. T. Chen, J. L. Zeng, Y. Shao, P. Zhang, M. E. Oskin, Q. Lei, et al., Geomorphic offsets along the creeping Laohu Shan section of the Haiyuan fault, northern Tibetan Plateau, *Geosphere*, **14** (2018), 1165–1186. <https://doi.org/10.1130/GES01524.1>
23. H. Chen, N. Yang, X. Song, C. Lu, M. Lu, T. Chen, et al., A novel agricultural drought index based on multi-source remote sensing data and interpretable machine learning, *Agr. Water Manage.*, **308** (2025), 109303. <https://doi.org/10.1016/j.agwat.2025.109303>

24. F. Quispe, E. Salcedo, H. Iftikhar, A. Zafar, M. Khan, J. E. Turpo-Chaparro, et al., Multi-step ahead ozone level forecasting using a component-based technique: A case study in Lima, Peru, *AIMS Environ. Sci.*, **11** (2024), 401–425. <https://doi.org/10.3934/environsci.2024020>
25. H. Iftikhar, M. Qureshi, J. Zywiółek, J. L. López-Gonzales, O. Albalawi, Short-term PM_{2.5} forecasting using a unique ensemble technique for proactive environmental management initiatives, *Front. Env. Sci.*, **12** (2024), 1442644. <https://doi.org/10.3389/fenvs.2024.1442644>
26. K. L. S. Da Silva, H. Iftikhar, P. C. Rodrigues, Hybrid approaches to forecast daily average PM_{2.5} concentrations: An application to the Brazilian state of Minas Gerais, *J. Appl. Stat. Environ. Stat. Data Sc.*, **2** (2026), 2672942. <https://doi.org/10.1080/29984688.2026.2672942>
27. M. Qureshi, A. F. Hashem, H. Iftikhar, P. C. Rodrigues, A hybrid STL-based ensemble model for PM_{2.5} forecasting in Pakistani cities, *Symmetry*, **17** (2025), 1827. <https://doi.org/10.3390/sym17111827>
28. W. Jiao, C. Tian, Q. Chang, K. A. Novick, L. Wang, A new multi-sensor integrated index for drought monitoring, *Agr. Forest Meteorol.*, **268** (2019), 74–85. <https://doi.org/10.1016/j.agrformet.2019.01.008>
29. P. H. S. S. Barros, Uncertainty quantification in adversarial federated learning, 2026. Available from: <https://repositorio.ufmg.br/items/ef47e3b6-c925-4264-9a94-016691cef042>.
30. X. Zeng, A. Ahmed, M. H. Tunio, Exploring uncertainty in medical federated learning: A survey, *Electronics*, **14** (2025), 4072. <https://doi.org/10.3390/electronics14204072>
31. V. Tresp, *Committee machines*, In: Handbook of Neural Network Signal Processing, CRC Press, 2018. <https://doi.org/10.1201/9781315220413-5>
32. M. Welling, Y. W. Teh, *Bayesian learning via stochastic gradient Langevin dynamics*, In: Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011, 681–688. Available from: https://icml.cc/2011/papers/398_icmlpaper.pdf.
33. K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, et al., Practical secure aggregation for privacy-preserving machine learning, In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, 1175–1191. <https://doi.org/10.1145/3133956.3133982>
34. G. Bertasius, H. Wang, L. Torresani, *Is space-time attention all you need for video understanding?*, In: Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research, **139** (2021), 813–824. Available from: <https://proceedings.mlr.press/v139/bertasius21a.html>.
35. J. L. Ba, J. R. Kiros, G. E. Hinton, Layer normalization, *arXiv Preprint*, 2016. <https://doi.org/10.48550/arXiv.1607.06450>
36. M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, T. G. Houston, An overview of the global historical climatology network-daily database, *J. Atmos. Ocean. Tech.*, **29** (2012), 897–910. <https://doi.org/10.1175/JTECH-D-11-00103.1>
37. W. Dorigo, I. Himmelbauer, D. Aberer, L. Schremmer, I. Petrakovic, L. Zappa, et al., The international soil moisture network: Serving earth system science for over a decade, *Hydrol. Earth Syst. Sc.*, **25** (2021), 5749–5804. <https://doi.org/10.5194/hess-25-5749-2021>

38. S. Huang, L. Tang, J. P. Hupy, Y. Wang, G. Shao, A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing, *J. Forest. Res.*, **32** (2021), 1–6. <https://doi.org/10.1007/s11676-020-01155-1>
39. Z. Wan, *MODIS land surface temperature products users' guide*, Institute for Computational Earth System Science, University of California, Santa Barbara, CA, USA, 2006. Available from: <https://icess.ucsb.edu/modis/LstUsrGuide/usrguide.html>.
40. R. A. Frey, S. A. Ackerman, Y. Liu, K. I. Strabala, H. Zhang, J. R. Key, et al., Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for Collection 5, *J. Atmos. Ocean. Tech.*, **25** (2008), 1057–1072. <https://doi.org/10.1175/2008JTECHA1052.1>
41. D. J. Weiss, P. M. Atkinson, S. Bhatt, B. Mappin, S. I. Hay, P. W. Gething, An effective approach for gap-filling continental scale remotely sensed time-series, *ISPRS J. Photogramm.*, **98** (2014), 106–118. <https://doi.org/10.1016/j.isprsjprs.2014.10.001>
42. S. Feng, Q. Hu, W. Qian, Quality control of daily meteorological data in China, 1951–2000: A new dataset, *Int. J. Climatol.*, **24** (2004), 853–870. <https://doi.org/10.1002/joc.1047>
43. W. Skierucha, M. Kafarski, A. Wilczek, A. Szyplowska, S. Jones, P. Sabouroux, et al., *Towards standardization of electromagnetic soil moisture measurements*, In: Proceedings of the 2021 13th International Conference on Electromagnetic Wave Interaction with Water and Moist Substances (ISEMA), IEEE, 2021, 1–4.
44. J. H. Stagge, L. M. Tallaksen, L. Gudmundsson, A. F. V. Loon, K. Stahl, Candidate distributions for climatological drought indices (SPI and SPEI), *Int. J. Climatol.*, **35** (2015), 4027–4040. <https://doi.org/10.1002/joc.4267>
45. A. Mokhtar, M. Jalali, H. He, N. Al-Ansari, A. Elbeltagi, K. Alsafadi, et al., Estimation of SPEI meteorological drought using machine learning algorithms, *IEEE Access*, **9** (2021), 65503–65523.
46. S. Fei, M. A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, et al., UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat, *Precis. Agric.*, **24** (2023), 187–212. <https://doi.org/10.1007/s11119-022-09938-8>
47. Y. Zhang, H. Yu, Uncertainty-aware explainable federated learning, *arXiv Preprint*, 2025. Available from: <https://arxiv.org/abs/2503.05194>.
48. X. Zhang, Y. Li, W. Li, K. Guo, Y. Shao, *Personalized federated learning via variational Bayesian inference*, In: Proceedings of the 39th International Conference on Machine Learning, Proceedings of Machine Learning Research, **162** (2022), 26293–26310. Available from: <https://proceedings.mlr.press/v162/zhang22o.html>.
49. H. Iftikhar, F. Khan, E. A. Torres Armas, P. C. Rodrigues, J. L. López-Gonzales, A novel hybrid framework for forecasting stock indices based on the nonlinear time series models, *Computation. Stat.*, **40** (2025), 4163–4186. <https://doi.org/10.1007/s00180-025-01614-5>
50. H. Iftikhar, S. F. Shah, F. E. Almuhayfith, P. Canas, A kernel regression-based hybrid forecasting framework for highly volatile energy commodity markets, *AIMS Math.*, **11** (2026), 24877–24912. <https://doi.org/10.3934/math.20261000>

51. C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, *On calibration of modern neural networks*, In: Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research, **70** (2017), 1321–1330. Available from: <https://proceedings.mlr.press/v70/guo17a.html>.
52. M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, et al., *Deep learning with differential privacy*, In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, 308–318. <https://doi.org/10.1145/2976749.2978318>
53. M. Park, J. Foulds, K. Chaudhuri, M. Welling, Variational Bayes in private settings (VIPS), *J. Artif. Intell. Res.*, **68** (2020), 109–157. <https://doi.org/10.1613/jair.1.11568>
54. M. Galdo, G. Bahg, B. M. Turner, Variational Bayesian methods for cognitive science, *Psychol. Methods*, **25** (2020), 535–559. <https://doi.org/10.1037/met0000242>
55. M. Ismail, M. Mohsin, A novel varying component Gaussian mixture model for improved drought assessment based on standardized-type drought indices, *Theor. Appl. Climatol.*, 157 (2026), 387. <https://doi.org/10.1007/s00704-026-06324-6>
56. Z. Ali, I. Hussain, M. Faisal, I. M. Almanjahie, I. Ahmad, D. M. Khan, et al., A probabilistic weighted joint aggregative drought index (PWJADI) criterion for drought monitoring systems, *Tellus A*, **71** (2019), 1588584. <https://doi.org/10.1080/16000870.2019.1588584>
57. E. Mathieu, M. Nickel, *Riemannian continuous normalizing flows*, In: Advances in Neural Information Processing Systems, **33** (2020), 2503–2515. Available from: <https://arxiv.org/abs/2006.10605>.



AIMS Press

©2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)