Geothermics 142 (2026) 103816

Contents lists available at ScienceDirect

Geothermics

journal homepage: www.elsevier.com/locate/geothermics

Learning with constraints: A physics-guided Bayesian framework for reliable geothermal temperature mapping in northeastern Italy

Danial Sheini Dashtgoli [a,b,c,*], Michela Giustiniani [b], Martina Busetti [b], Claudia Cherubini [a,b], Guillermo A. Narsilio [c]

[a] Department of Mathematics, Informatics, and Geosciences, University of Trieste, Italy
[b] National Institute of Oceanography and Applied Geophysics - OGS, Sgonico, Trieste, Italy
[c] Department of Infrastructure Engineering, The University of Melbourne, Parkville, Australia

ARTICLE INFO

Keywords:
Geothermal energy
Artificial intelligence
Machine learning
Deep learning
Geothermal reservoir characterization

ABSTRACT

Reliable deep-temperature estimation remains a central challenge in geothermal exploration, particularly in sedimentary–carbonate provinces where borehole data are sparse and vertically imbalanced. These conditions lead to unreliable extrapolation and poor uncertainty characterization in classical interpolation and purely data-driven models. This study develops a physics-guided Bayesian neural network (BNN) framework that integrates heteroscedastic probabilistic learning with conductive heat-transfer constraints to reconstruct the 3D geothermal structure of the Lower Friulian Plain (NE Italy) using 849 irregularly distributed temperature measurements. To explicitly address depth imbalance and evaluate extrapolation performance, a depth-aware training and validation strategy was implemented, including controlled withholding of deep observations (1000–1200 m). The architecture combines Monte Carlo-dropout Bayesian inference with two physics-informed formulations, a monotonicity constraint enforcing non-negative vertical gradients and a gradient-band constraint bounding geothermal slopes within regionally plausible conductive ranges. The resulting models achieve high fidelity in the data-supported domain (<1000 m), with coefficient of determination (R^2) ≈ 0.95–0.96 and root mean square error ≈ 2.6–2.8 °C, while classical and data-driven models show reduced reliability under extrapolation conditions. The monotonic BNN attains R^2 = 0.877 and RMSE = 1.33 °C, and the gradient-band BNN produces physically consistent geothermal gradients (≈0.014–0.018 °C/m), aligning with regional heat-flow constraints. Both physics-guided models deliver near-ideal uncertainty calibration (≈97–99% prediction-interval coverage) and eliminate nonphysical deep-zone oscillations. Independent validation against the Grado-1 well confirms agreement within ±1 °C across the full 100–1200 m profile. The proposed framework enables extrapolation-robust and uncertainty-aware 3D geothermal mapping, providing a transferable approach for geothermal resource assessment in data-limited subsurface systems.

Introduction

Geothermal energy is broadly acknowledged as a clean, reliable, and expandable pillar of future low-carbon energy systems (Ahmadi, 2025). Because subsurface temperature governs both the feasibility and performance of geothermal utilization, reliable deep-temperature estimates are essential for identifying viable resources and reducing drilling risk (Aljubran and Horne, 2024; Shahdi et al., 2021). For decades, geothermal temperature mapping has relied on borehole measurements combined with conductive heat-transfer models, providing valuable large-scale thermal frameworks (Frone et al., 2015; García-Gil et al., 2015; Hermans et al., 2015; Khaled et al., 2023; Nielson et al., 2012). However, these approaches often produce overly smoothed fields and struggle in regions with sparse observations, where interpolation and purely conductive assumptions obscure local anomalies and structural controls (Anvari et al., 2025; Mordret, 2018; Rodriguez-Gomez et al., 2023). These limitations have motivated a shift toward more flexible, data-driven alternatives (Ahmed et al., 2024; Li et al., 2023; L. Wang et al., 2023a; Zheng et al., 2025).

Artificial intelligence (AI) adoption in geothermal research has expanded dramatically since 2019, with geothermal-AI publication rates exceeding 75% per year, far outpacing the general AI growth rate, and

* Corresponding author at: Department of Mathematics, Informatics, and Geosciences, University of Trieste, Italy.
E-mail address: dsheinidashtgoli@ogs.it (D. Sheini Dashtgoli).

https://doi.org/10.1016/j.geothermics.2026.103816
Received 26 January 2026; Received in revised form 21 July 2026; Accepted 5 August 2026
Available online 14 August 2026

the field transitioning from classical models such as artificial neural networks (ANN), support vector machines and random forest, toward deep learning, Bayesian modeling, and physics-informed architectures (Sheini Dashtgoli et al., 2026a). Early applications demonstrated the feasibility of predicting temperature from limited information, including drilling-based estimates and fluid geothermometry (AlGaiar et al., 2024; Bassam et al., 2015; Fahimuddin et al., 2024; Kiran et al., 2022; Pérez-Zárate et al., 2019; L. Wang et al., 2023b), temperature prediction (Haklidir and Haklidir, 2020) and gradient-boosted ensemble methods for regional geothermal mapping (Shahdi et al., 2021). Subsequent studies further expanded these capabilities, introducing hybrid ANN–metaheuristic frameworks for complex geothermal settings (Altay et al., 2022) and natural gradient boosting models enriched with SHAP interpretability to highlight key geochemical drivers such as Silicon dioxide (Ibrahim et al., 2023). In northeastern Italy, Sheini Dashtgoli et al. (2024a) assessed multiple ML models for predicting carbonate-reservoir temperatures in the Lower Friulian Plain and demonstrated that extreme gradient boosting (XGBoost) delivers the most robust performance for this setting. Their sensitivity analysis highlighted bicarbonate as the dominant control on temperature, with magnesium, electrical conductivity, and subsurface depth also contributing meaningfully to thermal variability. Despite these advances, purely data-driven models remain statistical and black-box, rely heavily on dense training data, and degrade when extrapolating across regions with limited or uneven measurements, conditions characteristic of many geothermal provinces (Isah et al., 2025; Manfren et al., 2024; Qin et al., 2025). Integrating physical laws with ML has repeatedly been shown to stabilize geothermal predictions and prevent nonphysical behavior in data-scarce settings. For example, enforcing conductive and energy-balance equations within physics informed neural network (PINNs) markedly improves the consistency of temperature reconstructions (Piere Canilandi et al., 2025), while conditioning models on physics-constrained reservoir simulations reduces extrapolation uncertainty even when well data are sparse (Vogt et al., 2010). These findings highlight that physics-guided learning fundamentally improves geothermal temperature estimation by restricting models to physically admissible solutions. Building on these early insights, the development of physics-informed geothermal modeling has followed a clear chronological progression. Early Monte Carlo based geothermal studies demonstrated that incorporating conductive heat-flow constraints during uncertainty propagation significantly narrows the range of plausible subsurface temperatures (Vogt et al., 2010). A major breakthrough came with physics-informed neural networks. (Ishitsuka and Lin, 2023) developed one of the first geothermal PINNs that jointly inferred temperature, pressure, and permeability by enforcing mass- and energy-conservation equations, achieving far more reliable predictions in unobserved parts of the reservoir than conventional deep neural networks (DNNs). Extending this concept, (Shima et al., 2024) introduced a transfer-learning PINN that pre-trains on multiple physics-consistent geothermal simulations before fine-tuning with sparse local logs, substantially improving stability and accuracy in extrapolation domains. Meanwhile, PINNs were also applied to operational settings. Yuan et al. (2024) incorporated wellbore flow partial differential equations (PDEs) to predict drilling temperature and pressure under extreme downhole conditions, achieving much higher robustness than empirical ANN models.

Recent years have seen a major shift toward physics-guided geothermal temperature mapping, particularly at regional and continental scales. Among these, Aljubran & Horne (2024) produced one of the most advanced large-scale thermal models using a physics-informed graph neural network. By simultaneously predicting temperature-at-depth, heat flow, and thermal conductivity while approximating Fourier's law over a spatial graph enriched with geological, geophysical, and structural features, including gravity and magnetic anomalies, sediment thickness, radioactive heat production, seismicity, and fault proximity, their model generated a 0–7 km continental-scale thermal map of the United States with mean absolute error (MAE) ≈ 6.4 °C and preserved localized anomalies that classical interpolation methods systematically smooth out. Most recently, Ishitsuka et al. (2025) advanced physics-guided ML toward practical geothermal inverse modeling, developing a robust framework that maintains physical consistency while remaining computationally feasible for field applications.

Northeastern Italy, particularly the Lower Friulian Plain, hosts significant geothermal resources, yet its subsurface thermal regime remains incompletely characterized at depth. Temperature observations remain scarce and unevenly distributed, while the regional thermal regime is strongly influenced by hydrochemical variability, stratigraphic transitions, and the structural architecture of the Mesozoic carbonate platform. These gaps persist because drilling, downhole logging, and geochemical sampling are costly and logistically demanding, leading to measurements at only a few locations. In this setting, conventional ML or interpolative schemes compress the true thermal complexity into smoothed gradients, overlook conductive behavior, and misrepresent localized deviations governed by geological controls. What the region needs instead is a model that can extract structure from sparse observations, deliver both temperature and uncertainty estimates, and leverage governing physics to compensate for the limited and uneven field data. A physics-guided Bayesian framework provides this capability. The Bayesian component supplies principled regularization and credible uncertainty quantification, preventing the overconfident extrapolation typical of deterministic neural networks (Fernández et al., 2023; Jospin et al., 2022; Li et al., 2024; Olivieret al., 2021; Zhang et al., 2022; Sheini Dashtgoli et al., 2026b). Embedding geothermal heat-transfer constraints further restricts the inferred temperature field to thermodynamically plausible solutions and enforces consistency with regional geological controls. By uniting these two elements, physics-guided Bayesian learning interprets the sparse borehole measurements of northeastern Italy through the lens of conductive heat-transfer physics rather than purely statistical correlations. This study therefore develops a physics-guided Bayesian neural network tailored to the Lower Friulian Plain, yielding physically consistent deep-temperature predictions while quantifying epistemic uncertainty, an essential requirement for geothermal assessment in carbonate-dominated provinces with limited data. The resulting model delivers the first robust, uncertainty-aware deep-temperature map for northeastern Italy, providing a decision-ready foundation for exploration strategy, drilling-risk reduction, and long-term resource development.

Materials and methods

This section outlines the workflow used to reconstruct a probabilistic 3D temperature field from borehole measurements. It first summarizes the geological context of the Lower Friulian Plain, as the layered and karstified carbonate system controls the depth variability of the dataset. The following subsections describe the database structure, the depth-based train–test partitioning, the preprocessing steps, the Bayesian architecture with its embedded physical constraints, and the procedures used to evaluate predictive performance and uncertainty.

Geological and hydrogeological setting

The study area, located in the Lower Friulian Plain of northeastern Italy, lies on the southwestern margin of the Adriatic–Dinaric Carbonate Platform, where a 4–5 km thick Meso-Cenozoic carbonate succession is overlain by Eocene to Miocene terrigenous deposits (Brancatelli et al., 2023; Busetti, Volpi, Barison, et al., 2010, 2010; Cati et al., 1987). Recurrent subaerial exposure during the Upper Jurassic–Upper Cretaceous produced extensive karstification and enhanced secondary permeability in the upper 1–1.5 km of the carbonate sequence (Velić et al., 2000). The overlying Eocene to Miocene terrigenous deposits, composed mainly of marls and sandstones, act as regional aquitards due to their low-permeability lithologies(Corradin et al., 2025; Cucchi et al,. 2015; Zini et al., 2023). Two fractured carbonate intervals form the main low-enthalpy geothermal aquifer system tapped by the Grado-1 and Grado-2 wells: Paleogene carbonates at ~615–1005 m and Cretaceous carbonates below ~1000 m (Cimolino et al., 2010; Della Vedova et al., 2015; Sheini Dashtgoli et al., 2026c). These units exhibit karst conduits and fracture networks (e.g., at 736–740 m and >1040 m in Grado-1), enabling vertical fluid circulation and temperatures of 42–45 °C (Grado-1) and up to 49 °C (Grado-2) (Della Vedova et al., 2015). Regional thermal springs at Monfalcone and along the Istrian coast indicate basin-scale circulation of waters through faults and fracture systems (Giustiniani et al., 2022; Petrini et al., 2013; Šušmelj et al., 2024). This layered structure, karstified carbonates overlain by low-permeability clastics, produces sharp vertical contrasts in hydraulic and thermal properties. Such contrasts contribute to the strong depth-dependent variability observed in borehole temperatures across the area and motivate the need for physics-guided modeling approaches capable of handling sparse and vertically imbalanced geothermal datasets.

Database

The complete database comprises 849 borehole temperature measurements distributed irregularly across the Lower Friulian Plain (Fig. 1). Each entry includes projected spatial coordinates and a groundwater temperature value T (°C). The dataset shows strong spatial heterogeneity: dense clusters dominate the southern and central sectors, while observations become sparse toward the northeast and coastal margins. This uneven coverage presents challenges for constructing a spatially coherent subsurface thermal model. Depth distribution is similarly imbalanced. Most boreholes are shallow, with 437 samples in the 0–300 m interval, followed by 211 in the 300–600 m range, 110 in the 600–900 m range, and only 89 observations below 900 m. Two deep geothermal wells (Grado-2 and Lignano) provide nearly continuous profiles to approximately 1200 m, whereas the Grado-1 well (approximately 1110 m) is withheld for independent validation. The plan-view map (Fig. 1a) distinguishes the training and test subsets colored by measured temperature, while the 3D visualization (Fig. 1b) highlights the extreme vertical sparsity of deep data relative to the surface domain.

In addition to spatial imbalance, the dataset exhibits a marked depth-dependent thermal structure, as summarized in the depth-binned histograms (Fig. 2). The shallowest interval (0–300 m; N = 436) displays the greatest temperature variability, with a wide interquartile range (IQR ≈ 6.9 °C), a large P5–P95 spread (~21.8 °C), and the highest noise ratio (~0.46). Variability decreases with depth: intermediate intervals (300–600 m and 600–900 m) have progressively narrower IQRs (11.0 °C and 2.2 °C, respectively). The deepest bin (900–1200 m; N = 89) has the lowest noise ratio (~0.10), reflecting a stable, dominantly conductive thermal regime. These observations reveal a complementary imbalance: shallow data are abundant but highly heterogeneous, while deep data are sparse yet thermally consistent. This imbalance motivates the depth–aware weighting and physics-guided learning strategies described in the following sections.

Overall modeling strategy

Building on the spatial and depth-structured characteristics of the dataset described in Section 2.2, the objective of this study is to construct a probabilistic 3D model of subsurface groundwater temperature in the Lower Friulian Plain up to 1200 m depth under conditions of strong spatial sparsity and pronounced vertical imbalance. Because geothermal borehole observations in the region are unevenly distributed both laterally and with depth, the modeling strategy was explicitly designed to address three methodological requirements: (i) probabilistic prediction under limited and clustered sampling, (ii) physically consistent representation of vertical geothermal gradients, and (iii) explicit separation between interpolation within observed depths and prediction beyond them. The overall modeling workflow is summarized in Fig. 3, which illustrates the sequential pipeline from borehole data preparation to probabilistic temperature prediction and uncertainty estimation. Given these constraints, the modeling framework must account for input-dependent variability while maintaining physically plausible behavior outside the sampled depth range. The modeling framework is therefore based on a heteroscedastic Bayesian neural network (BNN) with Monte Carlo dropout, which provides approximate Bayesian inference while allowing predictive uncertainty to vary with both depth and data density. This formulation is required because shallow observations exhibit higher variability and noise, whereas deeper intervals are more stable but poorly sampled, making constant-variance assumptions inappropriate. Unlike PINNs, which explicitly enforce governing partial differential equations and typically require continuous

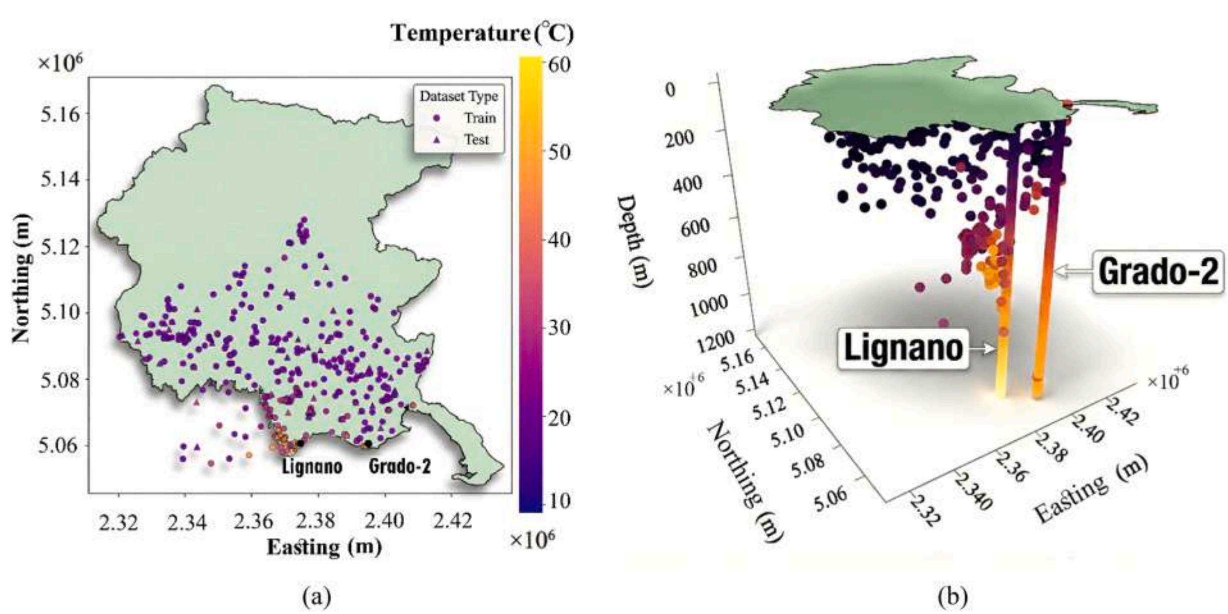

Fig. 1. Spatial distribution of boreholes in the region: a) map view and b) 3D view.

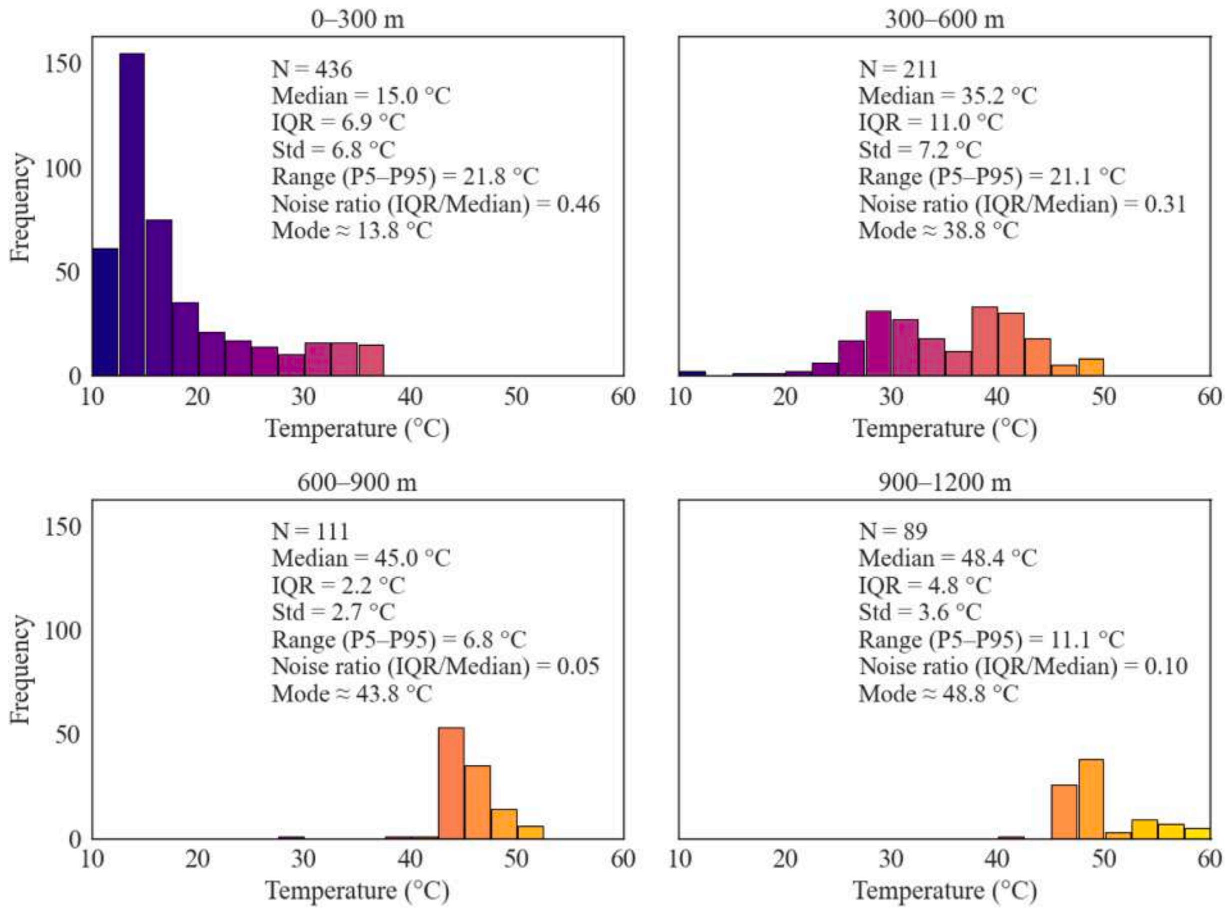

Fig. 2. Depth-binned temperature histograms (0–300, 300–600, 600–900, 900–1200 m) showing frequency distributions and key summary metrics (N, median, IQR, standard deviation, 5–95% range, noise ratio, mode), highlighting the progressive tightening and stabilization of temperatures with depth.

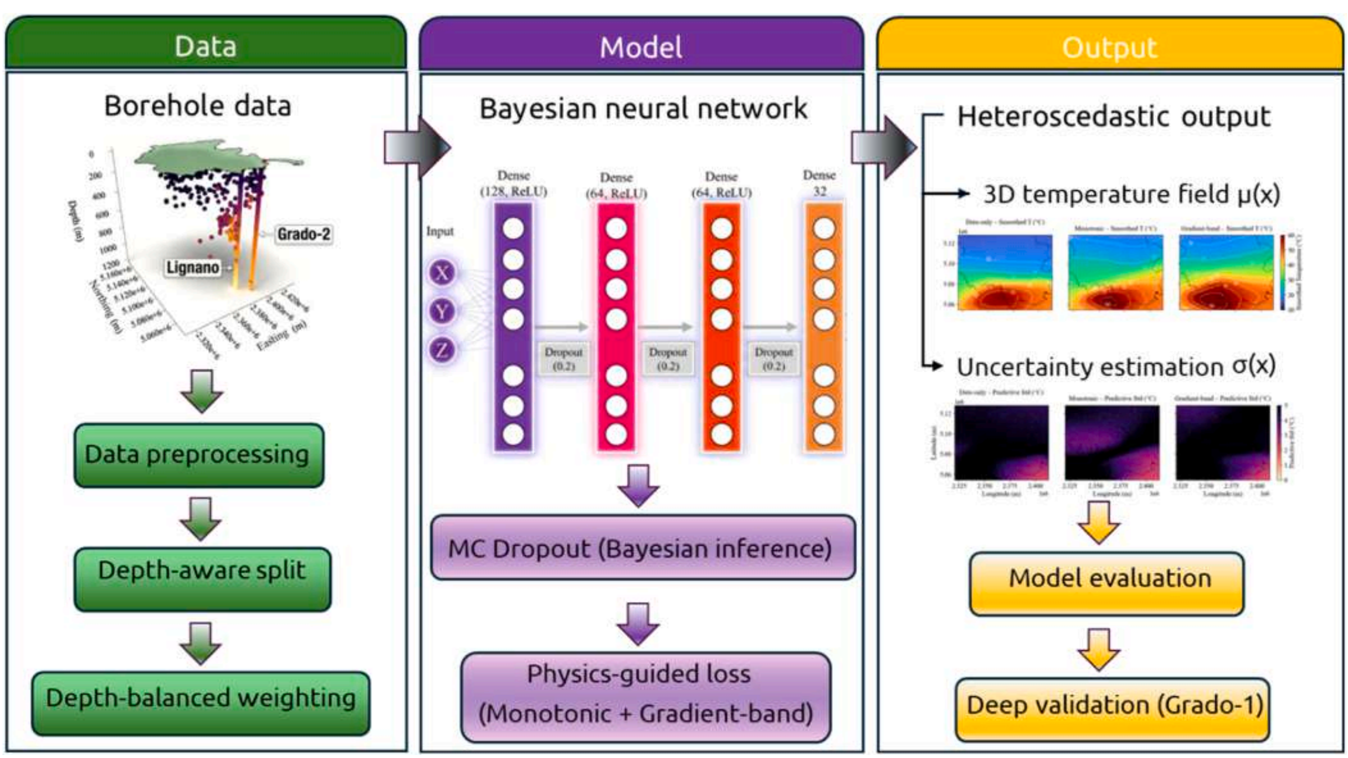

Fig. 3. Overall workflow of the proposed probabilistic geothermal temperature modeling framework.

spatial domains and boundary conditions, the present formulation incorporates physics as soft constraints within the loss function. This choice reflects the limited spatial coverage of the dataset, where strict enforcement of PDE residuals would be poorly constrained. Embedding conductive heat-transfer constraints as soft regularization terms stabilizes vertical temperature gradients and reduces non-physical oscillations that commonly arise in purely data-driven models trained on uneven subsurface datasets. These constraints are introduced to guide predictions in deeper intervals where observational support is limited. Three BNN variants are evaluated to progressively examine the influence of physical constraints: (i) a purely data-driven heteroscedastic BNN, (ii) a monotonic physics-informed BNN enforcing non-decreasing temperature with depth, and (iii) a gradient-band BNN constraining geothermal gradients within a physically plausible conductive range. This hierarchical formulation enables systematic assessment of how increasing levels of physical guidance affect predictive stability, uncertainty calibration, and generalization under data imbalance. It also provides a controlled comparison to isolate the contribution of each constraint. In parallel, representative classical and probabilistic spatial prediction approaches were implemented as methodological benchmarks, including Gaussian Process Regression (GPR), geostatistical kriging, and inverse-distance weighting (IDW). These methods are applied using the same depth-aware partitioning to ensure a consistent basis for comparison. This setup contrasts conventional interpolation with physics-guided probabilistic modeling under identical data conditions. Finally, interpolation (<1000 m) and extrapolation (>1000 m) domains were explicitly separated through depth-aware partitioning. This design provides a controlled framework for evaluating model behavior within the observed range and beyond it, ensuring that performance in deeper zones reflects true out-of-domain prediction.

Depth-aware splitting

To enable robust model generalization under strong vertical imbalance, the complete borehole dataset, comprising spatial coordinates and corresponding groundwater temperature measurements (°C), was filtered to retain only valid observations within the 0–1200 m depth range. This interval captures both the shallow hydrothermal domain and the deeper geothermal system targeted for extrapolation analysis. To explicitly disentangle interpolation within the data-supported domain from extrapolation beyond it, the dataset was divided into two distinct depth subsets. The shallow subset (0–1000 m) was used for model training and in-domain testing, while the deep subset (1000–1200 m) was completely withheld during calibration and served exclusively for extrapolation validation of the trained models. This separation is essential to ensure that model performance in the deep interval reflects true out-of-domain generalization rather than interpolation within partially observed depths. Given the pronounced dominance of shallow records, a depth-stratified sampling strategy was applied to the shallow subset to maintain balanced representation across the vertical profile. Four bins were defined at [0–300], [300–600], [600–800], and [800–1000] m, reflecting the natural clustering of available data. Within these bins, a stratified random split (using a fixed random seed for reproducibility) allocated approximately 80% of the samples for training and 20% for in-domain testing. This approach prevents shallow-data dominance and forces the model to learn depth-dependent thermal structure more uniformly across the vertical profile. The resulting partitioning enables a controlled evaluation of both interpolation performance within the observed domain and extrapolation capability in the deeper geothermal interval. The 1000–1200 m samples were intentionally excluded from the standard test set because they serve a different purpose: they provide a true extrapolation benchmark. This design explicitly defines a strict extrapolation benchmark, in which model predictions are evaluated in a depth interval that is entirely unseen during training. Because the borehole dataset exhibits spatial clustering, the adopted depth-stratified random split may still provide optimistic in-domain estimates in areas with strong local spatial autocorrelation. To mitigate this limitation, model transferability was additionally evaluated using the fully independent Grado-1 geothermal well, which was excluded from model development and retained solely for external validation. This independent well-based validation provides a stricter and more realistic test of spatial generalization than random partitioning alone.

Normalization and depth-balanced sample weights

All input variables were standardized using statistics derived exclusively from the shallow training subset, ensuring that no information from the test or extrapolation domains influenced normalization. For each feature *j*, the standardized variable is computed as:

$$X_{\mathrm{std}} = \frac{X - \mu_X}{\sigma_X} \tag{1}$$

where μ_X and σ_X denote the mean and standard deviation of the feature computed within the training set. The target temperature is standardized in the same way:

$$y_{\mathrm{std}} = \frac{y - \mu_y}{\sigma_y} \tag{2}$$

where μ_y and σ_y are the mean and standard deviation of the temperature target.

This normalization ensures that all input dimensions contribute on comparable numerical scales, improving training stability and avoiding dominance of depth-related magnitudes in the optimization process. To mitigate the depth imbalance observed in the borehole dataset, a continuous depth-balanced weighting scheme was introduced. Let $z_{\min}$ and $z_{\max}$ represent the minimum and maximum depths in the shallow training subset. For each training sample i with depth z_i, a weight w_i is defined as:

$$w_i = 1 + \alpha \, \frac{z_i - z_{\min}}{z_{\max} - z_{\min}} \tag{3}$$

where α is a scaling factor determining the degree of emphasis on deeper samples. In this study, α was set to 2, resulting in the deepest shallow samples receiving approximately three times the weight of the shallowest ones. This weighting ratio was selected empirically to counterbalance the strong dominance of shallow observations while avoiding excessive over-weighting of the sparsely distributed deep samples. All weights are normalized by their mean value to preserve the effective global learning rate during optimization. These weights enter the model loss function as multiplicative factors applied to each sample's negative log-likelihood (NLL) term, thereby enhancing the contribution of deeper data while retaining information from shallower layers. This approach ensures that the BNN learns temperature variations more uniformly across the vertical profile, reducing bias toward densely sampled shallow regions and improving representativeness of the geothermal gradient throughout the domain.

Bayesian neural network framework

BNN architecture consists of three fully connected hidden layers with 128, 64, and 32 neurons, each using the *ReLU* activation function and followed by dropout layers (0.2, 0.2, and 0.2, respectively) (Fig. 4). Two linear output heads predict the mean $\mu_n(x)$ and the raw scale $s(x)$ of the standardized temperature, where the standard deviation is obtained via the *softplus* transformation (see Eq. (4)). Dropout layers remain active during both training and inference (MC dropout) to approximate epistemic uncertainty, with S = 50 stochastic forward passes. In practice, dropout layers are explicitly forced to remain active during inference by evaluating the network in training mode, ensuring stochastic forward passes for Monte Carlo sampling. Predictive statistics (mean and

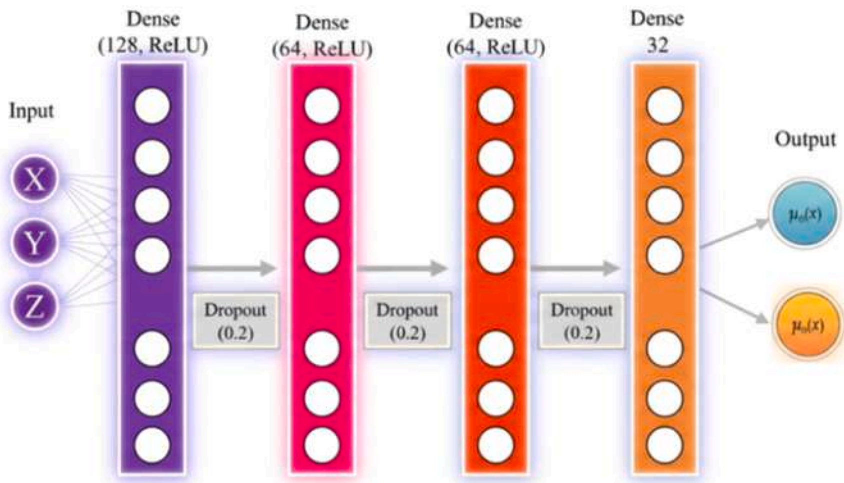

Fig. 4. BNN architecture.

variance) are computed from these stochastic realizations, and all uncertainty quantities are finally rescaled to physical temperature units using the inverse normalization of the target variable. To address vertical imbalance, each sample is weighted according to the depth-dependent scheme defined in Eq. (3). All input variables (X, Y, Z) and the target temperature are standardized using statistics computed exclusively from the training subset to avoid information leakage. The network is trained using the Adam optimizer (learning rate = 10^{-3}), batch size = 32, and up to 300 epochs with early stopping based on validation loss stability. During inference, predictive statistics (mean prediction, total uncertainty, epistemic uncertainty, and aleatoric uncertainty) are obtained from multiple stochastic forward passes and subsequently rescaled to physical temperature units using the corresponding target scaling factors. This unified configuration ensures consistency across all model variants and enables a reliable quantification of uncertainty across the entire depth range.

Heteroscedastic BNN with MC dropout

This section describes the probabilistic modeling framework used to predict subsurface temperature and quantify predictive uncertainty. The approach combines a heteroscedastic Bayesian neural network approximation with Monte Carlo dropout in order to jointly estimate temperature predictions and input-dependent uncertainty. The formulation of the probabilistic model, the associated training objectives, and the physics-informed constraints incorporated in the different model variants are presented below.

Model formulation and uncertainty estimation

The core predictive model is implemented as a BNN approximation using MC dropout. Instead of assigning explicit probability distributions to network weights, epistemic uncertainty is approximated by keeping dropout layers active during both training and inference, thereby performing multiple stochastic forward passes that effectively sample different sub-networks.

For a given input vector $=[x, y, z]$containing spatial coordinates and depth, the network predicts two quantities in the normalized space:

$\mu_n(x)$:the mean of the standardized temperature, and

$s(x)$: the raw scale parameter, which is converted into a strictly positive standard deviation through the softplus transformation:

$$\sigma_n(x) = \ln(1 + e^{s(x)}) + \varepsilon \tag{4}$$

where ε is a small positive constant added for numerical stability. Given these outputs, the conditional likelihood of the standardized target y_{std} is modeled as a Gaussian distribution (Kendall and Gal, 2017):

$$\mathrm{p}(y_{std}|x) = \mathcal{N}\left(\mu_n(x), [\sigma_n(x)]^2\right) \tag{5}$$

This formulation is heteroscedastic, meaning that the predictive variance varies with the input x through $\sigma_n(x)$, thereby allowing the model to capture input-dependent (aleatoric) uncertainty. Epistemic uncertainty is approximated using MC dropout: during both training and prediction, dropout layers remain active so that each forward pass samples a different sub-network. For a given input *x, S* stochastic passes produce $\{\mu_n^{(s)}(x), \sigma_n^{(s)}(x)\}$, for s = 1,…,S. The predictive statistics in normalized space are computed as (Kendall and Gal, 2017):

$$\widehat{\mu}_{\mathbf{n}}(\mathbf{x}) = \mathbf{S}^{-1}\boldsymbol{\Sigma}_{\mathbf{s}}\boldsymbol{\mu}_{\mathbf{n}}^{(\mathbf{s})}(\mathbf{x}) \tag{6}$$

$$\mathbf{Var}_{\mathbf{epistemic,n}}(\mathbf{x}) = \mathbf{S}^{-1}\boldsymbol{\Sigma}_{\mathbf{s}}\left[\boldsymbol{\mu}_{\mathbf{n}}^{(\mathbf{s})}(\mathbf{x}) - \widehat{\mu}_{\mathbf{n}}(\mathbf{x})\right]^2 \tag{7}$$

$$\mathrm{Var}_{\mathrm{aleatoric,n}}(\mathrm{x}) = \mathrm{S}^{-1}\Sigma_{\mathrm{s}}\left[\sigma_{\mathrm{n}}^{(\mathrm{s})}(\mathrm{x})\right]^2 \tag{8}$$

$$\mathrm{Var}_{\mathrm{total,n}}(\mathrm{x}) = \mathrm{Var}_{\mathrm{epistemic,n}}(\mathrm{x}) + \mathrm{Var}_{\mathrm{aleatoric,n}}(\mathrm{x}) \tag{9}$$

The corresponding quantities in physical units are obtained using the inverse scaling of the target variable:

$$\widehat{\boldsymbol{\mu}(\boldsymbol{x})} = \boldsymbol{\mu}_{\boldsymbol{y}} + \boldsymbol{\sigma}_{\boldsymbol{y}}\widehat{\boldsymbol{\mu}}_{\boldsymbol{n}}(\boldsymbol{x}) \tag{10}$$

$$\mathbf{Var}_{\boldsymbol{total}}(\boldsymbol{x}) = \boldsymbol{\sigma}_{\boldsymbol{y}}^2\mathbf{Var}_{\boldsymbol{total,n}}(\boldsymbol{x}) \tag{11}$$

$$\sigma_{total}(x) = \sqrt{Var_{total}(x)} \tag{12}$$

where $\sigma_{total}(x)$represents the total predictive uncertainty combining epistemic and aleatoric components.

Training loss functions for the three models

(a) Data-only model

For the purely data-driven model, the training objective is the depth-weighted heteroscedastic Gaussian negative log-likelihood (NLL) (Kendall and Gal, 2017):

$$L_{data} = \frac{\Sigma_i w_i NLL_i}{\Sigma_i w_i} \tag{13}$$

where, for each sample i in standardized space:

$$\mathrm{NLL}_i = 0.5\left[\ln(2\pi) + 2\ln(\sigma_{n,i}) + \left(y_{std,i} - \mu_{n,i}\right)^2 \sigma_{n,i}^{-2}\right] \tag{14}$$

(b) Physics-informed model I: monotonicity constraint

To improve the physical consistency of the predicted temperature field and to enhance model extrapolation capability, first-order heat conduction physics is embedded into the learning process as soft constraints. These constraints regularize the vertical temperature gradient, discouraging non-physical oscillations while maintaining model flexibility.

Under steady-state conditions, the one-dimensional heat conduction equation with depth-dependent thermal conductivity *k(z)* and volumetric heat generation *A(z)* is expressed as:

$$\frac{d}{dz}\left[k(z)\,\frac{dT}{dz}\right] + \mathrm{A}(z) = 0 \tag{15}$$

Assuming that thermal conductivity is locally constant, and that internal heat generation is negligible ($A \approx 0$), the expression simplifies to:

$$\mathrm{k}\frac{d^2T}{dz^2} = 0 \rightarrow \frac{dT}{dz} = \mathrm{q}k^{-1} = \mathrm{G} \tag{16}$$

where q is the vertical heat flux and G denotes the geothermal gradient.

The first physics-informed formulation enforces the fundamental principle that temperature should not decrease with depth in a conductive regime, i.e.:

$$\frac{dT}{dz} \geq 0 \tag{17}$$

Let $\mu_n(x)$ denote the normalized mean temperature predicted by the network and let z_{std} be the standardized depth. Using automatic differentiation, the vertical derivative of the prediction with respect to standardized depth is computed as:

$$g_{\mathrm{norm}}(x) = \frac{\partial \mu_n(x)}{\partial z_{\mathrm{std}}} \tag{18}$$

To recover the gradient in physical units, the normalization factors of temperature and depth (σ_y, σ_z) are applied:

$$\frac{dT}{dz_{real}} = g_{norm}(x)\left(\sigma_y \sigma_z^{-1}\right) \tag{19}$$

Negative gradients are penalized through a monotonicity regularization term:

$$L_{\mathrm{phys,mono}} = \mathrm{mean}\left[max\left(0,\ -\frac{\mathrm{d}T}{\mathrm{d}z_{\mathrm{real}}}\right)^2\right] \tag{20}$$

The total loss for this model is:

$$L_{total,mono} = L_{data} + \lambda_{mono} L_{phys,mono} \tag{21}$$

where, λ_{mono} controls the strength of the physics constraint relative to data fidelity.

(c) Physics-informed model II: gradient-band constraint

A more informative constraint is imposed in the second physics-informed formulation, which bounds the geothermal gradient within a physically plausible range based on prior knowledge of regional heat flux and thermal conductivity. Given the lower and upper limits of heat flux (q_{min}, q_{max}) and thermal conductivity (k_{min}, k_{max} *),* the permissible range of vertical gradients is defined as:

$$g_{min} = \frac{q_{min}}{k_{max}}, g_{max} = \frac{q_{max}}{k_{min}} \tag{22}$$

This range is informed by regional estimation: surface heat-flow across the Friulian Plain and northern Adriatic domain ranges between 40 and 60 mW/m^2 (Della Vedova et al., 2015). Considering the multilayered nature of the 0–1200 m column, which comprises Meso-Cenozoic carbonates, Eocene–Miocene marls and sandstones and Quaternary clastic deposits, an apparent bulk thermal conductivity in the range of 2–4 W/mK is reasonable based on published values for siliciclastic and carbonate rocks (Pasquale et al., 2011). Combining these bounds implies conductive geothermal gradients of about 0.01–0.03 °C/m, which we adopt as the physically plausible band for the gradient constraint.

For each sample, violations of this physical range are penalized according to:

$$v_{\mathrm{lower}} = \max\left(0,\ g_{\mathrm{min}} - \frac{\mathrm{d}T}{\mathrm{d}z_{\mathrm{real}}}\right) \tag{23}$$

$$v_{\mathrm{upper}} = \max\left(0,\ \frac{\mathrm{d}T}{\mathrm{d}z_{\mathrm{real}}} - g_{\mathrm{max}}\right) \tag{24}$$

The corresponding gradient-band loss function aggregates both lower and upper constraint violations, normalized by the number of valid samples:

$$L_{phys,band} = \frac{\Sigma_i mask_i\left(v_{lower,i}^2 + v_{upper,i}^2\right)}{\Sigma_i mask_i + \varepsilon} \tag{25}$$

and the total objective for the gradient-band physics-informed model is:

$$L_{total,band} = L_{data} + \lambda_{band} L_{phys,band} \tag{26}$$

Here, λ_{band} controls the influence of the gradient-bound constraint, while the mask term ensures that penalties are applied only to valid in-domain samples.

Implementation details of the loss functions

All loss terms are evaluated batch-wise during training using TensorFlow automatic differentiation. For each mini-batch, the network outputs the normalized predictive mean $\mu_n(x)$and a raw scale parameter, which is transformed into a strictly positive standard deviation as $\sigma_n(x) = \mathrm{softplus}(s(x)) + 10^{-6}$. The depth-weighted heteroscedastic negative log-likelihood is then computed over the batch using the normalized target values. For the two physics-guided models, the vertical temperature gradient is obtained by differentiating the predicted normalized mean temperature $\mu_n(x)$with respect to the standardized input variables using GradientTape, and the derivative with respect to the depth coordinate is extracted from the third input dimension. This normalized depth derivative is converted to physical units (°C/m) by multiplying by the ratio σ_y/σ_z, where $\sigma_y$and σ_z are the standard deviations used to normalize temperature and depth, respectively.

In the monotonic formulation, the physics penalty is calculated for each mini-batch as the mean squared ReLU violation of negative vertical gradients. In the gradient-band formulation, lower- and upper-bound violations relative to the admissible conductive gradient interval are first computed for each sample. These violations are then multiplied by a binary mask that activates the constraint only for samples deeper than the selected physical threshold ($z \geq 400$ m in real coordinates). The 400 m threshold was selected empirically based on the observed depth-dependent thermal behavior of the dataset. As shown in the depth-binned analysis (Fig. 2), shallow intervals exhibit substantially higher thermal variability and noise associated with near-surface heterogeneity and local hydrogeological effects, whereas temperatures become progressively more stable and conduction-dominated below approximately 300–400 m depth. Activating the gradient-band constraint only below this transition zone avoids over-constraining the highly variable shallow regime while enforcing physically meaningful conductive gradients in the deeper domain where conductive behavior becomes more dominant. The masked physics term is averaged over the valid masked samples in the batch, and the total loss is obtained by adding this term to the depth-weighted negative log-likelihood using the corresponding regularization coefficient.

To improve numerical stability, a small positive constant (10^{-6}) is added after the softplus transformation of the predictive scale and also in the denominator of the masked gradient-band average to avoid division-by-zero when no valid deep samples are present in a mini-batch. Depth-based sample weights are normalized by their mean value to preserve the overall optimization scale during training. During prediction, dropout remains active by evaluating the network in training mode, and Monte Carlo sampling is performed through repeated stochastic forward passes.

Training and evaluation

All three model variants, the purely data-driven BNN, the monotonic physics-informed BNN, and the gradient-band physics-informed BNN, share an identical network architecture and optimization setup. Each model is trained on the shallow subset (0–1000 m) using the Adam optimizer with early stopping based on validation loss stability. The monotonic and gradient-band formulations incorporate additional physics-based regularization terms weighted by λ_{mono} and λ_{band}, respectively. The deep subset (1000–1200 m) is entirely excluded from training and is used only for out-of-domain extrapolation assessment to evaluate the physical consistency of predictions beyond the calibrated range. Model performance is quantified using multiple complementary metrics that assess accuracy, reliability, and uncertainty calibration. These include the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), NLL, and 95% Prediction Interval (PI) Coverage (Kendall and Gal, 2017; Sheini Dashtgoli et al., 2024b; Wang, 2008).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{\mu}_i\right)^2} \tag{27}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \widehat{\mu}_i\right| \tag{28}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \widehat{\mu}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2} \tag{29}$$

$$\mathrm{PI}_{95} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{I}\left(y_i \in \left[\widehat{\mu}_i - 1.96\sigma_{\mathrm{total},i}, \widehat{\mu}_i + 1.96\sigma_{\mathrm{total},i}\right]\right) \times 100 \tag{30}$$

$$\mathrm{NLL}_{\mathrm{Gaussian}} = \frac{1}{2N}\sum_{i=1}^{N}\left[\ln\left(2\pi\sigma_{\mathrm{total},i}^2\right) + \frac{\left(y_i - \widehat{\mu}_i\right)^2}{\sigma_{\mathrm{total},i}^2}\right] \tag{31}$$

Where:

y_i: Observed (measured) groundwater temperature at sample i.

$\widehat{\mu}_i$: Predicted mean temperature obtained from BNN.

$\overline{y}$: Average of all observed temperatures in the evaluation set.

$\sigma_{\mathrm{total},i}$: Total predictive standard deviation combining epistemic and aleatoric components.

N: Total number of samples in the evaluation subset.

Classical benchmark methods

To contextualize the performance of the proposed physics-informed Bayesian neural networks, three widely used classical spatial prediction approaches were implemented as benchmarks. These methods represent distinct modeling philosophies: probabilistic kernel-based regression (GPR), geostatistical interpolation based on spatial autocorrelation (kriging), and deterministic distance-weighted interpolation (IDW). All benchmark models were trained using the same depth-aware partition (<1000 m for calibration and in-domain testing; 1000–1200 m reserved exclusively for extrapolation evaluation) to ensure a fair comparison with the BNN framework.

Gaussian process regression

GPR is a non-parametric probabilistic regression method that models spatial temperature variation using a covariance kernel that describes similarity between observation points (Deng et al., 2024). The Automatic Relevance Determination Radial Basis Function (ARD-RBF) kernel used here allows anisotropic length scales along spatial coordinates and depth, enabling flexible representation of spatial thermal gradients. GPR provides both predictive mean and uncertainty estimates, making it a useful probabilistic baseline for comparison with the Bayesian neural network approach (Liu et al., 2020).

Three-dimensional kriging

Kriging is a classical geostatistical interpolation technique that predicts subsurface properties based on spatial autocorrelation inferred from experimental variograms (Yang and Jia, 2024). In this study, 3D kriging was implemented to interpolate temperature as a function of spatial coordinates and depth. Unlike machine learning models, kriging does not learn global functional relationships but relies on local covariance structure, which may limit the reliability of extrapolation in sparsely sampled geothermal settings.

Inverse distance weighting (IDW)

Inverse Distance Weighting is a deterministic interpolation method in which predicted values are computed as weighted averages of neighboring observations, with weights inversely proportional to distance (Choi and Chong, 2022). IDW assumes spatial continuity without explicit probabilistic modeling or physical constraints. Although simple, it remains a commonly used baseline in hydrogeological and geothermal mapping studies and therefore provides a useful reference for evaluating more advanced probabilistic learning approaches (Brcković et al., 2025; Masoud et al., 2023).

Results and discussion

With the modeling framework established in Section 2, we now evaluate the performance of the three BNN variants and examine their ability to reconstruct the subsurface temperature field across the Lower Friulian Plain. The results emphasize (i) the stability of predictions across depth, (ii) the degree of uncertainty reduction achieved through physics-informed regularization, and (iii) the model's behavior in data-sparse deep intervals. Particular attention is given to the contrast between shallow zones, where data abundance introduces greater thermal variability, and deeper regions, where limited measurements but stronger physical control yield more stable gradients. Three model variants, Data-only, Monotonic, and Grad-band, were developed to evaluate how physical priors enhance predictive accuracy, uncertainty calibration, and spatial consistency. The following sections present and discuss their quantitative performance, uncertainty behavior, spatial predictions, and independent validation.

Performance evaluation

The predictive performance of the proposed BNN variants (Data-only, Monotonic, and Grad-band) was evaluated alongside classical baseline approaches, including GPR, 3-D geostatistical kriging, and IDW. A deep ensemble benchmark was also implemented by training multiple independently initialized networks with identical architectures and aggregating their predictions to obtain the ensemble mean and predictive variance. Because this comparison serves as a secondary robustness check rather than a primary modeling approach, detailed results are reported in the Supplementary Information. To rigorously assess extrapolation capability, all data deeper than 1000 m were excluded from model fitting, and predictive accuracy was evaluated separately within the in-domain interval (<1000 m) and the deep extrapolation interval (>1000 m). This experimental design allows comparison between interpolation behavior in data-supported regions and generalization capacity under strong depth imbalance. As shown in Table 1, the three BNN variants achieved comparably high accuracy within the data-supported domain (<1000 m). Training performance is similar across variants ($R^2 \approx 0.981$–0.982; RMSE ≈ 1.78–1.83 °C), and in-domain test performance remains consistent ($R^2 \approx 0.954$–0.955; RMSE ≈ 2.65–2.66 °C), indicating stable generalization within the sampled depth range. Within the data-supported domain (<1000 m), 3-D kriging achieved competitive accuracy ($R^2 = 0.951$; RMSE = 2.72 °C), closely matching the BNN variants. GPR produced slightly lower in-domain accuracy ($R^2 = 0.942$; RMSE = 3.00 °C), while IDW remained comparable in R^2 (0.949) but does not provide probabilistic uncertainty estimates. These results confirm that in data-rich shallow conditions, classical spatial interpolators remain strong baselines and can reproduce the observed thermal field with similar accuracy. Clear differences appeared in the deep extrapolation interval (>1000 m), where no training data were available. In this regime, kriging performance declined substantially ($R^2 = 0.417$; RMSE = 2.90 °C), and IDW showed limited predictive reliability ($R^2 = 0.573$). GPR maintained moderate extrapolation capability ($R^2 = 0.69$; RMSE = 2.00 °C), reflecting the smoothness assumptions embedded in its kernel structure. In contrast, under the 1000 m holdout design, the Monotonic and Grad-band BNN variants achieved stronger predictive consistency ($R^2 = 0.877$ and 0.810, respectively) and lower RMSE than the classical baselines. Overall, the results indicate that while classical interpolation techniques remain effective within well-sampled shallow domains, physics-informed probabilistic learning provides improved robustness under depth imbalance and data scarcity. Consistent trends were also observed in additional deep ensemble experiments reported in the Supplementary Information. The following sections therefore examine in greater detail the uncertainty behavior, spatial consistency, and gradient stability of the BNN formulations.

Sensitivity to the 1000 m depth cutoff used for extrapolation testing

Because extrapolation performance may depend on how the depth threshold separating training and withheld data is defined, a sensitivity analysis was conducted by varying the cutoff depth around the reference value of 1000 m (900, 950, 1050, and 1100 m). For each cutoff, all deeper samples were excluded from model fitting and used exclusively for extrapolation evaluation, while the remaining data were partitioned using the same depth-stratified protocol. Fig. 5 presents the variation of extrapolation R^2 and RMSE as a function of the cutoff depth. The results show moderate variability in performance across thresholds, primarily driven by changes in the size and composition of the withheld deep subset. The largest deviation occurs at 900 m, where the unconstrained Data-only model exhibits a clear degradation in performance, while the physics-constrained models remain comparatively stable. Across the range 950–1100 m, performance trends are more consistent, with the Monotonic model generally providing the most accurate extrapolation in terms of RMSE and R^2, and the Grad-band model maintaining stronger control over physically plausible gradient behavior with only minor trade-offs in error metrics. These results indicate that the proposed framework is robust to reasonable variations in the extrapolation boundary. The reference cutoff of 1000 m was retained as it provides a balanced configuration, ensuring a meaningful out-of-domain test while preserving a sufficient number of deep samples for evaluation. This choice is consistent with the observed reduction in data density beyond approximately 900 m, marking the transition to a sparsely sampled regime.

Sensitivity to the physics-constraint weight (λ)

The relative contribution of the physics-based regularization term is controlled by the weight parameter λ, which determines the balance between data fidelity and physical consistency. To assess the sensitivity of model performance to this parameter, both Monotonic and Grad-band BNN variants were trained across a range of λ values spanning several orders of magnitude. Model selection was based on validation RMSE within the in-domain interval (<1000 m), while test and extrapolation metrics were reserved for independent performance evaluation. Fig. 6 summarizes the variation of validation RMSE and gradient violation rate as functions of λ. For both constraint formulations, very small λ values ($\approx 10^{-5}$–10^{-4}) result in weak enforcement of physical constraints, leading to elevated violation rates and less stable extrapolation behavior. As λ increases, violation rates decrease substantially, indicating stronger adherence to the imposed physical priors. At larger λ values (≥ 0.1–0.3), a mild increase in RMSE is observed in some cases, suggesting the onset of over-regularization where the model becomes increasingly constrained relative to the data. However, this effect remains moderate and does not lead to systematic degradation across all subsets. A stable performance regime is observed for intermediate values ($\lambda \approx 0.03$–0.1), where validation RMSE remains low and violation rates are significantly reduced. Within this range, $\lambda = 0.05$ was selected as a representative value that provides a balanced compromise between predictive accuracy and physical consistency. Notably, the relatively flat behavior of the sensitivity curves in this interval indicates that the model is not overly sensitive to the precise choice of λ. This selection falls within the stability plateau of the sensitivity analysis and ensures consistent regularization across both in-domain and extrapolation settings.

Depth-dependent predictive behavior

The detailed sample-sorted plots in Fig. 7 reveal that all three networks reproduce the geothermal trend, with minor deviations at intermediate depths. The *Monotonic* model maintains tighter uncertainty bands compared to *Data-only*, while *Grad-band* produces smoother transitions where thermal gradients naturally stabilize between 400 and 800 m, reflecting improved control over gradient variability. The overlap between the observed and predicted values, together with high 95% PI coverage and low NLL, indicates that both physics-guided models effectively capture the underlying geothermal structure within the data-supported domain. The extrapolation subset (1000–1200 m; Fig. 7c) provides the most critical test of model robustness. Here, despite the complete absence of deep-zone data during training, the *Monotonic* BNN achieves substantially improved predictive consistency ($R^2 = 0.877$, RMSE = 1.33 °C) compared with *Data-only* ($R^2 = 0.750$, RMSE =

Table 1
Performance metrics of the three BNN variants for training, in-domain testing (<1000 m), and deep extrapolation validation (>1000 m).

Model	Train				Test				Validation on the depth>1000m			
	R^2	RMSE	MAE	NLL	R^2	RMSE	MAE	NLL	R^2	RMSE	MAE	NLL
Data-only	0.982	1.78	1.14	2.031	0.954	2.66	1.49	2.247	0.75	1.90	1.089	2.09
Monotonic	0.982	1.79	1.01	1.972	0.955	2.65	1.42	2.191	0.877	1.33	0.715	1.99
Grad-band	0.981	1.83	1.12	1.953	0.955	2.65	1.49	2.234	0.810	1.66	0.916	2.02
GPR (ARD-RBF)	-	-	-	-	0.942	3.00	1.61	2.29	0.69	2.00	1.63	2.21
3-D Kriging	-	-	-	-	0.951	2.72	1.55	-	0.417	2.9	2.46	-
IDW	-	-	-	-	0.949	2.78	1.58	-	0.573	2.07	2.07	-

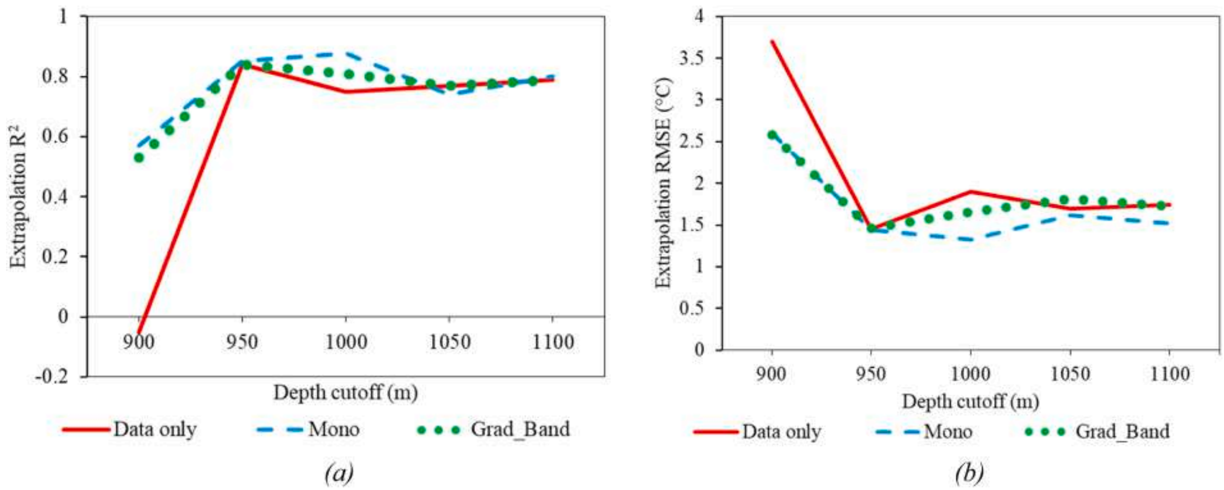

Fig. 5. Sensitivity of extrapolation performance to the depth cutoff for the three BNN variants: (a) R^2; (b) RMSE. Samples deeper than the cutoff are excluded from training. Numbers denote the size of the withheld subset.

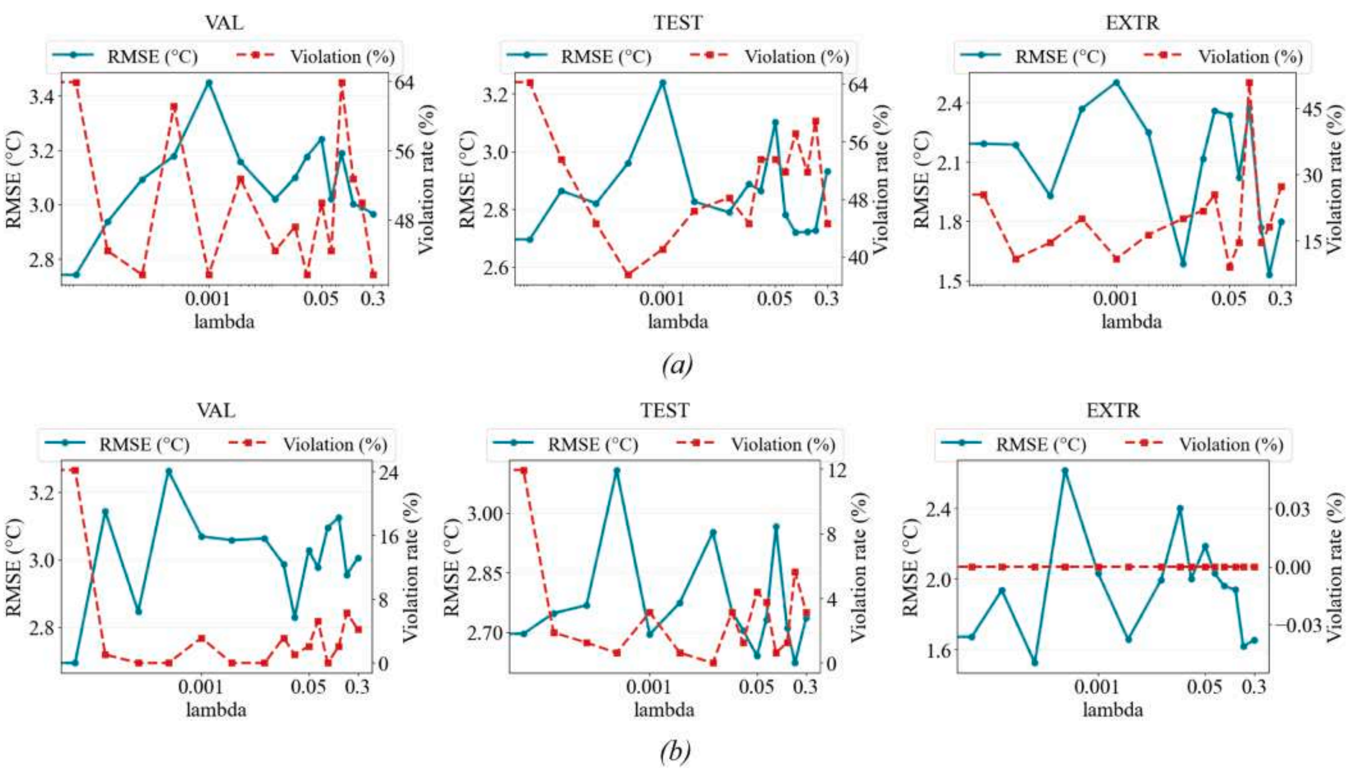

Fig. 6. Sensitivity of model performance to the regularization weight λ for (a) gradient-band and (b) monotonic constraints. RMSE (left axis) and violation rate (right axis) are shown for validation (VAL), in-domain test (TEST), and deep extrapolation (EXTR) subsets.

1.90 °C), while maintaining a 95% PI coverage of ≈ 98% and a reduced NLL (≈ 1.99). The *Grad-band* model also performs robustly ($R^2 = 0.81$), exhibiting a physically realistic gradient trend bounded within the expected geothermal range ($g_{min} \leq dT/dz \leq g_{max}$), ensuring physically plausible temperature evolution with depth. These results indicate that incorporating monotonicity and gradient-band constraints improves extrapolation stability and enforces physically consistent behavior in data-sparse deep zones. The *Data-only* BNN, in contrast, shows larger spread and occasional thermal inversion, indicating that unconstrained learning fails to extrapolate reliably when deeper information is withheld.

Overall, under the 1000 m holdout experiment (Table 1), the Monotonic BNN provides the highest extrapolation accuracy, whereas the Grad-band BNN enforces stricter physical consistency on vertical temperature gradients. This highlights a clear trade-off between statistical accuracy and physical control: the Monotonic formulation is preferable when minimizing extrapolation error is the primary objective, while the Grad-band formulation is more appropriate when adherence to physically bounded gradients is essential. These results confirm that physics-informed priors substantially improve predictive reliability and uncertainty behavior in depth-imbalanced geothermal systems. In practical terms, the Monotonic formulation is more suitable for applications where minimizing predictive error is the primary objective (e.g., temperature estimation for drilling decisions), whereas the Grad-band formulation is preferable for applications requiring physically constrained geothermal gradients (e.g., regional resource assessment and

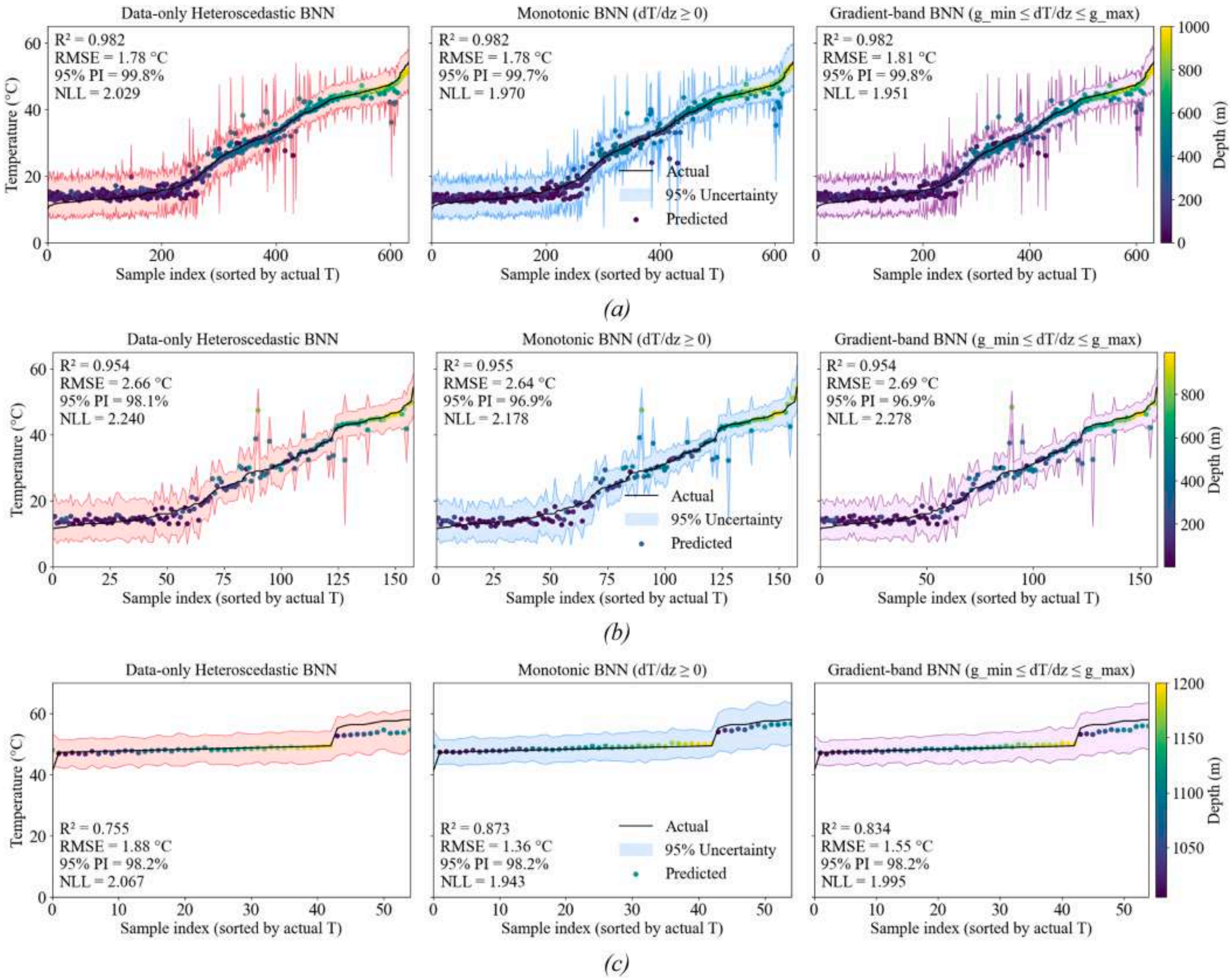

Fig. 7. Comparison of predictive performance for the three BNN variants: (a) training set (<1000 m), (b) test set (<1000 m), and (c) extrapolation validation set (>1000 m).

heat-flow consistency analysis).

Uncertainty analysis

The probabilistic behavior of the three BNN variants was analyzed to assess how including physical constraints affects predictive uncertainty, calibration reliability, and depth-dependent error propagation (Fig. 8). Because the geothermal dataset is sparse and vertically unbalanced, a robust model must provide not only accurate mean predictions but also uncertainty estimates that reflect the true level of epistemic and aleatoric variability across depth. Depth-stratified residuals (Fig. 8a–b) show that the *Monotonic* and *Grad-band* models substantially reduce bias compared to the *Data-only* baseline, particularly below 600 m. For both the training and in-domain test subsets, residuals remain symmetrically distributed around zero across all depth bins, indicating that the physics-guided priors prevent systematic over- or underestimation. In the extrapolation range (Fig. 8c), where no data deeper than 1000 m were used in training, both physics-informed models maintain small residuals and tight interquartile ranges, while the *Data-only* BNN exhibits a clear upward bias. This result confirms that the monotonic and gradient-band constraints effectively preserve the physical continuity of the geothermal gradient, stabilizing predictions in data-scarce deep zones.

Uncertainty magnitude and depth-dependence

The relative uncertainty ratios plotted in Fig. 9 compare the predictive standard deviation of the physics-informed models to the *Data-only* baseline across depth. Both *Monotonic* and *Grad-band* models yield consistently lower or comparable uncertainty (σ ratio < 1) throughout the training and in-domain test domains, showing that physics guidance acts as a regularizer that suppresses spurious variance at shallow depths. At greater depths (1000–1200 m), the uncertainty ratios remain close to unity, implying that both models express a realistic increase in uncertainty without diverging excessively. At the same time, many deep-zone samples exhibit uncertainty ratios moderately below unity (≈0.8–0.95), indicating that the physics-guided constraints suppress part of the non-physical variance amplification observed in the Data-only model during extrapolation. This reduction does not reflect artificial overconfidence, as the physics-informed models simultaneously maintain lower NLL values, high prediction-interval coverage, and stable agreement with the independent Grado-1 validation profile. Notably, the *Grad-band* BNN exhibits slightly higher σ ratios (> 1.1) in a few deep-zone points, corresponding to samples where geothermal gradients approach the upper physical bound (g_{max}). This adaptive increase suggests that the model successfully links uncertainty to physically meaningful thermal variability rather than to numerical instability.

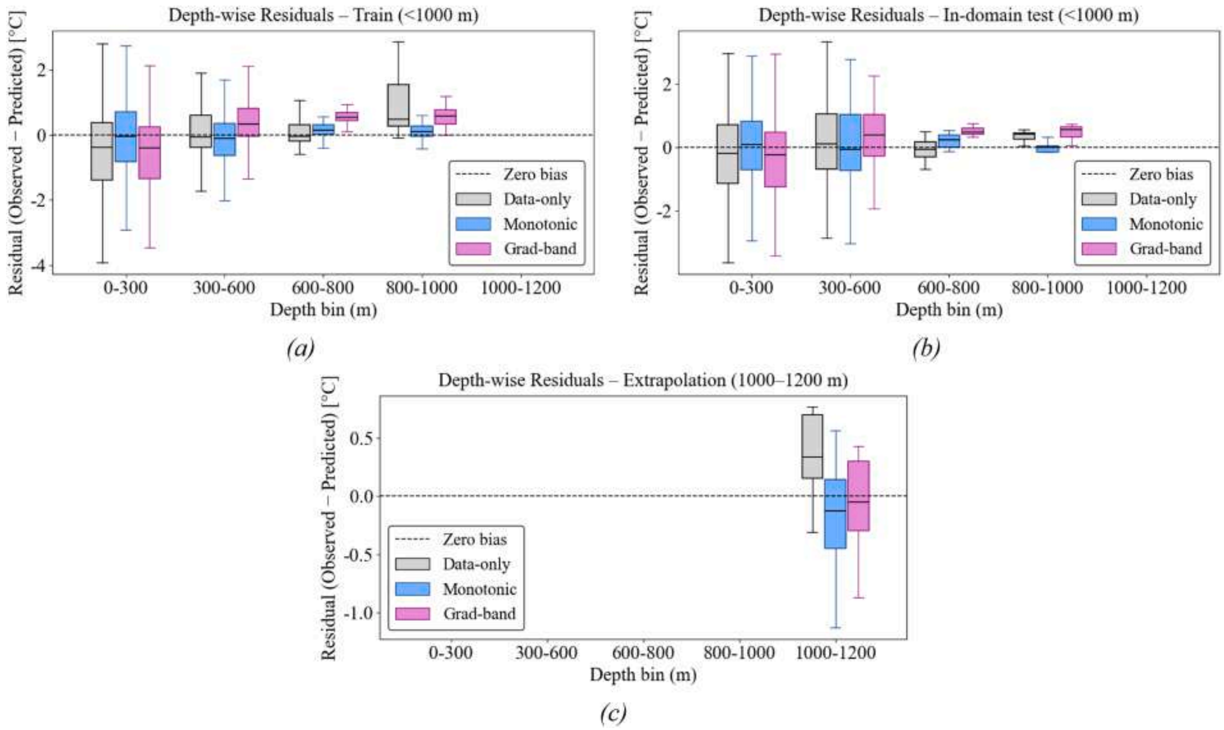

Fig. 8. Depth-wise residuals (Observed – Predicted) for the three BNN variants across (a) training (<1000 m), (b) in-domain testing (<1000 m), and (c) extrapolation validation (1000–1200 m).

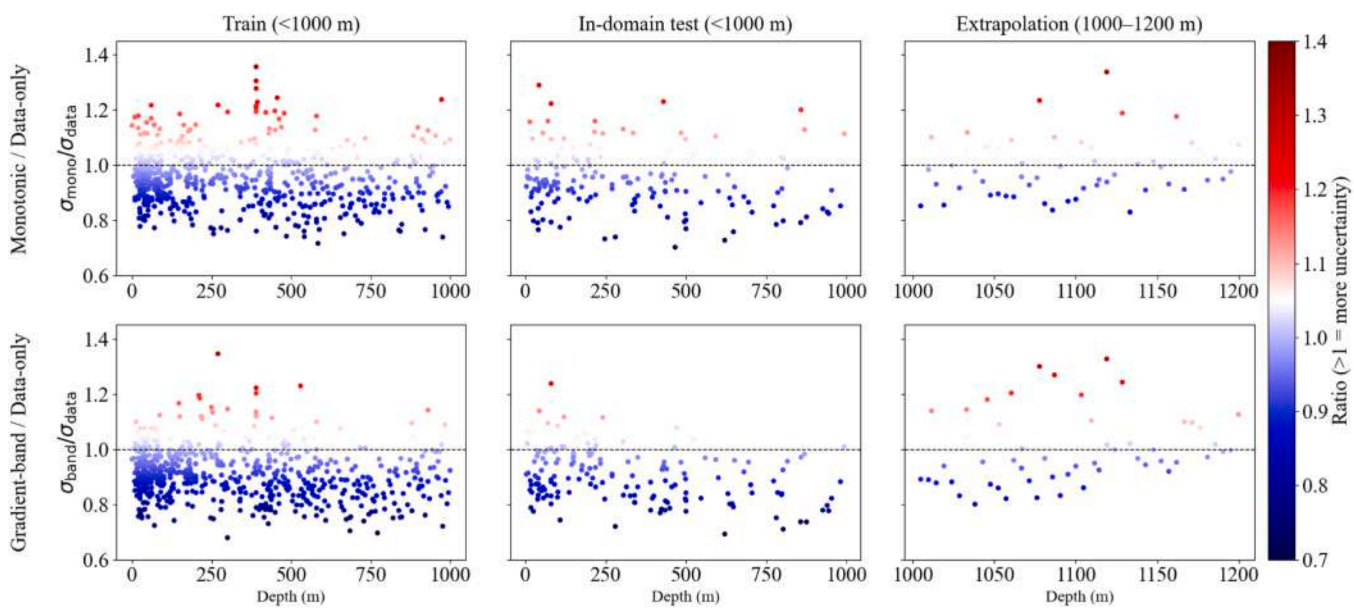

Fig. 9. Depth-wise uncertainty ratio of physics-informed BNNs relative to the Data-only model for training (<1000 m), in-domain testing (<1000 m), and extrapolation (1000–1200 m).

Calibration of predictive intervals

The calibration analysis in Fig. 10 further quantifies the statistical reliability of the predicted uncertainty. The *Monotonic* and *Grad-band* networks both closely follow the ideal 1:1 line, demonstrating near-perfect empirical coverage across all nominal confidence levels. Specifically, the 95% prediction intervals (PIs) achieved empirical coverage of approximately 97–99%, confirming that the Bayesian variance estimates are neither overconfident nor under-dispersed. This high coverage does not indicate overly conservative prediction intervals, as it is supported by consistent NLL values and independent validation against the Grado-1 well, indicating that the uncertainty estimates remain well-calibrated rather than inflated. Correspondingly, NLL values reported in Fig. 7 remain in the narrow range of 1.9–2.2 across all subsets, further supporting well-calibrated probabilistic behavior. In contrast, the *Data-only* model, exhibits slight overconfidence at mid-to-deep depths, consistent with its higher NLL and residual scatter.

Overall, the uncertainty analysis confirms that physically informed BNNs not only improve mean accuracy but also produce uncertainties that are geologically interpretable and statistically consistent. Under the

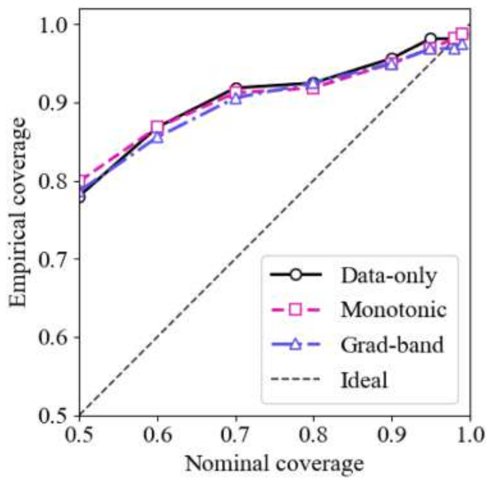

Fig. 10. Predictive interval calibration for the in-domain test set (<1000 m).

adopted split, the Monotonic model exhibits slightly tighter uncertainty bands, while both physics-informed variants remain well-calibrated and avoid overconfident behavior relative to the Data-only baseline. Together, these findings show that integrating monotonic and gradient-band priors yields models capable of distinguishing between genuine geothermal variability and model-induced noise—an essential requirement for reliable geothermal mapping in the Lower Friulian Plain.

Spatial interpretation

The depth-slice temperature maps (Fig. 11−15) illustrate how the three BNN variants reconstruct the geothermal field at multiple depths, integrating mean predictions, smoothed contours, residual noise, and predictive uncertainty. Along with the gradient distributions (Fig. 16), these spatial visualizations reveal the vertical thermal architecture and highlight the stabilizing influence of physics-informed constraints in reproducing geologically coherent temperature patterns.

Noise structure and filtering

At shallow depths (approximately 100–300 m; Fig. 11c−e, and Fig. 12c−e), all models show small-scale fluctuations in the raw mean maps, especially the Data-only BNN, which exhibits pixel-level noise caused by sparse surface observations. After Gaussian smoothing, the large-scale conductive trends remain intact, while the high-frequency artifacts are suppressed, as evidenced by the smoother contour spacing and more stable temperature gradients. The residual noise panels confirm that the Monotonic and Grad-band models inherently produce lower random variance, demonstrating how the imposed physical priors serve as implicit regularizers that prevent overfitting to noisy shallow data.

Spatial uncertainty patterns

The predictive standard deviation maps (bottom rows in each figure) show the lateral and vertical evolution of model uncertainty. Across all layers, uncertainty remains below 2 °C for most of the domain, increasing only near areas with sparse coverage or strong variability due to thermal contrasts. At shallow levels (100–300 m; Figs. 11 and 12), uncertainty is lowest (< 1 °C) in the central and northern sectors where borehole density is highest. At mid-depths (∼ 600 m; Fig. 13), localized uncertainty bands appear along the southern thermal front, corresponding to zones of steep geothermal gradients. In the deep layers (≈ 900–1200 m; Figs. 14 and 15), uncertainty increases moderately (up to 4–5 °C) but remains spatially coherent, closely tracking the thermal anomaly rather than showing random scatter. This pattern matches the uncertainty ratio analysis (Fig. 9), confirming that the variance primarily reflects epistemic uncertainty due to limited deep data, rather than instability in model behavior.

Vertical temperature evolution and hot-zone localization

The mean and smoothed temperature maps across depths (Fig. 11–Fig. 15) delineate a consistent geothermal trend, with temperature increasing from approximately 15 °C at 100 m to 55–60 °C at 1200 m. All models converge in identifying a persistent high-temperature zone in the southern and southeastern portion of the Lower Friulian Plain, coinciding with the carbonate reservoir and fault-controlled upflow zones reported in regional hydrothermal studies. This anomaly intensifies with depth, peaking between 1000 and 1200 m, where the Grad-band model produces the sharpest thermal gradients and most defined isotherm boundaries. The Monotonic variant yields smoother vertical transitions, consistent with conductive–convective equilibrium expected in the deep carbonate system. Overall, the depth-wise progression across Fig. 11–Fig. 15 demonstrates that the physics-informed BNNs—particularly the Grad-band formulation—capture both the physical continuity and spatial variability of the geothermal structure, bridging the gap between data-driven interpolation and physically constrained temperature mapping.

Geothermal gradient

The predicted geothermal gradient distributions (Fig. 16, and Table 2) exhibit a systematic evolution from noisy and heterogeneous near-surface behavior to stable and physically consistent conductive gradients with depth. At 0–300 m, all BNN variants are dominated by low positive gradients (≈ 0.003–0.01 °C/m), consistent with shallow lithological variability and surface perturbations. The Data-only model shows the broadest spread (σ ≈ 0.018 °C/m), reflecting its sensitivity to sparse and noisy shallow data. The Monotonic BNN presents similar central tendencies (mean ≈ 0.014 °C/m) but with smoother transitions, enforcing a physically plausible vertical temperature trend. The Grad-band variant achieves the most coherent behavior, filtering out unstable oscillations and concentrating the distribution around the conductive range (≈ 0.003–0.01 °C/m), thereby stabilizing near-surface predictions. Between 300–600 m, the gradients of all models converge toward realistic conductive values typical of the sedimentary–carbonate succession of the Lower Friulian Plain. The Monotonic and Grad-band BNNs both exhibit dominant modes between 0.025 and 0.035 °C/m, fully consistent with regional conductive regimes, whereas the Data-only model retains a broader tail up to ≈ 0.06 °C/m. This depth interval marks the best balance between regional heat flow and vertical conduction, with the physics-informed formulations exhibiting notably higher internal coherence. In the 600–900 m range, where observational data become sparser, the stabilizing effect of physics-informed priors becomes more pronounced. The Monotonic and Grad-band models maintain consistent mean gradients (≈ 0.015–0.016 °C/m) and narrow dispersion (σ ≈ 0.008 °C/m), while the Data-only model slightly underestimates the conductive slope (≈ 0.013 °C/m). The Grad-band BNN confines gradients within 0.010–0.016 °C/m, closely following the expected regional conductive regime and avoiding unrealistic flattening of deep temperature profiles. At 1000–1200 m, entirely beyond the training domain, the extrapolation capability of each model becomes diagnostic. The Grad-band BNN maintains a narrow, stable gradient range (≈ 0.014–0.018 °C/m) with mean ≈ 0.016 °C/m, demonstrating strong generalization and physical continuity. The Monotonic model shows comparable behavior, whereas the Data-only variant, despite a seemingly tight range (≈ 0.011–0.015 °C/m), indicates mild underestimation due to the absence of physical regulation. Across all depth intervals, the distributions evolve from multi-peaked, noise-influenced shapes in the upper layers to single, sharply defined modes at depth, marking a transition from data-driven variability to physically grounded

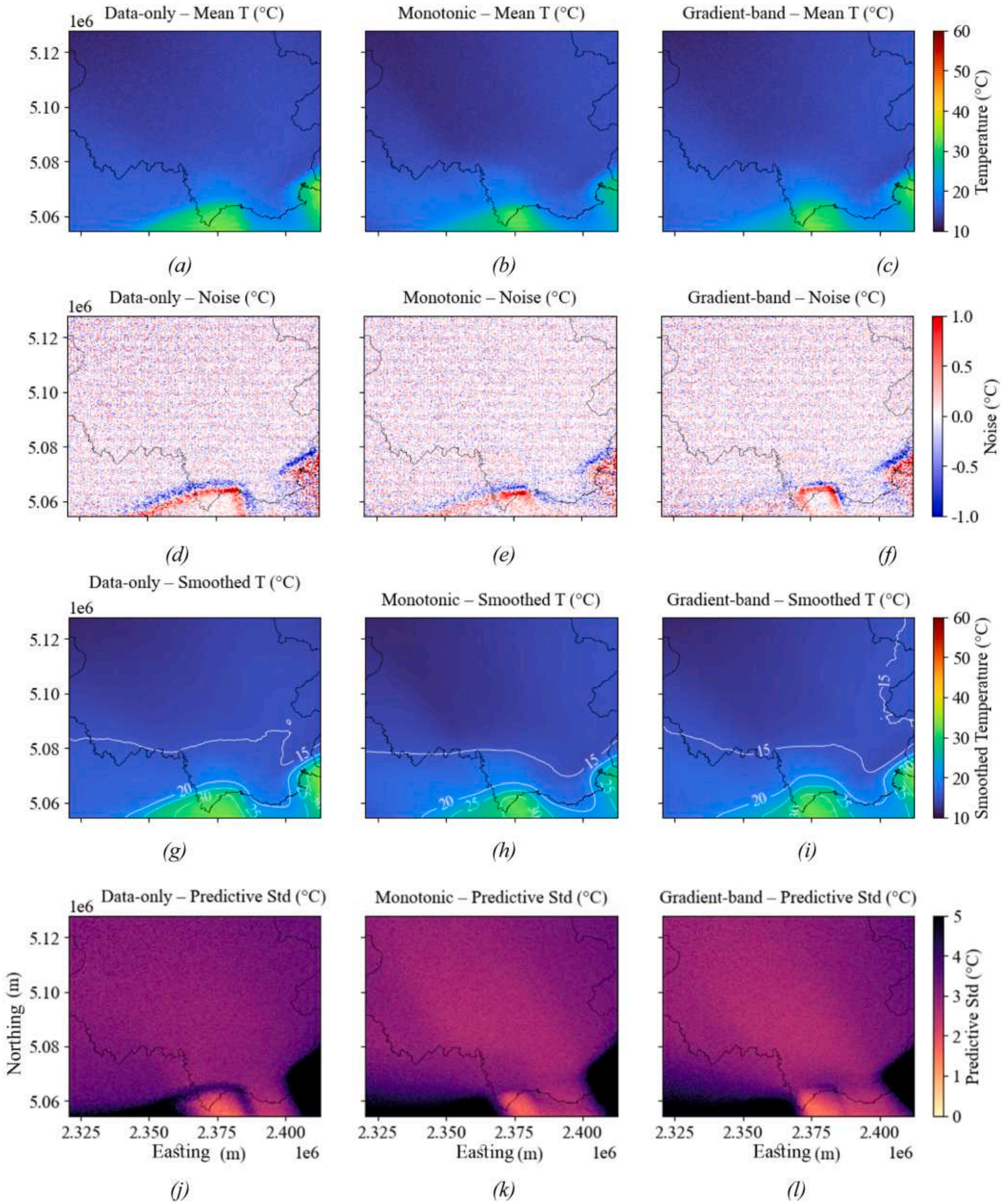

Fig. 11. Depth-slice maps at 100 m showing predicted temperature fields from the three BNN variants. (a–c) Mean temperature distributions; (d–f) local residual noise relative to smoothed fields; (g–i) smoothed temperature isolines highlighting conductive patterns; (j-l) predictive standard deviation illustrating spatial uncertainty.

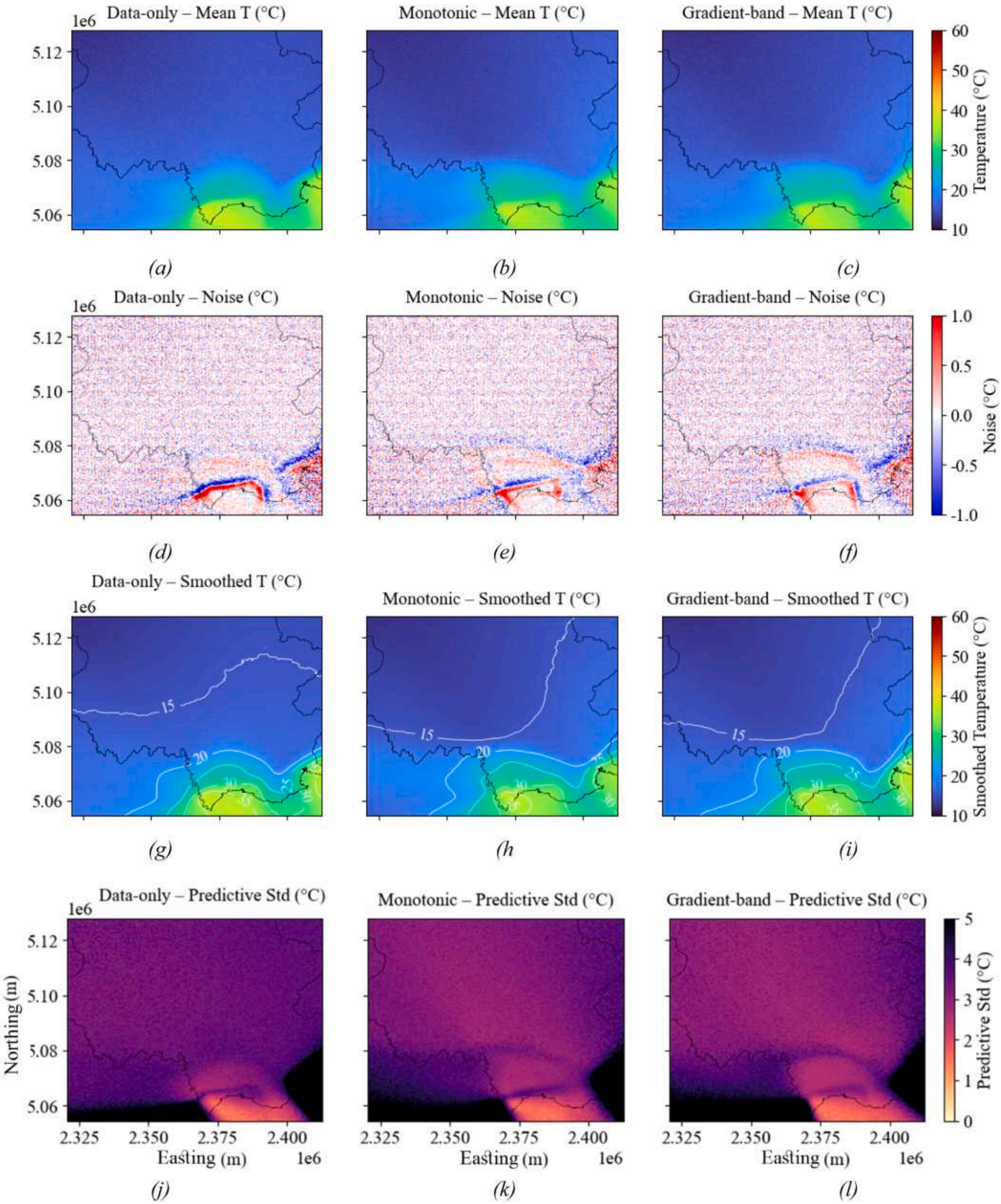

Fig. 12. Depth-slice maps at 300 m showing predicted temperature fields from the three BNN variants. (a–c) Mean temperature distributions; (d–f) local residual noise relative to smoothed fields; (g–i) smoothed temperature isolines highlighting conductive patterns; (j–l) predictive standard deviation illustrating spatial uncertainty.

thermal structure. The dominant geothermal gradients thus stabilize progressively from $\approx 0.005\,^{\circ}$C/m in the shallow zone to $\approx 0.016\,^{\circ}$C/m at 1200 m, corresponding to regional heat fluxes of $\approx$ 65–80 mW/m^2 (reported by Della Vedova et al., 2015) . In summary, the Grad-band BNN demonstrates the highest physical fidelity by maintaining depth-wise consistency and bounding dT/dz within realistic conductive

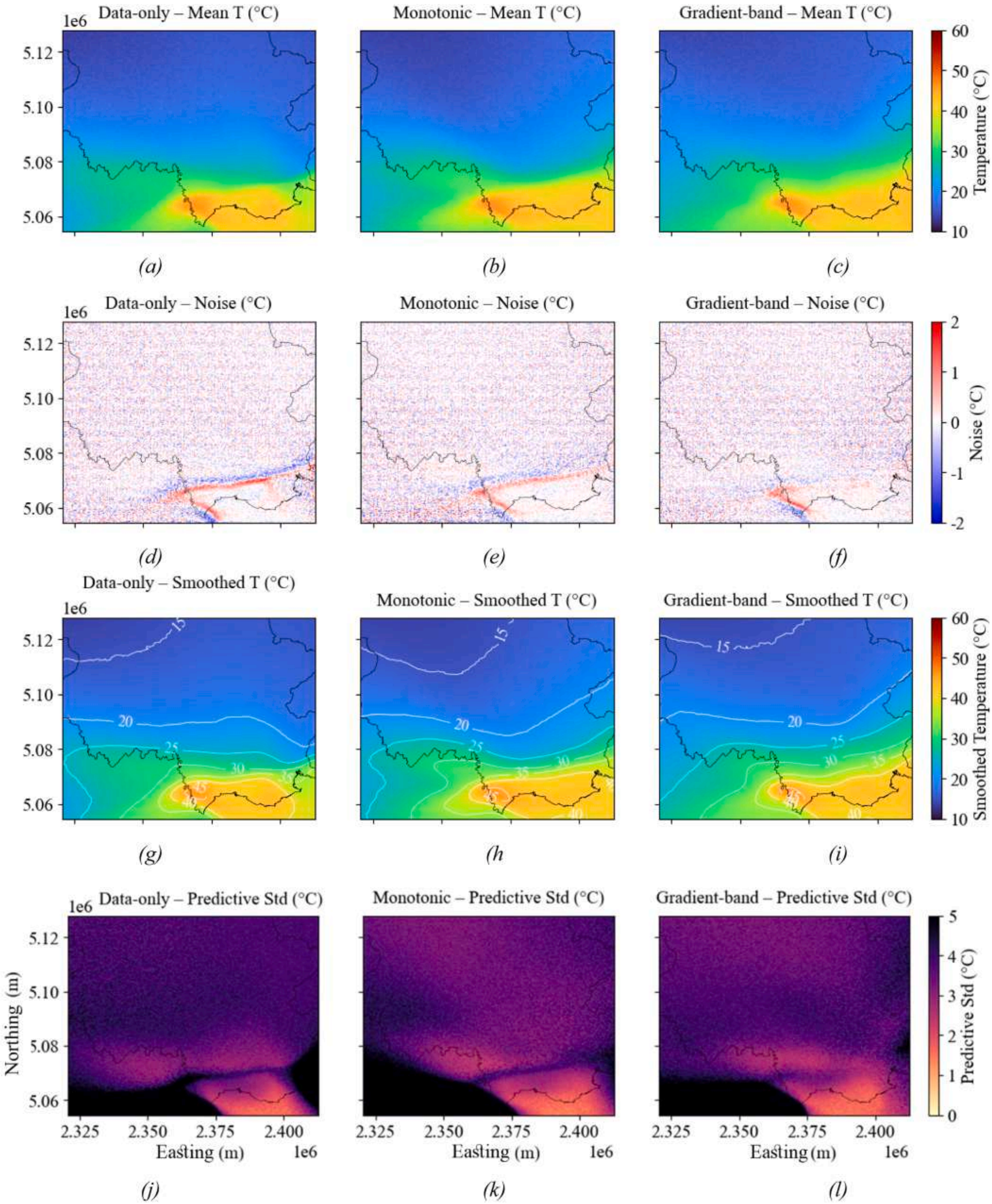

Fig. 13. Depth-slice maps at 600 m showing predicted temperature fields from the three BNN variants. (a–c) Mean temperature distributions; (d–f) local residual noise relative to smoothed fields; (g–i) smoothed temperature isolines highlighting conductive patterns; (j–l) predictive standard deviation illustrating spatial uncertainty.

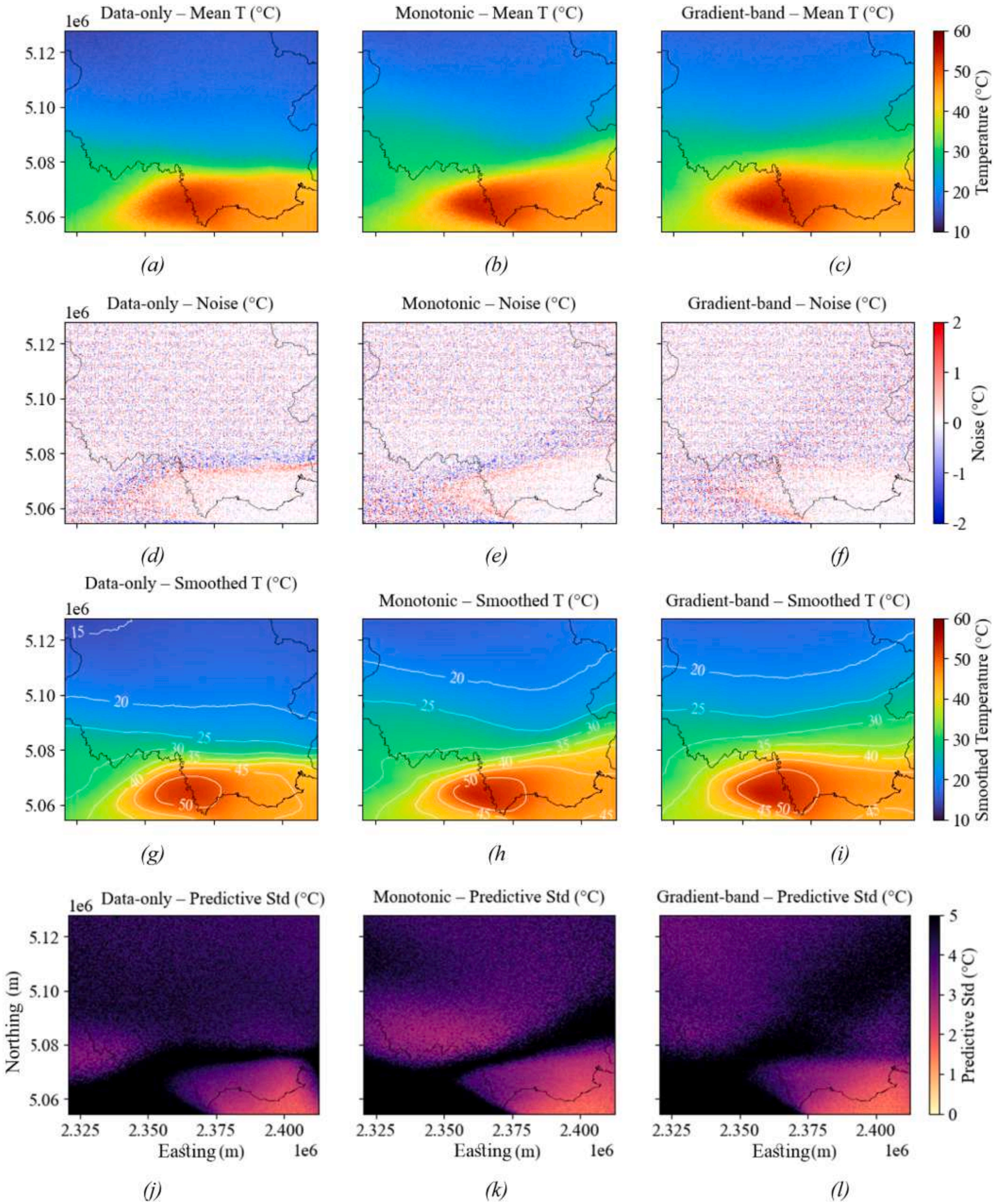

Fig. 14. Depth-slice maps at 900 m showing predicted temperature fields from the three BNN variants. (a–c) Mean temperature distributions; (d–f) local residual noise relative to smoothed fields; (g–i) smoothed temperature isolines highlighting conductive patterns; (j–l) predictive standard deviation illustrating spatial uncertainty.

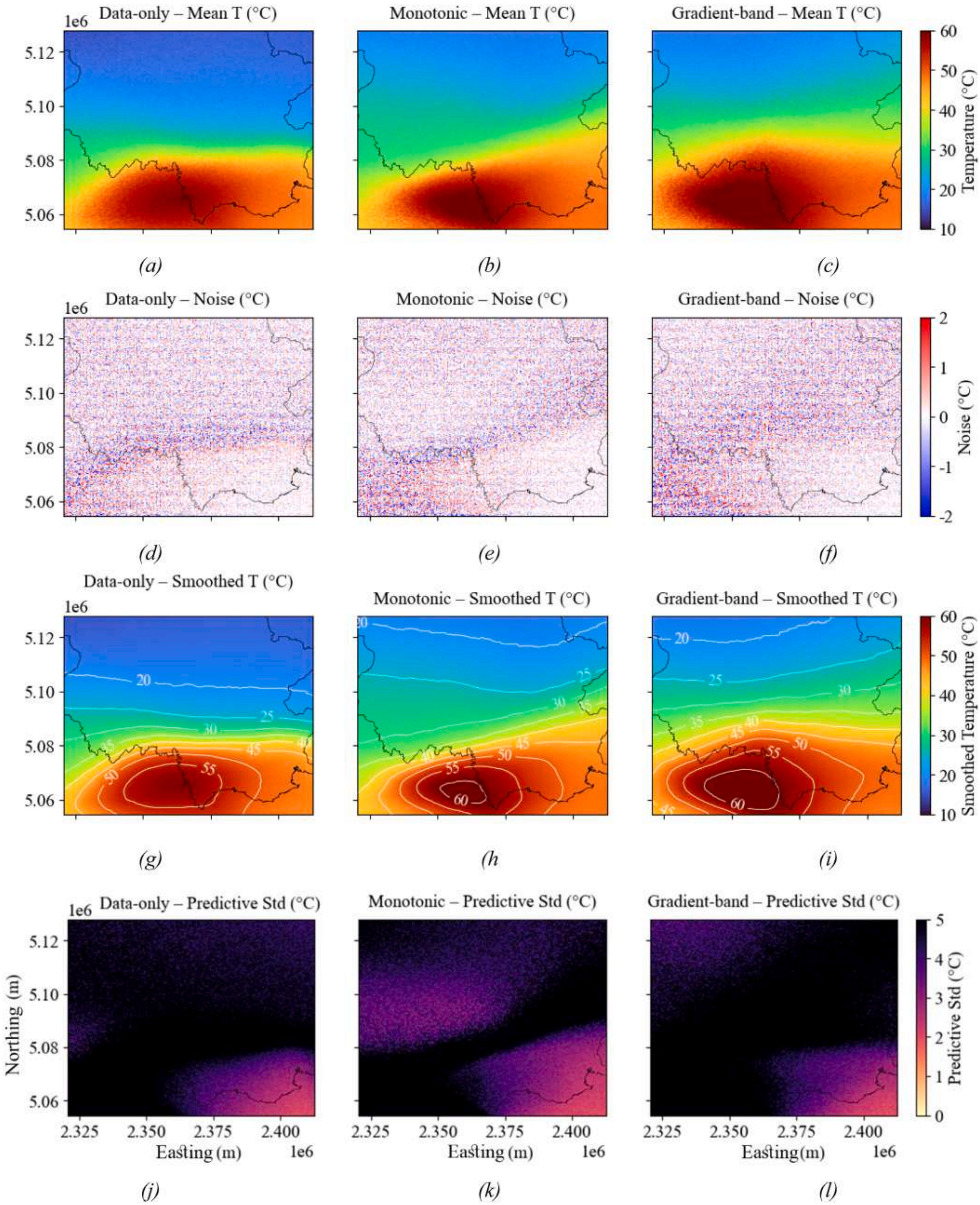

Fig. 15. Depth-slice maps at 1200 m showing predicted temperature fields from the three BNN variants. (a–c) Mean temperature distributions; (d–f) local residual noise relative to smoothed fields; (g–i) smoothed temperature isolines highlighting conductive patterns; (j–l) predictive standard deviation illustrating spatial uncertainty.

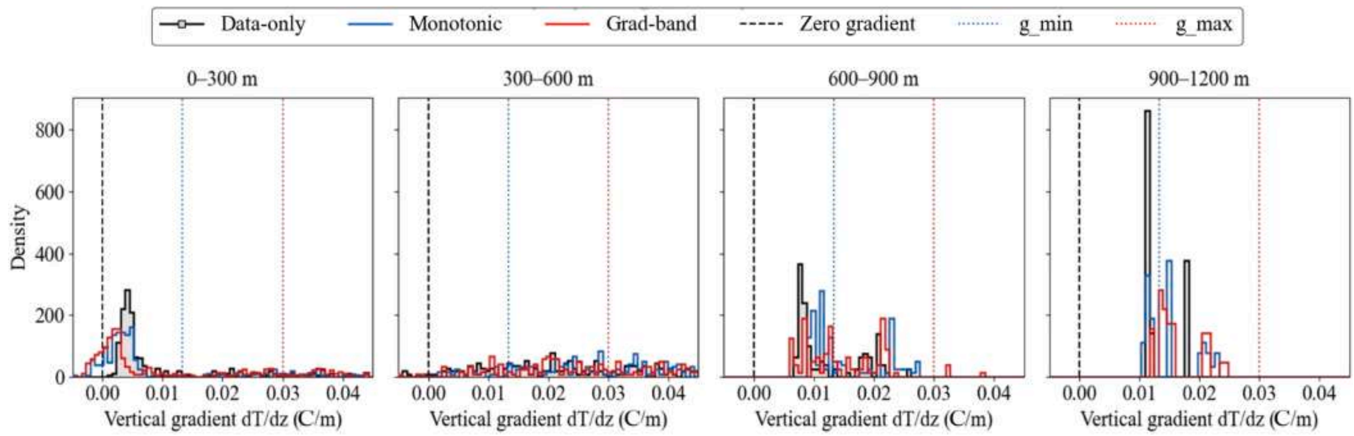

Fig. 16. Predicted geothermal gradient distributions stratified by depth range (0–1200 m) for the three BNN variants.

Table 2
Depth-wise ranges of the most frequent geothermal gradients (dT/dz) predicted by the three BNN variants.

Depth range (m)	Most frequent dT/dz (°C/m)		
	Data-only	Monotonic	Grad-band
0–300	0.004–0.008	0.004–0.010	0.003–0.007
300–600	0.02–0.03	0.025–0.035	0.025–0.035
600–900	0.008–0.012	0.012–0.018	0.010–0.016
900–1200	0.011–0.015	0.013–0.017	0.014–0.018

limits, while the Monotonic BNN ensures smooth vertical continuity. Together, these physics-informed formulations transform the model from a purely statistical interpolator into a physically interpretable framework, capable of reconstructing subsurface thermal regimes with geological reliability, numerical stability, and prediction accuracy.

Validation on the Grado-1 well

The Grado-1 geothermal well, located approximately 1 km west of the Grado-2 well along the coastal margin of the town of Grado, lies within the southern carbonate reservoir of the Lower Friulian Plain and provides an independent benchmark to evaluate model reliability at depths extending beyond the training domain. Fig. 17 compares the predicted temperature–depth profiles from the three BNN variants against the observed borehole log. All models reproduce the general conductive trend, with temperature increasing almost linearly from ≈ 20 °C near 100 m to ≈ 58 °C at 1200 m, consistent with steady conductive heat transfer through the carbonate succession. The Data-only BNN exhibits the widest 95% predictive interval (≈ ± 4 °C) and several local deviations from the measured profile, indicating residual sensitivity to noise and the absence of physical guidance at depth. The Monotonic BNN produces a smoother and physically coherent trend that preserves the expected monotonic temperature rise with depth, though it slightly underestimates the gradient between 400 and 700 m. The Grad-band BNN achieves the closest overall agreement with the measured log, maintaining deviations within ± 1 °C across most of the well and presenting a tightly calibrated 95% uncertainty band that fully encompasses the observations. Below ≈ 1000 m, beyond the trained depth range, the Grad-band model sustains physically consistent extrapolation with no artificial flattening or overshooting of the thermal gradient. Its inferred dT/dz values remain within the conductive range of ≈ 0.01–0.02 °C/m, confirming that the gradient-band prior effectively constrains temperature slopes and preserves realistic deep conductive behavior. The close alignment between predicted and observed profiles indicates that the physics-informed BNNs not only replicate the local stratigraphic structure but also maintain the regional heat-flow regime characteristic of the Grado carbonate reservoir. The Grado I validation confirms that the physics-informed models, especially the Grad-band variant, combine predictive accuracy with strong physical consistency. The Grad-band BNN demonstrates superior extrapolation stability and uncertainty calibration, validating its capability for reliable geothermal temperature mapping in deep carbonate systems of the Friulian coastal region.

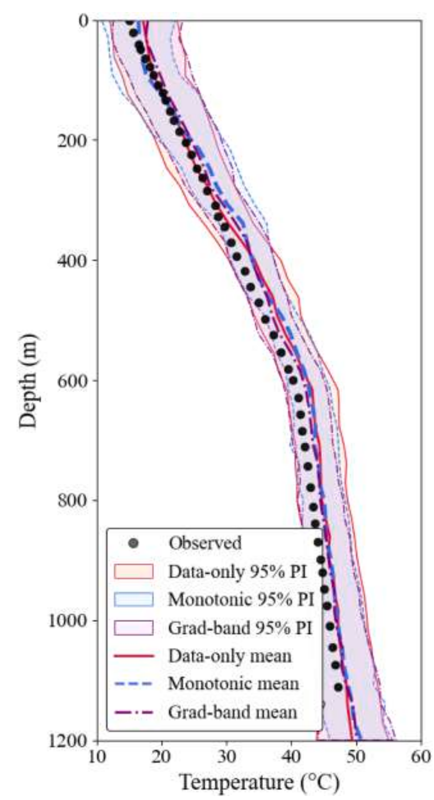

Fig. 17. Comparison between observed and predicted temperature–depth profiles at the Grado-1 well using the three BNN variants.

Limitations and scope of applicability

While the proposed physics-guided Bayesian framework provides stable temperature predictions and well-calibrated uncertainty estimates, several limitations should be acknowledged. The approach adopted in this study is designed as a physics-guided probabilistic regression model rather than a full thermo-hydraulic simulator. The monotonic and gradient-band constraints act as soft physical priors that stabilize the learning process and guide the model toward physically plausible geothermal trends. However, these constraints represent simplified conductive behavior and do not explicitly resolve groundwater flow, fracture-controlled advection, or localized hydrothermal circulation. In heterogeneous carbonate systems, groundwater flow and fracture-controlled circulation may locally perturb the vertical temperature structure. However, the depth-dependent analysis presented in this study indicates that, despite the presence of karstification and fracture networks, the observed temperature profiles exhibit predominantly conductive behavior at the regional scale, as reflected by stable geothermal gradients and consistent temperature–depth trends in the dataset and in the independent Grado-1 validation. Accordingly, the imposed constraints should be interpreted as regional-scale regularization mechanisms that remain appropriate under conduction-dominated conditions, rather than complete descriptions of local thermo-hydraulic processes. In settings where advective heat transport becomes dominant over conduction, such as strongly convective or highly fault-controlled systems, the applicability of these constraints may be reduced and should be considered with caution. Another limitation relates to the structure of the available dataset. Borehole observations are spatially clustered and strongly depth-imbalanced. The adopted validation strategy therefore focuses on depth-based extrapolation testing, which is the primary challenge in geothermal exploration in this basin. Due to the limited availability of deep geothermal wells (only three wells exceeding ~1000 m), a full spatial block cross-validation could not be reliably implemented; instead, Grado-2 and Lignano wells were included within the training/testing framework, while Grado-1 was entirely excluded and used as an independent validation dataset. Although this design provides a rigorous evaluation of deep prediction capability, successful depth extrapolation should not be interpreted as evidence of equivalent spatial transferability, since geothermal conditions may vary laterally due to differences in geological structure, fault systems, groundwater recharge, and regional heat-flow patterns. Future studies could therefore complement the present framework with spatial block cross-validation to further assess transferability across different locations and sensitivity to lateral sampling bias. Finally, the uncertainty estimates are obtained using MC dropout as an approximate Bayesian inference technique. While supplementary comparisons with deep ensembles confirm that the main depth-dependent uncertainty trends remain consistent across inference approaches, MC dropout does not represent the full posterior distribution. In addition, the present formulation was developed for a low-enthalpy carbonate basin dominated by conductive thermal behavior; application to volcanic or strongly convective geothermal systems may require modified physical constraints and additional process representations. As a consequence, model performance in spatially distant or geologically distinct areas that are not represented by the available wells may be lower than reported here.

Conclusions

This study reconstructed the deep geothermal temperature structure of the Lower Friulian Plain using a sparse dataset of 849 borehole temperature measurements with strong spatial and depth imbalance, including only two wells extending to ~1200 m. To mitigate this imbalance and enable reliable deep forecasting, a depth-aware training strategy was implemented, combining stratified sampling of the 0–1000 m interval, depth-balanced likelihood weighting to prevent shallow-data dominance, and full withholding of the 1000–1200 m measurements to provide a true out-of-domain extrapolation test. Building on this depth-aware experimental design, the proposed framework integrates physics-constrained Bayesian learning with controlled extrapolation validation, enabling direct assessment of model behavior beyond the data-supported domain. Three heteroscedastic MC-dropout Bayesian neural networks were then developed, a data-only BNN, a monotonicity-constrained BNN, and a gradient-band BNN, in which geothermal heat-transfer physics was explicitly embedded as model constraints to improve extrapolation reliability and physical consistency. Classical benchmark approaches, including Gaussian Process Regression, 3-D kriging, and inverse-distance weighting, were additionally evaluated to place the proposed framework within the broader context of geothermal temperature interpolation methods. The main findings are as follows:

- Comparable interpolation accuracy: Within the data-supported domain (<1000 m), both physics-informed BNNs and classical interpolators achieve similar accuracy, indicating that the primary advantage of the proposed framework emerges under extrapolation conditions.
- Improved extrapolation robustness: In the main 1000–1200 m holdout, physics-informed BNNs improved deep prediction reliability relative to the unconstrained model and classical interpolators; the Monotonic variant achieved the best extrapolation accuracy ($R^2 = 0.877$; RMSE = 1.33 °C), whereas the Grad-band variant ensured physically bounded gradients. This highlights the ability of physics constraints to regularize predictions in depth intervals completely unseen during training.
- Physical consistency: The gradient-band BNN effectively constrained conductive gradients to ~0.014–0.018 °C/m, aligning with regional geological expectations and removing deep non-physical oscillations.
- Reliable uncertainty quantification: Physics-guided models achieved near-ideal calibration (≈98–99% PI coverage) and lower NLL, enabling uncertainty-aware geothermal mapping and risk-informed decision-making.
- Uncertainty-resolved geothermal mapping: The framework produces a region-wide 3D temperature field with spatially explicit uncertainty, providing a decision-ready basis for geothermal exploration and drilling risk assessment.
- Improved spatial coherence: Physics constraints suppressed shallow artifacts, stabilized mid-depth structure, and preserved conductive deep trends in the 3D temperature field.
- Independent validation: The gradient-band model reproduces the Grado-1 well temperature profile within ±1 °C across 100–1200 m depth, supporting predictive reliability at an independent well location.
- Controlled extrapolation framework: The depth-aware holdout strategy enables a clear separation between interpolation and extrapolation, providing a rigorous benchmark for evaluating model generalization in depth-imbalanced geothermal datasets.

In summary, the proposed physics-guided Bayesian neural network provides a unified framework for uncertainty-aware and physically consistent geothermal temperature reconstruction under data-sparse and depth-imbalanced conditions. By explicitly integrating geothermal heat-transfer constraints within a probabilistic learning architecture and validating performance under controlled extrapolation, the model demonstrates improved robustness beyond the training domain. The resulting approach enables reliable 3D temperature mapping and risk-informed decision-making, offering a transferable methodology for geothermal resource assessment in carbonate and similarly structured sedimentary systems.

CRediT authorship contribution statement

Danial Sheini Dashtgoli: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michela Giustiniani:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition. **Martina Busetti:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation. **Claudia Cherubini:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation. **Guillermo A. Narsilio:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was carried out within the PNRR research activities of the consortium iNEST (Interconnected North-East Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.5 – D.D. 1058 23/06/2022, ECS_00000043). This manuscript reflects only the Authors' views and opinions.

The authors thanks the "Servizio Geologico della Direzione centrale difesa dell'ambiente, energia e sviluppo sostenibile della Regione Autonoma Friuli Venezia Giulia" for providing part of the data used in this study.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.geothermics.2026.103816.

Data availability

Data will be made available on request.

References

Ahmadi, M., 2025. Interpretable machine learning for high-accuracy reservoir temperature prediction in. Geotherm. Energy Syst. Energy 18 (13), 3366. https://doi.org/10.3390/en18133366.

Ahmed, N., Assadi, M., Zhang, Q., Śliwa, T., 2024. Data-driven insights for improved heating and cooling predictions: impact of input parameters on multivariate deep learning algorithms using geothermal borehole field data. Appl. Therm. Eng. 245, 122870. https://doi.org/10.1016/j.applthermaleng.2024.122870.

AlGaiar, M., Hossain, M., Petrovski, A., Lashin, A., Faisal, N., 2024. Applications of artificial intelligence in geothermal resource exploration: a review. Deep Undergr. Sci. Eng. 3 (3), 269–285. https://doi.org/10.1002/dug2.12122.

Aljubran, M.J., Horne, R.N., 2024. Thermal earth model for the conterminous United States using an interpolative physics-informed graph neural network (InterPIGNN) (Version 1). ArXiv. https://doi.org/10.48550/ARXIV.2403.09961.

Ezgi Altay, E.V.G., Altay, O., Dikici, A., 2022. Hybrid artificial neural network based on a metaheuristic optimization algorithm for the prediction of reservoir temperature using hydrogeochemical data of different geothermal areas in Anatolia (Turkey). Geothermics 104, 102476. https://doi.org/10.1016/j.geothermics.2022.102476.

Anvari, K., Benndorf, J., Gerber, G., Alisch, U., 2025. Hybrid geostatistical and deep learning framework for geochemical characterization in historical mine tailings. Sci. Rep. 15 (1), 35004. https://doi.org/10.1038/s41598-025-19441-5.

Bassam, A.A., del, C.A, Garcia-Valladares, O., Santoyo, E., 2015. Determination of pressure drops in flowing geothermal wells by using artificial neural networks and wellbore simulation tools. Appl. Therm. Eng. 75, 1217–1228. https://doi.org/10.1016/j.applthermaleng.2014.05.048.

Brancatelli, G., Busetti, M., Dal Cin, M., Forlin, E., 2023. Reprocessing the CROP95-M18 vintage multichannel seismic data acquired in the northern Adriatic Sea: the case of high penetration crustal profile recorded in shallow waters. Bull. Geophys. Oceanogr. 64 (3), 213–236.

Brcković, A., Malvić, T., Orešković, J., Kapuralić, J., 2025. Comparison of neural network, ordinary kriging, and inverse distance weighting algorithms for seismic and well-derived depth data: a case study in the Bjelovar subdepression. Croat Geosci. 15 (6), 206. https://doi.org/10.3390/geosciences15060206.

Busetti, M., Volpi, V., Barison, E., Giustiniani, M., Marchi, M., Ramella, R., Wardell, N., Zanolla, C., 2010a. Meso-cenozoic seismic stratigraphy and the tectonic setting of the Gulf of Trieste (northern Adriatic). GeoActa 3, 1–14.

Busetti, M., Volpi, V., Nicolich, R., Barison, E., Romeo, R., Baradello, L., Giustiniani, M., Marchi, M., Zonlla, C., Wardell, N., 2010b. Dinaric tectonic features in the Gulf of Trieste (Northern Adriatic). Boll. Geofis. Teor. Appl. 51 (2–3), 117–128.

Piere Canilandi, M., Sutopo, Ashat, A., Dominicus Sinurat, P., 2025. The application of physics-informed neural networks (PINNS) for geothermal well thermal recovery post-injection. IOP Conf. Ser. Earth Environ. Sci. 1456 (1), 012006. https://doi.org/10.1088/1755-1315/1456/1/012006.

Cati, A., Sartorio, D., Venturini, S., 1987. Carbonate platforms in the subsurface of the northern Adriatic area. Mem. Soc. Geol. Ital. 40, 295–308.

Choi, K., Chong, K., 2022. Modified inverse distance weighting interpolation for particulate matter estimation and mapping. Atmosphere 13 (5), 846. https://doi.org/10.3390/atmos13050846.

Cimolino, A., Della Vedova, B., Nicolich, R., Barison, E., Brancatelli, G., 2010. New evidence of the outer dinaric deformation front in the Grado area. NE-Italy Rend Lincei 21 (S1), 167–179. https://doi.org/10.1007/s12210-010-0096-y.

Corradin, C., Thomas, A.T., Camerlenghi, A., Zini, L., Giustiniani, M., Busetti, M., Foglia, L., Bertoni, C., Micallef, A., 2026. Characterization of an onshore–offshore aquifer system in the Venetian Friulian Plain and north Adriatic Basin: a 3D modeling approach. Hydrogeol. J. 34, 301–324. https://doi.org/10.1007/s10040-025-02952-w.

Cucchi, F., Zini, L., Calligaris, C., 2015. Le Acque Del Carso Classico: Progetto Hydrokarst. EUT Edizioni Università di Trieste.

Sheini Dashtgoli, D., Sadeghian, R Reza Mahboubi Ardakani, A., Mohammadnezhad, H., Giustiniani, M., Busetti, M., Cherubini, C., 2024b. Predictive modeling of shallow tunnel behavior: leveraging machine learning for maximum convergence displacement estimation. Transp. Geotech. 47, 101284. https://doi.org/10.1016/j.trgeo.2024.101284.

Deng, Y., Eden, M., Cremaschi, S., 2024. A Gaussian process embedded feature selection method based on automatic relevance determination. Comput. Chem. Eng. 191, 108852. https://doi.org/10.1016/j.compchemeng.2024.108852.

Fahimuddin, A.K.M., Hussein, W., Mezghani, M., Anifowose, F., 2024. Data driven subsurface workflow for geothermal resource exploration: applications of machine learning methodology. ADIPEC. https://doi.org/10.2118/222254-MS.

Fernández, J., Chiachío, J., Chiachío, M., Barros, J., Corbetta, M., 2023. Physics-guided bayesian neural networks by ABC-SS: application to reinforced concrete columns. Eng. Appl. Artif. Intell. 119, 105790. https://doi.org/10.1016/j.engappai.2022.105790.

Frone, Z.S., Blackwell, D.D., Richards, M.C., & Hornbach, M.J. (2015). Heat Flow and Thermal Modeling of the Appalachian Basin, West Virginia. Geosphere, 11(5), 1279–1290.

García-Gil, A., Vázquez-Suñe, E., Alcaraz, M.M., Juan, A.S., Sánchez-Navarro, J.Á., Montlleó, M., Rodríguez, G., Lao, J., 2015. GIS-supported mapping of low-temperature geothermal potential taking groundwater flow into account. Renew. Energy 77, 268–278. https://doi.org/10.1016/j.renene.2014.11.096.

Giustiniani, M., Busetti, M., Dal Cin, M., Barison, E., Cimolino, A., Brancatelli, G., Baradello, L., 2022. Geophysical and geological views of potential water resources in the North-Eastern Adriatic Sea. Geosciences 12 (3), 139. https://doi.org/10.3390/geosciences12030139.

Haklidir, F.S.T., Haklidir, M., 2020. Prediction of reservoir temperatures using hydrogeochemical data, Western Anatolia Geothermal systems (Turkey): a machine learning approach. Nat. Resour. Res. 29 (4), 2333–2346. https://doi.org/10.1007/s11053-019-09596-0. SPRINGER.

Hermans, T., Wildemeersch, S., Jamin, P., Orban, P., Brouyère, S., Dassargues, A., Nguyen, F., 2015. Quantitative temperature monitoring of a heat tracing experiment using cross-borehole ERT. Geothermics 53, 14–26. https://doi.org/10.1016/j.geothermics.2014.03.013.

Ibrahim, B.K., Owusu, J., Ahenkorah, I., 2023. Predicting reservoir temperature of geothermal systems in Western Anatolia, Turkey: a focus on predictive performance and explainability of machine learning models. Geothermics 112, 102727. https://doi.org/10.1016/j.geothermics.2023.102727.

Isah, A., Tariq, Z., Mustafa, A., Mahmoud, M., Okoroafor, E.R., 2025. A review of data-driven machine learning applications in reservoir petrophysics. Arab. J. Sci. Eng. 50, 20343–20377. https://doi.org/10.1007/s13369-025-10329-0.

Ishitsuka, K., Lin, W., 2023. Physics-informed neural network for inverse modeling of natural-state geothermal systems. Appl. Energy 337, 120855. https://doi.org/10.1016/j.apenergy.2023.120855. ELSEVIER SCI LTD.

Ishitsuka, K., Ishizu, K., Watanabe, N., Yamaya, Y., Suzuki, A., Bandai, T., Ohta, Y., Mogi, T., Asanuma, H., Kajiwara, T., Sugimoto, T., 2025. Reliable and practical inverse modeling of natural-State geothermal systems using physics-informed neural networks: three-dimensional model construction and assimilation with magnetotelluric data. J. Geophys. Res. Mach. Learn. Comput. 2 (3), e2025JH000683. https://doi.org/10.1029/2025JH000683.

Jospin, L.V., Laga, H., Boussaid, F., Buntine, W., Bennamoun, M., 2022. Hands-on bayesian neural networks—a tutorial for deep learning users. IEEE Comput. Intell. Mag. 17 (2), 29–48. https://doi.org/10.1109/MCI.2022.3155327.

Kendall, A., Gal, Y., 2017. What uncertainties do we need in bayesian deep learning for computer vision? Advances in Neural Information Processing Systems, p. 30.

Khaled, M.S., Chen, D., Ashok, P., Van Oort, E., 2023. Drilling heat maps for active temperature management in geothermal wells. SPE J. 28 (04), 1577–1593. https://doi.org/10.2118/210306-PA.

Kiran, R., Dansena, P., Salehi, S., Rajak, V.K., 2022. Application of machine learning and well log attributes in geothermal drilling. Geothermics 101, 102355. https://doi.org/10.1016/j.geothermics.2022.102355.

Li, Y., Ali, G., Rehman Akbar, A., 2023. Advances in geothermal energy prospectivity mapping research based on machine learning in the age of big data. Sustain. Energy Technol. Assess. 60, 103550. https://doi.org/10.1016/j.seta.2023.103550.

Li, P., Grana, D., Liu, M., 2024. Bayesian neural network and Bayesian physics-informed neural network via variational inference for seismic petrophysical inversion. Geophysics 89 (6), M185–M196. https://doi.org/10.1190/geo2023-0737.1.

Liu, K., Li, Y., Hu, X., Lucu, M., Widanage, W.D., 2020. Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries. IEEE Trans. Ind. Inform. 16 (6), 3767–3777. https://doi.org/10.1109/TII.2019.2941747.

Manfren, M., Gonzalez-Carreon, K.M., James, P.A.B., 2024. Interpretable data-driven methods for building energy modelling—A review of critical connections and gaps. Energies 17 (4), 881. https://doi.org/10.3390/en17040881.

Masoud, A.A., Kubo, T., Koike, K., 2023. Detection of thermal features through interpolation of well-log data in low-to-medium enthalpy geothermal system, Gulf of Suez. Egypt Nat. Resour. Res. 32 (3), 955–980. https://doi.org/10.1007/s11053-023-10160-0.

Mordret, A., 2018. Uncovering the Iceland hot spot track beneath Greenland. J. Geophys. Res. Solid Earth 123 (6), 4922–4941. https://doi.org/10.1029/2017JB015104.

Nielson, D.L., Delahunty, C., Shervais, J.W., 2012. Geothermal systems in the Snake River Plain, Idaho, characterized by the hotspot project. Geotherm. Resour. Counc. Trans. 36, 727–730.

Olivier, A., Shields, M.D., Graham-Brady, L., 2021. Bayesian neural networks for uncertainty quantification in data-driven materials modeling. Comput. Methods Appl. Mech. Eng. 386, 114079. https://doi.org/10.1016/j.cma.2021.114079.

Pasquale, V., Gola, G., Chiozzi, P., Verdoya, M., 2011. Thermophysical properties of the Po Basin rocks: properties of the Po Basin rocks. Geophys. J. Int. 186 (1), 69–81. https://doi.org/10.1111/j.1365-246X.2011.05040.x.

Pérez-Zárate, D., Santoyo, E., Acevedo-Anicasio, A., Díaz-González, L., García-López, C., 2019. Evaluation of artificial neural networks for the prediction of deep reservoir temperatures using the gas-phase composition of geothermal fluids. Comput. Geosci. 129, 49–68. https://doi.org/10.1016/j.cageo.2019.05.004.

Petrini, R., Italiano, F., Ponton, M., Slejko, F.F., Aviani, U., Zini, L., 2013. Geochemistry and isotope geochemistry of the Monfalcone thermal waters (northern Italy): inference on the deep geothermal reservoir. Hydrogeol. J. 21 (6), 1275–1287. https://doi.org/10.1007/s10040-013-1007-y.

Qin, Z., Liu, Y., Zheng, F., Jafarpour, B., 2025. A fluid flow-based deep learning (FFDL) architecture for subsurface flow systems with application to geologic CO2 storage. Water. Resour. Res. 61 (1), e2024WR037953. https://doi.org/10.1029/2024WR037953.

Rodriguez-Gomez, C., Kereszturi, G., Whitehead, M., Reeves, R., Rae, A., Pullanagari, R., 2023. Point pattern analysis of thermal anomalies in geothermal fields and its use for inferring shallow hydrological processes. Geothermics 110, 102664. https://doi.org/10.1016/j.geothermics.2023.102664.

Shahdi, A., Lee, S., Karpatne, A., Nojabaei, B., 2021. Exploratory analysis of machine learning methods in predicting subsurface temperature and geothermal gradient of Northeastern United States. Geotherm. Energy 9 (1), 18. https://doi.org/10.1186/s40517-021-00200-4.

Sheini Dashtgoli, D., Giustiniani, M., Busetti, M., Cherubini, C., 2024a. Artificial intelligence applications for accurate geothermal temperature prediction in the lower Friulian Plain (north-eastern Italy). In: Journal of Cleaner Production, 460. ELSEVIER SCI LTD. https://doi.org/10.1016/j.jclepro.2024.142452.

Sheini Dashtgoli, D., Giustiniani, M., Busetti, M., Cherubini, C., Alessandrini, G., Narsilio, G.A., 2026a. Artificial Intelligence in deep geothermal energy: trends, insights, and future. Renew. Energy Focus 56, 100781. https://doi.org/10.1016/j.ref.2025.100781.

Sheini Dashtgoli, D., Giustiniani, M., Busetti, M., Cherubini, C., Narsilio, G., 2026b. Long-term hydrothermal evolution and sustainable operation of a deep carbonate geothermal reservoir: a field-calibrated 3D framework for the lower Friulian Plain. Italy Energy Convers. Manag. 357, 121377. https://doi.org/10.1016/j.enconman.2026.121377.

Sheini Dashtgoli, D., Narsilio, G.A., Alqawasmeh, Q.I., Giustiniani, M., Busetti, M., Cherubini, C., 2026c. Artificial intelligence for shallow geothermal systems: a review of ground heat exchanger performance modeling. J. Rock Mech. Geotech. Eng., S1674775526002878 https://doi.org/10.1016/j.jrmge.2026.03.025.

Shima, A., Ishitsuka, K., Lin, W., Bjarkason, E.K., Suzuki, A., 2024. Modeling unobserved geothermal structures using a physics-informed neural network with transfer learning of prior knowledge. Geotherm. Energy 12 (1), 38. https://doi.org/10.1186/s40517-024-00312-7.

Šušmelj, K., Čenčur Curk, B., Kanduč, T., Rožič, B., Verbovšek, T., Vreča, P., Žagar, K., Žvab Rožič, P., 2024. Hydrogeochemical conditions of submarine and terrestrial karst sulfur springs in the Northern Adriatic. Environ. Earth Sci. 83 (7), 214. https://doi.org/10.1007/s12665-024-11476-7.

Della Vedova, B., Cimolino, A., Castelli, E., Brancatelli, G., 2015. Geothermal heating and cooling in the FVG region: the Grado district heating and the Pontebba ice rink plants. Status and future in the Peri–Adriatic Area–Veli Lošinj (Croatia). In: Proceedings of the Workshop on Geothermal Energy.

Velić, I., Tišljar, J., Matičec, D., Vlahović, I., 2000. Introduzione alla geologia dell'Istria. Guida Alla Escursioni-80a Riun. Estiva 237–245.

Vogt, C., Mottaghy, D., Wolf, A., Rath, V., Pechnig, R., Clauser, C., 2010. Reducing temperature uncertainties by stochastic geothermal reservoir modelling. Geophys. J. Int. 181 (1), 321–333. https://doi.org/10.1111/j.1365-246X.2009.04498.x.

Wang, L., Yu, Z., Zhang, Y., Yao, P., 2023a. Review of machine learning methods applied to enhanced geothermal systems. Environ. Earth Sci. 82 (3), 69. https://doi.org/10.1007/s12665-023-10749-x.

Wang, L., Yu, Z., Zhang, Y., Yao, P., 2023b. Review of machine learning methods applied to enhanced geothermal systems. Environ. Earth Sci. 82 (3), 69. https://doi.org/10.1007/s12665-023-10749-x.

Wang, H., 2008. Coverage probability of prediction intervals for discrete random variables. Comput. Stat. Data Anal. 53 (1), 17–26. https://doi.org/10.1016/j.csda.2008.07.017.

Yang, Y., Jia, M., 2024. 3D spatial interpolation of soil heavy metals by combining kriging with depth function trend model. J. Hazard. Mater. 461, 132571. https://doi.org/10.1016/j.jhazmat.2023.132571.

Yuan, Y., Li, W., Bian, L., Lei, J., 2024. A prediction model for pressure and temperature in geothermal drilling based on physics-informed neural networks. Electron 13 (19), 3869. https://doi.org/10.3390/electronics13193869.

Zhang, P., Yin, Z.-Y., Jin, Y.-F., 2022. Bayesian neural network-based uncertainty modelling: application to soil compressibility and undrained shear strength prediction. Can. Geotech. J. 59 (4), 546–557. https://doi.org/10.1139/cgj-2020-0751.

Zheng, S.-F., Li, X., Wang, M., 2025. Multivariate prediction model of geothermal parameters based on machine learning. Energy 316, 134497. https://doi.org/10.1016/j.energy.2025.134497.

Zini, L., Calligaris, C., Treu, F., Zavagno, E., Iervolino, D., Lippi, F., 2023. Groundwater sustainability in the Friuli Plain. Acque. Sotter - Ital. J. Groundw. 12 (3), 77–92. https://doi.org/10.7343/as-2023-713.