

Reconstructing Soil Temperature Time Series with Prolonged Gaps Across Depths: An Interpretable LSTM-Based Approach

Leticia Barros Dias Soares

`leticia.soares@ufpel.edu.br`

Universidade Federal de Pelotas

Natássia Siqueira Medeiros

Universidade Federal de Pelotas

João Inácio Moreira Bezerra

Universidade Federal de Pelotas

Geilson de Almeida Soares

Universidade Federal de Pelotas

Jairo Valões de Alencar Ramalho

Universidade Federal de Pelotas

Marlon Maurício Hernandez Cely

Universidade Federal de Pelotas

Honório Joaquim Fernando

Universidade Federal Fluminense

Research Article

Keywords: Soil temperature, Time series imputation, LSTM, Explainable AI, Environmental monitoring

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10919304/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Reconstructing Soil Temperature Time Series with Prolonged Gaps Across Depths: An Interpretable LSTM-Based Approach

Leticia Barros Dias Soares^{1*}, Natássia Siqueira Medeiros^{1†},
João Inácio Moreira Bezerra^{1†}, Geilson de Almeida Soares^{1†},
Jairo Valões de Alencar Ramalho^{1†}, Marlon Maurício Hernandez Cely^{1†},
Honório Joaquim Fernando^{2†}

¹ Universidade Federal de Pelotas (UFPel), Rua Gomes Carneiro, 1, Pelotas, 96010-610, Rio Grande do Sul, Brazil.

² Universidade Federal Fluminense, Campus Aterrado (UFF), Rua Desembargador Ellis Hermydio Figueira, 783, Volta Redonda, 27213-145, Rio de Janeiro, Brazil.

*Corresponding author(s). E-mail(s): leticia.soares@ufpel.edu.br;

Contributing authors: nrmsiqueira@inf.ufpel.edu.br; jimbezerra@inf.ufpel.edu.br;
geilson.soares@ufpel.edu.br; jairo.ramalho@ufpel.edu.br; marlon.cely@ufpel.edu.br;
honoriofernando@id.uff.br;

[†]These authors contributed equally to this work.

Abstract

Shallow geothermal energy remains underexplored in Uruguay despite high soil thermal potential. Assessing this requires continuous *in situ* monitoring, but prolonged missing data across soil depths compromises system modeling. Addressing this operational challenge, we propose a data-driven framework to reconstruct missing segments in soil temperature series using LSTM networks, Optuna hyperparameter optimization, and SHAP/TimeSHAP explainability. Using primary data from a monitoring station in Rivera, Uruguay, our pipeline integrates local sensor measurements with ERA5 global reanalysis data and temporal attributes across superficial, intermediate, and deep soil layers. Results show optimization substantially improved reconstruction in superficial and intermediate layers, reducing mean absolute errors from 0.16 to 0.06 and 0.29 to 0.07, respectively. The deep layer exhibited a trade-off between global trajectory adherence and pointwise calibration. Crucially, explainability analysis revealed depth-dependent model behavior, bridging predictive accuracy and physical interpretability by showing how the model relies on temporal and reference-derived attributes. This study demonstrates that LSTM-based imputation, when evaluated jointly with explainable AI, provides a reliable computational tool for recovering real-world environmental time series.

Keywords: Soil temperature, Time series imputation, LSTM, Explainable AI, Environmental monitoring

1 Introduction

Technological advances have played a significant role in promoting energy efficiency and addressing contemporary climate challenges. Although innovation contributes to the mitigation of carbon emissions, part of these gains may be offset by the so-called rebound effect, in which the expansion of economic activity and the intensive use of digital technologies, such as data centers and artificial intelligence applications, increase total energy demand [de Vries \(2023\)](#). In a context of accelerated global warming and recurrent thermal extremes [Brazil \(2024\)](#), the transition to clean energy sources and the improvement of low-energy-consumption technologies have become increasingly necessary.

Among renewable alternatives, geothermal energy stands out for exploiting heat from the subsurface. One application of interest is the Earth-to-Air Heat Exchanger (EAHE) system. This system uses low-power fans to promote air circulation through buried ducts, taking advantage of the relative thermal stability of

the soil to cool or heat the air supplied to buildings [Liu et al. \(2023\)](#). Owing to their low energy consumption and potential integration with passive or hybrid climate-control strategies, EAHE systems have received increasing attention in performance, modeling, and optimization studies [Imanloozadeh and Lubitz \(2026\)](#); [Soares et al. \(2025\)](#); [Domingues et al. \(2026\)](#); [Ramalho et al. \(2022\)](#); [Domingues et al. \(2021\)](#).

Despite these advances, the thermal analysis of EAHE systems and the underground environment still depends heavily on data availability and quality. In practice, particularly in continuous field monitoring stations, these time series are frequently affected by sensor failures, communication losses, equipment maintenance, or power supply interruptions. Such operational realities result in observational gaps that compromise the characterization of subsurface thermal behavior and, consequently, the modeling, forecasting, and performance assessment of these systems [Perlaky et al. \(2025\)](#); [Alejo-Sanchez et al. \(2025\)](#). Recent studies have shown that deep learning algorithms can support the modeling of soil-related environmental variables by integrating geo-environmental information and remote sensing indicators [Nasiri Khiavi \(2025\)](#), reinforcing the relevance of auxiliary environmental data sources in data-driven soil modeling.

In this context, data imputation and reconstruction are no longer treated merely as preprocessing steps but have instead been recognized as central problems in climatic and environmental time series [Kumar and Varija \(2024\)](#); [Park et al. \(2023\)](#); [Pastorini et al. \(2024\)](#). This challenge is particularly critical when gaps are long, because the local information from the series itself is no longer sufficient to restore the lost temporal dynamics. Kumar and Varija (2024), for example, show that the reconstruction of prolonged gaps in thermal series requires strategies different from those used for isolated missing values, with greater dependence on auxiliary sources and spatial or multivariate relationships [Kumar and Varija \(2024\)](#). Similarly, [Park et al. \(2023\)](#) demonstrate that deep learning-based approaches can outperform traditional gap-filling methods in time series with continuous missing segments [Park et al. \(2023\)](#).

Recent literature shows that different strategies have been employed to reconstruct time series with missing data, including statistical methods, classical machine learning techniques, and approaches based on deep neural networks [Kumar and Varija \(2024\)](#); [Park et al. \(2023\)](#); [Alejo-Sanchez et al. \(2025\)](#). Among these alternatives, recurrent models, especially those based on *Long Short-Term Memory* (LSTM), have received increasing attention due to their ability to capture short- and long-term temporal dependencies in sequential data, a particularly important feature in thermal time series reconstruction problems [Park et al. \(2023\)](#). Recent studies have also adopted LSTM-based architectures in environmental and energy-related time series, including evapotranspiration estimation supported by remote sensing inputs and short-term wind speed prediction combined with feature extraction and hyperparameter optimization [Hashemi et al. \(2025\)](#); [Tian et al. \(2025\)](#). In soil temperature series, this potential is even more relevant, since the observed temporal structure results from processes involving memory, damping, and time lag among different depths.

Despite the progress of these approaches, the evaluation of reconstruction models in environmental contexts still tends to focus predominantly on predictive performance metrics. Although this aspect is important, it is not sufficient to ensure that the model is learning relationships consistent with the dynamics of the phenomenon under analysis [Pastorini et al. \(2024\)](#); [Santamaría-Bonfil et al. \(2022\)](#); [Adekanmbi and Sizmur \(2022\)](#). Thus, beyond accuracy, it becomes important to investigate how the model mobilizes the available information to estimate missing values, opening space for discussions on transparency, interpretability, and the reliability of the reconstructions produced.

This issue becomes even more critical in the context of deep neural network (DNN) models, which, despite their high predictive capacity, still pose interpretability as a central challenge [Park et al. \(2023\)](#); [Lundberg and Lee \(2017\)](#). Due to these obstacles, the black-box problem emerges [Ali et al. \(2023\)](#). However, simpler machine learning models, such as decision trees, tend to provide greater interpretability, since their operating logic can be understood directly from their structure and parameters, without requiring external explanation mechanisms. In this sense, such approaches are frequently characterized as white-box models, in contrast to black-box models, whose decision-making process is less transparent [Ali et al. \(2023\)](#).

In the literature, this distinction can be understood based on different levels of internal transparency, which usually classify machine learning models into white-box, gray-box, and black-box approaches [Ali et al. \(2023\)](#). From this perspective, Figure 1 summarizes these categories according to the degree of model interpretability and the need for external explanation mechanisms.

	Transparency Level	Practical Applicability	Role of XAI
White-box 	Full (Self-interpretable) Lower in complex non-linear problems	Limited for very complex data, but essential for regulatory requirements	Not necessary.
Gray-box 	Partial Typical Accuracy Significant	Feasible in critical applications, provided it is rigorously designed.	Optional or complementary.
black-box 	None (Opaque) Typical Accuracy Hight	Impractical in critical domains without the use of auxiliary explanatory tools.	Essential for revealing decision logics, building trust, and ensuring transparency.

Fig. 1 Levels of Transparency and Interpretability in Machine Learning Models

Above all, Explainable Artificial Intelligence (XAI) provides useful tools for enhancing the interpretability of complex models [de la Peña et al. \(2025\)](#); [Zhan and Chen \(2025\)](#). Among these approaches, SHapley Additive exPlanations (SHAP) enables the contribution of input variables to the model output to be quantified at both global and local levels, supporting the identification of the patterns that most strongly influence the estimates produced [Lim et al. \(2021\)](#); [Geng et al. \(2024\)](#); [Alizamir et al. \(2024\)](#). In time series problems, this characteristic is particularly valuable, as it makes it possible to investigate which variables and which time steps within the temporal window exert the greatest influence on reconstruction, thereby providing evidence on the model's internal logic and on the plausibility of the relationships learned.

Although there is growing interest in the use of deep learning and XAI techniques for environmental data, studies that integrate the reconstruction of soil temperature time series at multiple depths with explainability-based interpretation remain limited [Kumar et al. \(2024\)](#); [Pastorini et al. \(2024\)](#); [Alizamir et al. \(2024\)](#); [Mampitiya et al. \(2024\)](#). This gap is relevant because, in applications involving data from different layers, it is not sufficient to recover missing observations with good accuracy; it is also necessary to examine whether the model coherently exploits the temporal dependencies and inter-depth relationships that characterize soil thermal dynamics. In this sense, the combination of LSTM and SHAP constitutes a promising strategy for bringing together predictive capacity and interpretability within the same analytical framework.

Given these challenges, this study proposes the reconstruction of missing data in experimental soil temperature time series measured at different layers. Rather than relying on synthetic gaps in benchmark datasets, this research addresses real-world sensor failures from an environmental monitoring station in Rivera, Uruguay, deployed to assess local shallow geothermal potential. Using an LSTM network as the predictive model and the SHAP method as an explainability tool, the analysis is conducted at two complementary scales: global, to identify the most relevant patterns and attributes for the model's overall behavior; and local, to interpret specific reconstruction cases. Thus, the contribution of this work is not limited to obtaining imputed estimates, but also includes an investigation of how the model uses temporal and inter-layer information to produce these estimates. In doing so, this study aims to provide an approach that combines thermal time series reconstruction with greater interpretive transparency, contributing to the more reliable use of deep learning-based models in operational applications involving soil temperature.

2 Related works

2.1 Reconstruction of soil temperature series with gaps

Soil temperature reconstruction has been widely investigated due to its relevance in agricultural, environmental, hydrological, and geothermal applications. In particular, modeling at different depths is important because soil thermal dynamics combine a shared seasonal behavior with depth-specific attenuation and phase-lag effects along the vertical profile [Khosravi et al. \(2023\)](#); [Alizamir et al. \(2024\)](#). In contexts related to geothermal energy, Earth-to-Air Heat Exchanger (EAHE) systems, generate soil temperature time series data that support thermal analyses, sizing procedures, and operational assessments [Imanloozadeh and Lubitz \(2026\)](#); [Perlaky et al. \(2025\)](#).

In climate-related applications, the literature shows that reconstruction performance strongly depends on the domain and on the missingness pattern. Santamaría-Bonfil et al. [Santamaría-Bonfil et al. \(2022\)](#) proposed an equivalent imputation methodology to reconstruct a geochemical database with missing data, showing that structured methods may still be competitive in specific scientific scenarios. In another direction, Niako et al. [Niako et al. \(2024\)](#) investigated how different imputation methods affect the analysis and forecasting of univariate blood pressure time series using an LSTM model, demonstrating that the choice of reconstruction strategy directly affects subsequent predictive performance. Therefore, reconstruction should not be treated merely as a preliminary data-processing step, but as a methodological choice that may directly influence the behavior and performance of the final model [Niako et al. \(2024\)](#); [Alejo-Sanchez et al. \(2025\)](#). Beyond imputation-specific studies, LSTM-based architectures have also been adopted in environmental and energy-related time series modeling, especially when combined with feature extraction and hyperparameter optimization strategies [Tian et al. \(2025\)](#).

This challenge has driven the development of more sophisticated methods, especially within deep learning models. Particularly, models based on Recurrent Neural Networks (RNNs) and variants with attention mechanisms have been increasingly employed in temporal imputation tasks [Du et al. \(2023\)](#); [Lai et al. \(2025\)](#). LSTM networks remain particularly attractive because they can model short- and medium-term temporal memory with relatively stable training, making them suitable for temporal reconstruction tasks [Park et al. \(2023\)](#); [Du et al. \(2023\)](#). In a more advanced line of research, Lai et al. [Lai et al. \(2025\)](#) proposed Ratai, a recurrent autoencoder with imputation units and temporal attention for multivariate time series. Complementarily, Zhao et al. [Zhao et al. \(2023\)](#) presented a multivariate imputation method based on generative adversarial networks with attention mechanisms, showing gains in modeling dependencies between variables and time steps. However, the adoption of these more sophisticated architectures does not eliminate the relevance of LSTM networks, especially in contexts where experimental stability and methodological design are valued.

Furthermore, these discussions become particularly important in scientific applications in which it is not sufficient to reconstruct accurately; it is also necessary to understand *why* the model produced a given estimate. In this context, SHAP has become one of the most widely used post-*hoc* explanation approaches, as it provides attributions based on Shapley value theory [Lundberg and Lee \(2017\)](#); [Bento et al. \(2021\)](#). The official documentation of the method describes the *KernelExplainer* [SHAP Developers \(2026b\)](#); [Feedzai \(2026\)](#) as a model-agnostic strategy based on locally weighted linear regression to estimate additive feature contributions. This characteristic makes it particularly useful in time series, where the goal is to interpret both the global importance of attributes and the local contribution of specific temporal windows [Alizamir et al. \(2024\)](#).

The integration between LSTM and explainability methods is still relatively recent in the literature. Zhan and Chen [Zhan and Chen \(2025\)](#) discuss the combination of SHAP and LIME in LSTM-based models for time series, arguing that global and local explanations are complementary for making recurrent models more understandable [Lundberg and Lee \(2017\)](#). More broadly, this movement follows the growing demand for temporal models that are not only accurate but also transparent. However, few studies have articulated, within a single experimental pipeline, LSTM-based imputation, systematic hyperparameter optimization, and global and local explainability oriented toward the reconstruction process.

Another underexplored point is the relationship between hyperparameter tuning and interpretability. In recurrent models, hyperparameters such as temporal window size, number of hidden units, *dropout*, learning rate, and batch size may significantly alter not only predictive performance, but also the pattern of importance assigned to variables and time steps. In this context, optimization frameworks such as Optuna provide a systematic and reproducible way to search for hyperparameters in black-box optimization problems [Optuna Project \(2026\)](#). Libraries such as NeuralForecast, in turn, illustrate the consolidation of ecosystems dedicated to neural time series modeling with an emphasis on modularity and reproducibility [Nixtla \(2026\)](#). Even so, the combination of LSTM-based imputation, hyperparameter optimization, and SHAP-based post-*hoc* interpretation remains insufficiently explored, especially in the context of reconstructing temperature series with prolonged gaps.

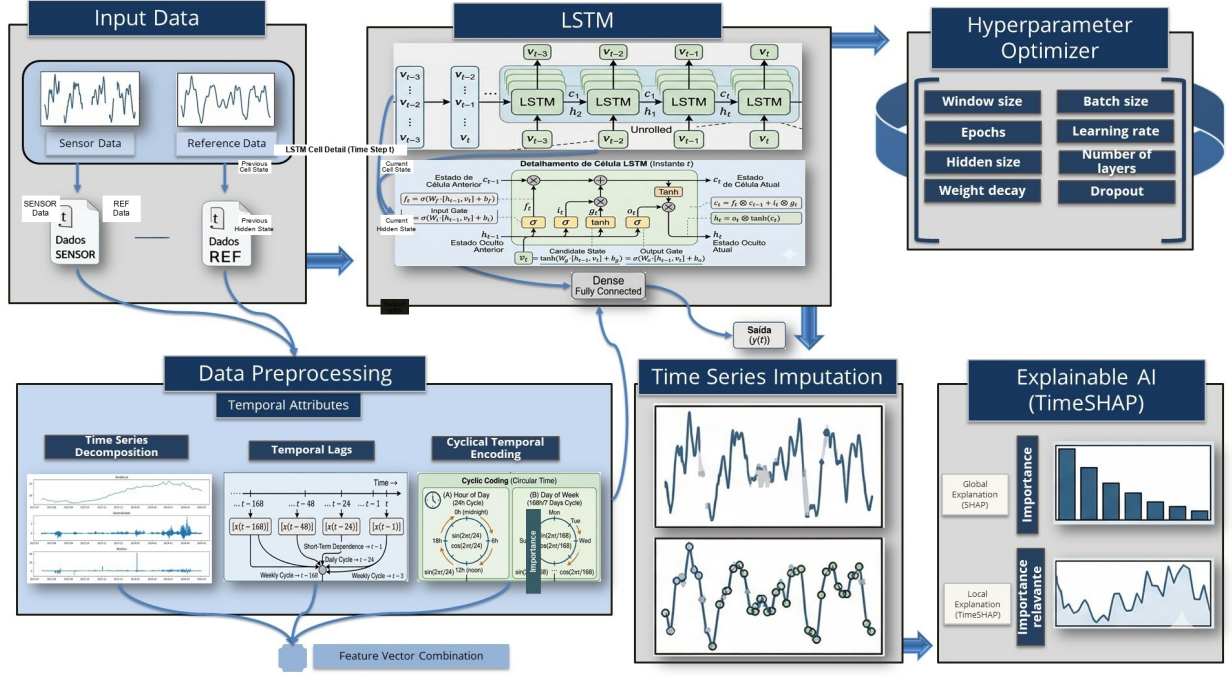


Fig. 2 Preprocessing, optimization, and temporal data imputation pipeline using LSTM.

Despite the advances observed, an important gap remains in the literature. Most studies focus on only one family of models or on pure forecasting problems [Kan et al. \(2026\)](#); [Alizamir et al. \(2024\)](#); [Khosravi et al. \(2023\)](#). The main gap identified is not merely the proposal of a new imputation model, but the integration of three components that are usually treated separately: (i) temporal reconstruction of missing data with LSTM; (ii) systematic hyperparameter optimization to avoid arbitrary architectural choices; and (iii) global and local explainability to clarify which variables and which temporal positions support the reconstruction. This is precisely the gap that the present study seeks to address.

3 Methodology

The methodology of this study was organized as a sequential workflow that integrates data preparation, predictive modeling, hyperparameter optimization, time series reconstruction, and explainability analysis. This sequence was adopted to ensure that the data reconstruction process was conducted systematically, from the structuring of the model inputs to the interpretation of the results presented. Figure 2 presents an overview of the modeling system adopted in this study.

3.1 Experimental data acquisition and quality control

The primary data used in this study stems from an *in situ* soil thermal monitoring system deployed at UTEC in Rivera, Uruguay [Barboza Pereira \(2024\)](#). The system operated continuously from August 1, 2023, to July 31, 2024, recording data approximately every 20 minutes. Data acquisition was performed using Arduino Uno and ESP8266 microcontrollers communicating via the MQTT protocol, with measurement circuits systematically activated by relays to prevent sensor self-heating.

The experimental setup consisted of three redundant measurement points spaced 3 meters apart. Each point was instrumented with five platinum resistance thermometers (PT100) inserted into polypropylene random (PPR) tubes, allowing thermal profiling at depths of 0.4 m, 0.8 m, 1.2 m, 1.6 m, and 2.0 m. Additionally, a surface SHT-30 sensor recorded air temperature and relative humidity near the ground, protected from direct solar radiation and rain. Figure 3 shows different pictures of the experimental setup.

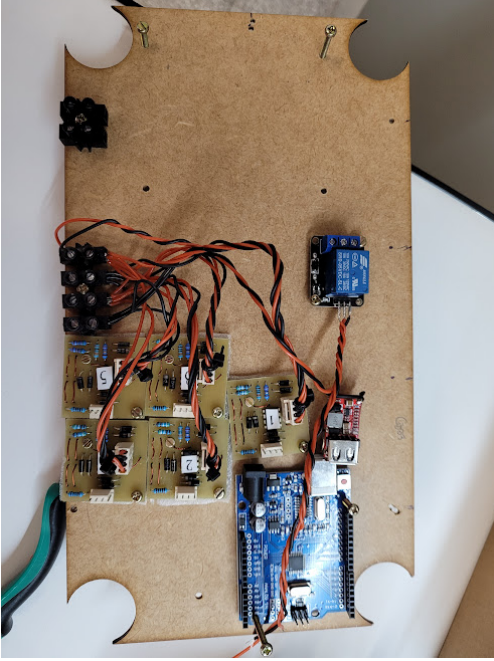
To ensure the physical and statistical consistency of the time series prior to thermal modeling, the raw data from the acquisition system underwent a strict sequential quality control protocol. First, a physical limits filter removed instrumental anomalies by restricting the data to a climatically viable range for the study region (-10°C to 50°C). Second, a dynamic consistency algorithm was applied to identify and remove outliers incompatible with soil thermal inertia. Each reading was compared to a 4-hour moving average, with a maximum allowable variation threshold (ΔT_{max}) adjusted according to depth: 6°C for the surface air and 0.4 m; 4.5°C for 0.8 m and 1.2 m; and 3°C for 1.6 m and 2.0 m. Finally, the validated records were



(a)



(b)



(c)



(d)

Fig. 3 Hardware implementation and field deployment of the soil thermal monitoring system. (a) Field installation highlighting the spatial distribution of the three redundant measurement points. (b) PT100 sensors mounted on PPR tubes for vertical thermal profiling. (c) Custom data acquisition circuitry, including microcontrollers and relay modules. (d) Protective outdoor enclosure housing the local acquisition unit and electrical components.

converted into hourly averages to attenuate high-frequency instrumental noise, and the series from the three redundant points were synchronized and consolidated into a single unified dataset.

3.2 Dataset and preprocessing

The experimental formulation considered the three representative layers of the soil thermal profile as the units of analysis: surface, intermediate, and deep layers. In each of them, imputation was treated as a supervised temporal reconstruction problem, in which the models used historical information from the target series itself, exogenous variables derived from ERA5, and temporal attributes to estimate missing values. In the baseline study, this process was implemented through feature engineering based on STL decomposition, cyclic temporal encoding, and relevant lags of 24 h and 168 h, while fully preserving the observed values and applying the model only to the gaps. A minimum threshold was also adopted for gaps with block durations of $H_{min} > 48$. Additionally, the article considers critical long-duration scenarios, with gaps of up to 552 hours, as they represent particularly challenging situations for reconstruction at different depths.

The dataset used in this study comprised two sources: (i) local soil temperature measurements obtained from sensors installed at UTEC, in Rivera, Uruguay, and (ii) ERA5 climate reanalysis data, used as an external reference to support the temporal continuity of the analysis and the construction of target variables during periods when local data were unavailable [Barboza Pereira \(2024\)](#); [Hersbach et al. \(2020\)](#); [ECMWF \(2026\)](#).

In the local dataset, 6,056 valid records were preserved within a total hourly grid of 8,784 time steps, corresponding to 68.94% of the series. Consequently, 2,728 time steps, equivalent to 31.06%, presented observational absence and constituted the potential reconstruction universe. The complementary ERA5

dataset was used because it provides continuous and consistent mean values for soil temperature layers, serving as support for temporal analysis and for the construction of exogenous attributes [Hersbach et al. \(2020\)](#); [ECMWF \(2026\)](#).

The target variable of this study is soil temperature, organized into three representative layers of the vertical profile. For this purpose, the local sensors were aggregated by depth, resulting in three series: surface layer, intermediate layer, and deep layer. To avoid ambiguity, this study adopts the ERA5 nomenclature, in which soil temperature layers are identified by the acronym STL (soil temperature level). Thus, the surface layer corresponds to the sensor located at 0.0 m and was associated with ERA5 STL1 (0–7 cm); the intermediate layer was obtained by averaging the sensors at 0.4 m and 0.8 m, approximating STL3 (28–100 cm); and the deep layer was constructed from the average of the sensors at 1.2 m, 1.6 m, and 2.0 m, approximating STL4 (100–289 cm). This strategy made it possible to render the local series comparable to the depths available in the reanalysis and to establish an approximate layer-wise correspondence between the two data sources [Hersbach et al. \(2020\)](#); [Jacobs et al. \(2024\)](#).

Temporal resolution involved the chronological ordering of the data, consistency checking, treatment of inconsistent records, and explicit identification of missing time steps, distinguishing effectively observed records from the complete temporal grid. This step was necessary to preserve the original temporal structure of the series and to adequately characterize prolonged failure segments, in line with good practices in time series modeling and with recent approaches that emphasize the importance of explicitly representing trend and seasonality in forecasting and reconstruction problems [Hyndman and Athanasopoulos \(2021\)](#); [Hao et al. \(2024\)](#).

In addition, the compatibility between the aggregated local series and the corresponding ERA5 layers was evaluated using Pearson correlation and simple linear regression, considering only the time steps in which both sources were simultaneously available [Yilmaz \(2023\)](#); [Schönauer et al. \(2025a\)](#); [Silva et al. \(2025b\)](#). This step made it possible to verify whether the consolidated series preserved the regional thermal behavior represented by the reanalysis. This procedure is consistent with recent literature, which uses ERA5 both to assess the regional climatic representativeness of observations and to support gap filling in temperature series [Bodjrenou et al. \(2023\)](#); [Jacobs et al. \(2024\)](#). The resulting series then served as the basis for the subsequent feature construction stage.

For each target layer, temporal windows were constructed using lagged observations from the series itself and from the exogenous variables, in order to capture short- and medium-term temporal dependencies. Derived temporal attributes, such as seasonal markers and cyclic variables, were also included to explicitly represent intra-daily and calendar patterns [Hyndman and Athanasopoulos \(2021\)](#).

For the treatment of seasonality, Seasonal and Trend decomposition using Loess (STL) was adopted as a strategy to separate trend, seasonality, and residual components, preserving the hourly dynamics of soil temperature and supporting the physically coherent reconstruction of long gaps [Cleveland et al. \(1990\)](#); [Wu et al. \(2024\)](#). This choice is particularly relevant because soil thermal variation presents a more pronounced intra-daily cycle in shallow layers, whereas in deeper layers the thermal amplitude tends to be damped and the temporal response delayed [Adekanmbi and Sizmur \(2022\)](#); [Lembrechts et al. \(2022\)](#). Consequently, preprocessing sought not only to reduce noise but also to make the temporal structure of the process explicit, rendering it usable by statistical, machine learning, and deep learning models.

3.3 Feature engineering

The preprocessing step aimed to preserve the temporal coherence of the series and to make patterns useful for the reconstruction process explicit. Initially, the local and reference series were aligned on the same temporal grid. Then, chronological ordering, consistency checking, and explicit missing-value marking routines were applied.

The attributes were organized into four groups. First, cyclic attributes were constructed to encode daily and weekly patterns through sine and cosine transformations, a strategy used to represent temporal periodicity continuously in recent predictive modeling applications [Durap \(2025\)](#); [Zhao et al. \(2025\)](#). The transformations applied were:

- Daily cycle:

$$\sin\left(\frac{2\pi t}{24}\right) \quad \text{and} \quad \cos\left(\frac{2\pi t}{24}\right), \quad (1)$$

- Weekly cycle:

$$\sin\left(\frac{2\pi t}{168}\right) \quad \text{and} \quad \cos\left(\frac{2\pi t}{168}\right), \quad (2)$$

where the time t is given in hours.

Second, the reference series was decomposed using STL into trend, seasonality, and residual components in order to make the structural components of thermal behavior explicit [Alejo-Sanchez et al. \(2025\)](#).

$$x^{ref}(t) = T(t) + S(t) + R(t), \quad (3)$$

where $T(t)$ is the trend, $S(t)$ is the seasonal component with a 24-hour period, and $R(t)$ is the stochastic residual. Third, temporal lags of the reference series selected through autocorrelation were added, including short-term and seasonal lags. Fourth, auxiliary attributes were constructed for the imputation process, including an interpolated version of the observed series and a missingness indicator variable.

After this step, the samples were converted into sliding temporal windows. For a window of size W , each sample was represented by a tensor $\mathbf{X}_t \in \mathbb{R}^{W \times F}$, where F denotes the number of attributes. The target variable was defined as the reference value at the final time step of the window.

Consider an input sequence $\mathbf{X}_t = \{\mathbf{x}_{t-W+1}, \dots, \mathbf{x}_t\}$, where each vector $\mathbf{x}_\tau$ contains temporal attributes, STL components, lags, and auxiliary reconstruction variables. The model learns a parametric function f_θ such that

$$\hat{y}_t = f_\theta(\mathbf{X}_t), \quad (4)$$

where $\hat{y}_t$ represents the reconstructed value at time t . Final imputation is performed only at the time steps that were originally missing, while the observed values are fully preserved. This formulation avoids altering valid segments of the series and restricts the model intervention exclusively to the gaps.

3.4 LSTM modeling and hyperparameter optimization

Temporal reconstruction was performed using an LSTM network followed by a linear scalar output layer. The architecture receives as input a multivariate sequence in the format (N, W, F) , where N is the number of samples, W is the window size, and F is the number of attributes. The last hidden state of the LSTM is projected through a dense layer to generate the final estimate.

The 8,569 sequential instances were chronologically split into:

- 5,826 for training (approximately 68.0%)
- 1,029 for validation (approximately 12.0%)
- 1,714 for testing (approximately 20.0%)

The inputs and the target variable were standardized based exclusively on the training set. Training was conducted by minimizing the mean squared error using the Adam optimizer. The final model state was selected according to the lowest loss on the validation set. The choice of LSTM is justified by its ability to model nonlinear temporal dependencies and to simultaneously incorporate historical information, seasonality, and exogenous attributes in reconstruction tasks [Niako et al. \(2024\)](#); [Zhan and Chen \(2025\)](#).

After training the baseline model, a hyperparameter optimization step was performed using the Optuna framework, originally introduced by Akiba et al. [Akiba et al. \(2019a\)](#). The choice of this optimizer was also supported by recent evidence in temporal modeling with deep networks, in which Optuna-based strategies have been successfully applied to the tuning of LSTM models for time series forecasting [Hanifi et al. \(2024\)](#). In addition, the use of hyperparameter optimization is consistent with recent Earth science applications in which LSTM performance was improved by tuning learning rate, regularization terms, and hidden-layer parameters [Tian et al. \(2025\)](#). For implementation purposes, the official repository [Optuna Project \(2026\)](#) was used as an initial basis; however, the code was adapted to the present study.

In the optimization process, the search space included the temporal window size, batch size, learning rate, number of hidden units, number of recurrent layers, *dropout* rate, and *weight decay* regularization term. This step was motivated by the fact that, in recurrent networks, small changes in these parameters can modify training stability and the way the model distributes importance across the temporal window.

3.5 Global and local explainability

The explainability stage was based on the SHAP framework, originally proposed by Lundberg and Lee [Lundberg and Lee \(2017\)](#), and adapted to the sequential context based on TimeSHAP, a method specifically developed for recurrent models [Bento et al. \(2021\)](#). The official repositories of the SHAP and TimeSHAP libraries were consulted to support implementation and experimental adaptation to the problem investigated in this study [SHAP Contributors \(2026a\)](#); [Feedzai \(2026\)](#).

The interpretation of the final model was conducted at two levels. At the global level, SHAP-style aggregate contributions were computed to quantify: (i) the mean absolute importance of each variable; (ii) the mean absolute importance of each temporal position within the window; and (iii) the time-by-variable

matrix, allowing visualization of where in the sequence each attribute exerts the greatest influence [SHAP Developers \(2026b\)](#). This analysis was performed on a representative subset of the test sequences.

At the local level, TimeSHAP was used as a methodological reference to explain individual predictions in recurrent models [Zhan and Chen \(2025\)](#). Three types of cases were selected: one example with low error, one example with high error, and one example belonging to a long missing-data block. For each case, local explanations were generated by temporal event and by variable. This strategy made it possible to interpret not only the average behavior of the model, but also the specific mechanisms associated with individual reconstructions.

3.6 Evaluation metrics

Performance was quantified using four metrics: mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and Pearson correlation (r). Formally,

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (7)$$

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}. \quad (8)$$

These metrics were computed on the test set and used both to compare the baseline model with the optimized model and to evaluate reconstruction quality across different layers of the thermal profile.

4 Results and discussion

4.1 Compatibility Performance Assessment: Sensor Data and Reference Data

The compatibility between the reference series and the ERA5 reanalysis was evaluated through residual analysis (ε_t), defined as the difference between the observed values in the reference series (y_t) and the corresponding values obtained from ERA5 ($\hat{y}_t$):

$$\varepsilon_t = y_t - \hat{y}_t, \quad (9)$$

Based on these residuals, the degree of agreement between the series was analyzed using MAE, RMSE, the coefficient of determination (R^2), and Pearson’s linear correlation coefficient (r), calculated between (y_t) and ($\hat{y}_t$). Together, these metrics make it possible to evaluate both the magnitude of the deviations and the linear association between the observed data and the reanalysis data.

Table 1 Compatibility performance metrics between the reference series and the ERA5 soil temperature layers (STL).

Layer	MAE (°C)	RMSE (°C)	R^2	r (Pearson)
Surface (S)	1,06	1,52	0,93	0,97
Intermediate (M)	0,46	0,66	0,97	0,99
Deep (P)	0,68	0,94	0,84	0,95

The results presented in Table 1 indicate very strong linear associations ($r > 0.94$) and a high explanatory capacity for data variability ($R^2 \geq 0.84$) across the three analyzed layers. These results confirm that the reference series preserves regional climatic consistency, validating its use as a basis for comparison and training in the subsequent imputation steps. As illustrated in Figures 4, 5, and 6, the series are shown for the different layers.

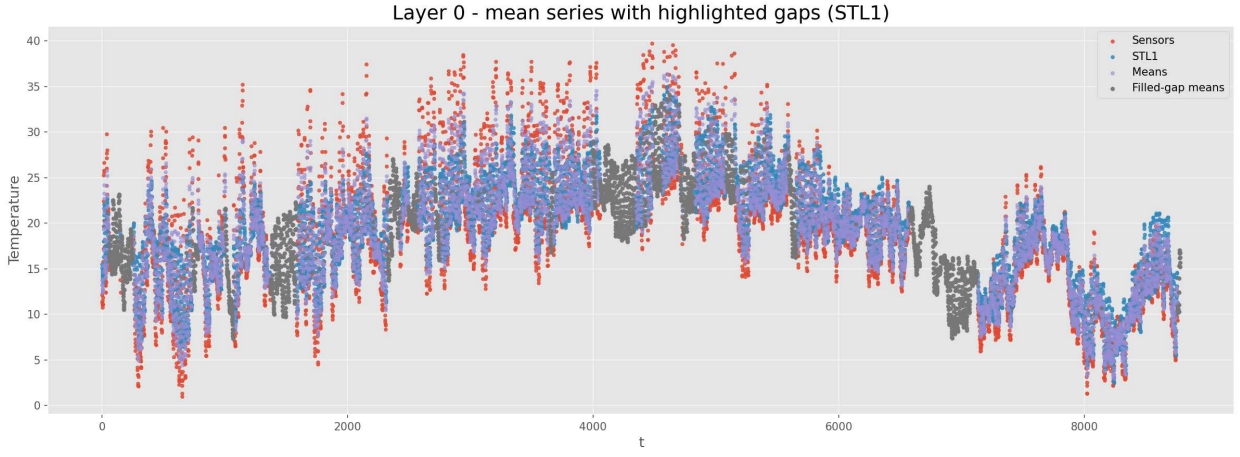


Fig. 4 Time series data of the soil thermal profile in the surface layer: sensor data and reference series.

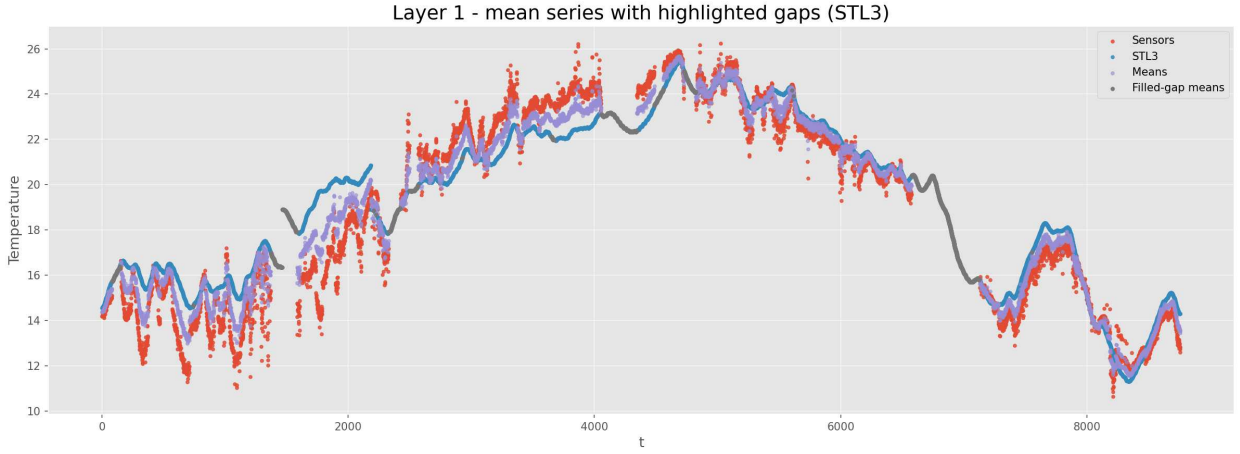


Fig. 5 Time series data of the soil thermal profile in the intermediate layer: sensor data and reference series.

4.2 Global reconstruction performance across the three layers

The global evaluation showed that the pipeline performance was not homogeneous across the three layers, which is consistent with the stratified nature of the problem and with differences in the thermal regime along depth. The comparison between the baseline configuration and the Optuna-optimized version revealed two main behaviors. In layers C_S and C_M , optimization produced substantial and simultaneous gains in absolute error, squared error, and global adherence. In contrast, in layer C_P , the effect was mixed: there was a relevant improvement in RMSE, R^2 , and Pearson correlation, but with a slight worsening in MAE. Consequently, the interpretation of the results should not be formulated in terms of uniform superiority of the optimized model, but rather as a reconfiguration of the fitting regime according to the depth analyzed.

In layer C_S , optimization produced the most expressive gain among the three depths. MAE decreased from 0.166 to 0.062, corresponding to an improvement of approximately 62.6%, while RMSE was reduced from 0.320 to 0.09, with a gain of about 70.5%. In addition, R^2 remained at 0.99, and Pearson correlation also maintained the same value of 0.99. These results indicate a reconfiguration of the model's predictive capacity. From an interpretive perspective, the baseline version was far from a saturation regime; on the contrary, there was still substantial room for improvement, and Optuna was able to explore this space very efficiently. Figure 7 presents the imputation of the surface layer in the global series.

In layer C_M , the effect of optimization was much more expressive. MAE decreased from 0.29 to 0.07, corresponding to an improvement of approximately 75.6%, while RMSE was reduced from 0.62 to 0.08, representing an improvement of 87.1%. In addition, R^2 increased from 0.87 to 0.99, and Pearson correlation remained at 0.99. This set of results indicates that, in the intermediate layer, optimization did not act merely as an incremental adjustment, but effectively reorganized the learning regime of the network, with a clear impact on both error magnitude and the ability to reproduce signal variance. From a computational



Fig. 6 Time series data of the soil thermal profile in the deep layer: sensor data and reference series.

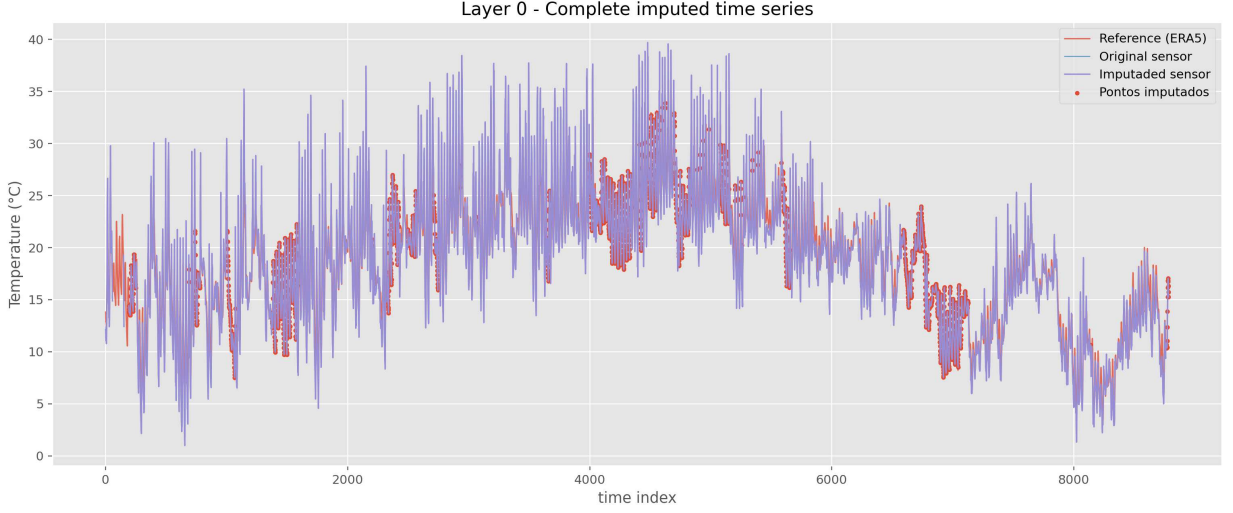


Fig. 7 Time series data of the soil thermal profile in the surface layer: sensor data and reference series.

perspective, C_M was the layer in which the combination of expanded temporal memory, representational capacity, and regularization showed the greatest positive net effect. Figure 8 presents the imputation of the intermediate layer in the global series.

In layer C_P , the behavior was more complex. MAE increased from 0.07 to 0.09, whereas RMSE decreased from 0.21 to 0.16, indicating an improvement of 25.3%. R^2 remained unchanged at 0.99. Pearson correlation also remained stable at 0.99. This pattern shows that optimization produced a model that was more aligned with the global shape of the trajectory, but not necessarily better calibrated in terms of amplitude and pointwise error. Figure 9 presents the imputation of the deep layer in the global series.

Taken together, the global results show that optimization with Optuna should be interpreted as a mechanism for reconfiguring the fitting regime, rather than as a step that necessarily improves all metrics at all depths. However, this is a strength of the article: the analysis does not need to defend a universal superiority of the optimized model, but rather to demonstrate in which layers and under which criteria this optimization adds value, and in which situations it imposes a trade-off among correlation, error, and variance recovery. Table 2 presents the comparison of each metric.

4.2.1 Hyperparameter optimization

The analysis of the selected hyperparameters showed that optimization converged to distinct configurations across the three layers. In layer C_S , the best configuration found combined a reduction in the temporal

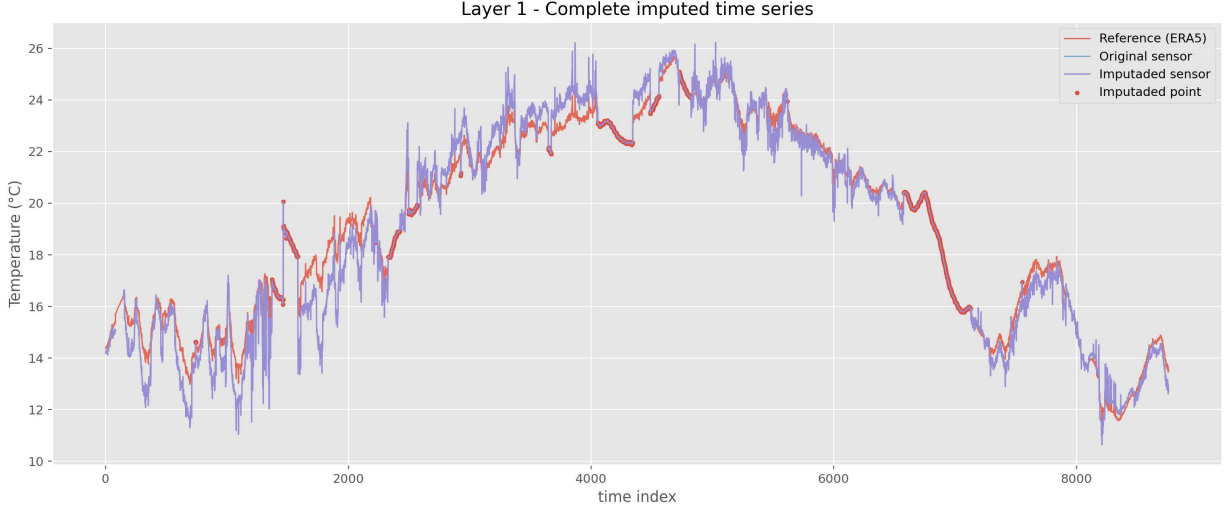


Fig. 8 Time series data of the soil thermal profile in the intermediate layer: sensor data and reference series.



Fig. 9 Time series data of the soil thermal profile in the deep layer: sensor data and reference series.

window with moderate regularization. Table 3 presents the comparison between the respective configurations. This suggests that, at this depth, the search space favored a more compact structure, capable of prioritizing recent information with additional control over overfitting.

In layer C_M , on the other hand, optimization preserved the wider temporal window, but substantially increased the representational capacity of the network by increasing the number of hidden units while maintaining moderate regularization. Table 4 presents the comparison between the baseline configuration and the optimized hyperparameters. The most plausible interpretation is that, at this depth, the search space favored an architecture more capable of exploiting longer temporal dependencies without sacrificing generalization control.

In layer C_P , optimization selected a deeper configuration with greater modeling capacity: a 24-step window, batch size of 32, learning rate $\approx 1.03 \times 10^{-4}$, 256 hidden units, 2 recurrent layers, dropout ≈ 0.420 , and weight decay $\approx 2.59 \times 10^{-5}$. Table 5 presents the comparison between the configurations. At this depth, the interpretation is that the search space favored a more complex structure to capture more damped and less reactive temporal patterns.

From a methodological perspective, this result is important. Hyperparameter optimization has been incorporated as a systematic refinement step in deep models, both through dedicated frameworks, such as Optuna [Akiba et al. \(2019b\)](#), and in specific predictive applications, in which fine-tuning configurations can produce relevant performance gains [Hanifi et al. \(2024\)](#); [Durap \(2025\)](#). However, the results obtained here

Table 2 Global performance comparison between baseline and Optuna-optimized models across the three layers.

Layer	Model	MAE	RMSE	R^2	r (Pearson)
C_S	Baseline	0.16	0.32	0.99	0.99
	Optuna	0.06	0.09	0.99	0.99
C_M	Baseline	0.29	0.62	0.87	0.95
	Optuna	0.07	0.08	0.99	0.99
C_P	Baseline	0.07	0.21	0.99	0.99
	Optuna	0.09	0.16	0.99	0.99

Table 3 Comparison between baseline and Optuna-optimized hyperparameters for the superficial layer (C_S).

Hyperparameter	Baseline	Optuna
Window size (W)	72	24
Batch size	32	16
Learning rate (η)	1.0×10^{-3}	2.11×10^{-3}
Hidden units	64	64
Number of layers	1	1
Dropout	0.00	0.439
Weight decay (λ)	0.00	4.02×10^{-5}
Epochs	25	25

Table 4 Comparison between baseline and Optuna-optimized hyperparameters for the intermediate layer (C_M).

Hyperparameter	Baseline	Optuna
Window size (W)	72	72
Batch size	32	16
Learning rate (η)	1.0×10^{-3}	2.34×10^{-3}
Hidden units	64	256
Number of layers	1	1
Dropout	0.00	0.352
Weight decay (λ)	0.00	3.33×10^{-6}
Epochs	25	25

show a distinct scenario: each layer converged to its own combination of temporal window, network capacity, and regularization. This interpretation reveals that the behavior of the pipeline varies as a function of depth and of the dynamic regime of the series, making any homogeneous interpretation of the optimization process inappropriate.

4.3 Interpretation of global explainability

4.3.1 Global analysis by time

In the global analysis, TimeSHAP indicates which variables and which temporal positions, on average, support the reconstruction produced by the model. For the surface and intermediate layers, however, the results reveal a strong concentration in the final time steps of the input window. As shown in Figures 10 and 11, the mean absolute SHAP values increase abruptly only in the last temporal steps, with a pronounced peak at the terminal time step. This behavior indicates that the model operates under the dominance of the most recent observations, while most of the historical window contributes almost negligibly.

In the deep layer, this pattern becomes different, as temporal dependencies are more gradually distributed. Figure 12 presents this distribution. This behavior is consistent with the physical nature of the deep layer, which is characterized by greater thermal damping and a delayed temporal response.

From a modeling perspective, this behavior suggests that the model does not effectively internalize the full temporal structure available, operating under a short-memory regime even when exposed to a wider window. Methodologically, this raises a central issue: the model’s ability to exploit long-range dependencies remains limited.

Table 5 Comparison between baseline and Optuna-optimized hyperparameters for the deep layer (C_P).

Hyperparameter	Baseline	Optuna
Window size (W)	72	24
Batch size	32	32
Learning rate (η)	1.0×10^{-3}	1.03×10^{-4}
Hidden units	64	256
Number of layers	1	2
Dropout	0.00	0.420
Weight decay (λ)	0.00	2.59×10^{-5}
Epochs	25	25

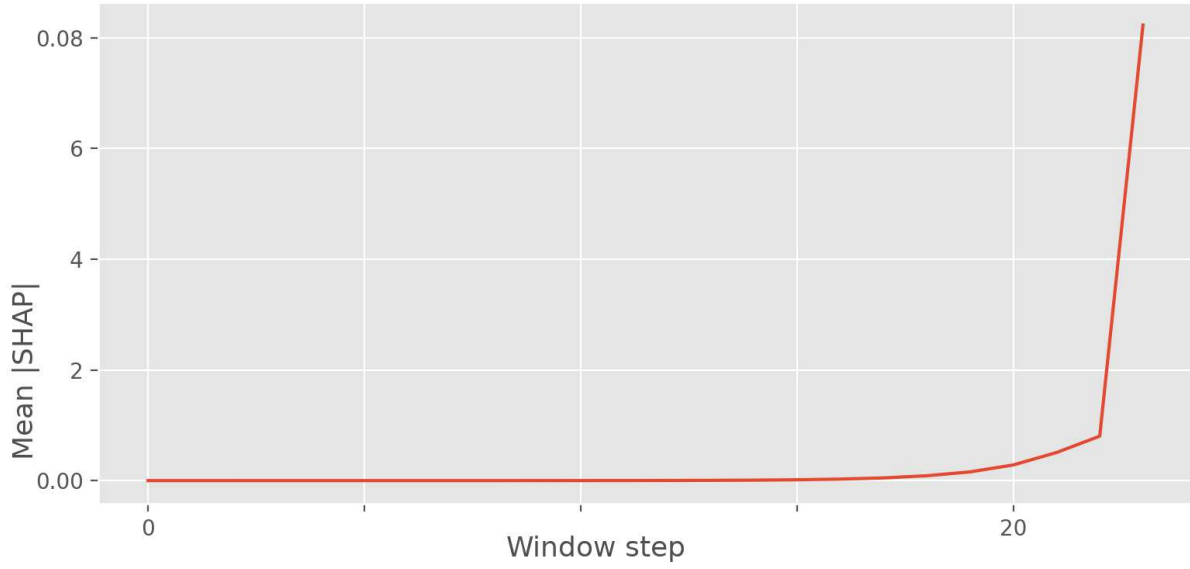


Fig. 10 Global time-based analysis for the surface layer.

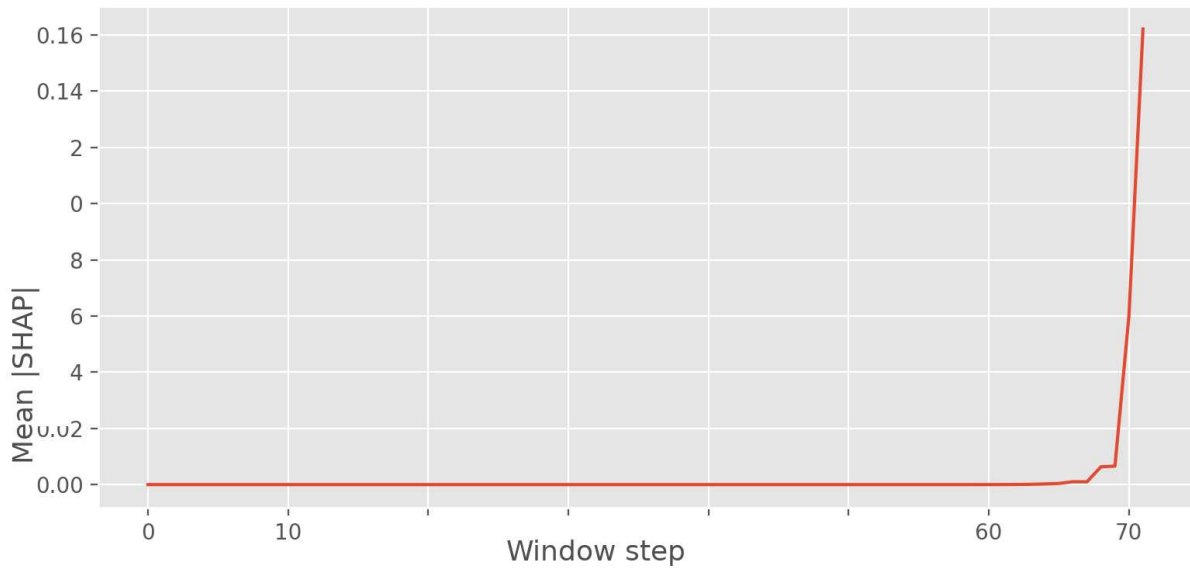


Fig. 11 Global time-based analysis for the intermediate layer.

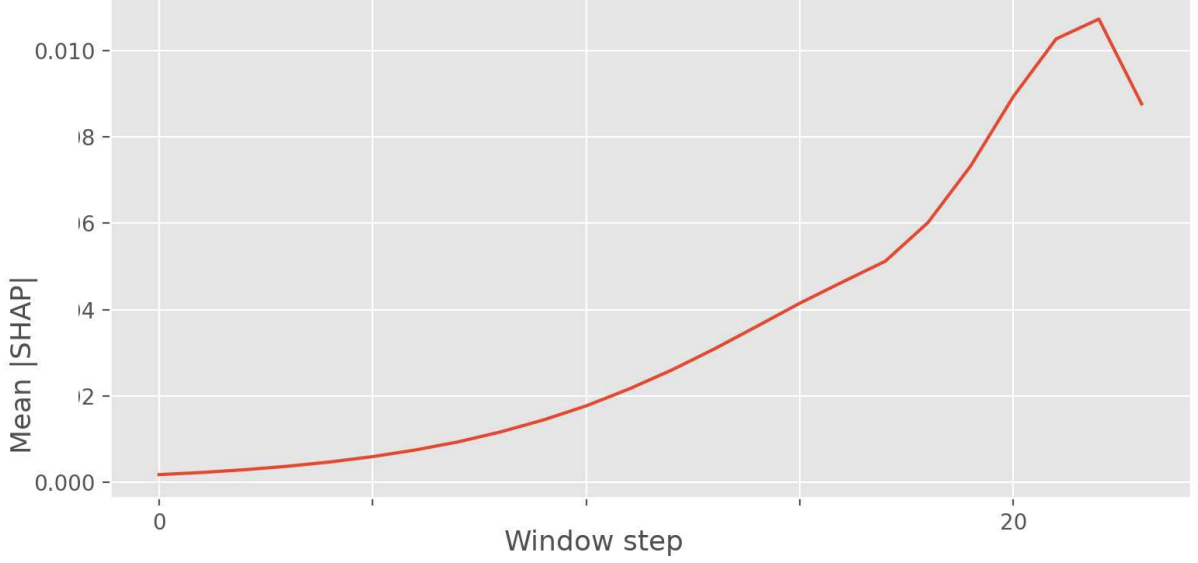


Fig. 12 Global time-based analysis for the deep layer.

4.3.2 Global analysis by variable

The global analysis by variable reinforces the previous conclusion: in the surface and intermediate layers, the models are concentrated on a restricted set of attributes: *ref_trend*, *ref_lag1*, *sensor_interp*, *ref_seasonal* and *ref_lag2*. This pattern indicates that prediction does not emerge from a distributed integration among multiple signals, but rather from a concentrated dependence on variables derived from the reference series and short-term signals. Figures 13, 14, and 15 present the distribution of these variables.

In the deep layer, the predictive logic is concentrated on the following attributes: *time_since_obs*, *sensor_interp*, *ref_trend*, *missing_lag*, *ref_lag1*, and *ref_rolling_mean24*. This pattern is interpretively relevant because it indicates that reconstruction does not depend only on the thermal dynamics themselves, but also on the structure of observational absence. In particular, the high importance of *time_since_obs* and *missing_lag* suggests that the model responds strongly to the geometry of the gap, which may be useful from an operational perspective.

The prominence of *ref_trend* is particularly relevant. Although the reference trend is potentially informative, its dominance may indicate that the model anchors reconstruction in a dominant structural proxy rather than integrating multiple system signals in a more balanced way. In this case, good predictive performance should be interpreted with caution, as it may reflect excessive dependence on an auxiliary variable.

The prominent presence of *sensor_interp* also requires attention. Since this is a variable derived from the sensor itself, its weight may suggest reinforcement of a previously smoothed estimate. From a methodological perspective, this may characterize informational redundancy. Finally, the low relative relevance of attributes associated with longer cycles, such as 24- or 48-step lags and hourly encodings, suggests limited exploitation of the periodic structure of the series.

4.4 Local explainability

The local explainability analysis was conducted with the objective of shifting the interpretation from the aggregate level to the mechanistic level of individual predictions. This step is particularly important because, in temporal reconstruction problems, good average indicators are not sufficient to demonstrate inferential reliability. A model may present high global performance and still make local decisions supported by fragile, excessively concentrated, or physically uninformative mechanisms. In this sense, the methodological choice of evaluating three distinct regimes—low error, high error, and long missing-data block—made it possible to examine not only when the model succeeds, but also how it fails and to what extent it remains operational in effectively critical scenarios.

In the local results used in this step, only the long missing-data block cases correspond to imputations at effectively missing points. The low-error and high-error cases are predictions performed on observed time steps. The first two cases test local calibration at known points, whereas the third tests the pipeline capacity in the target application scenario, that is, reconstruction under real and prolonged absence. If this distinction

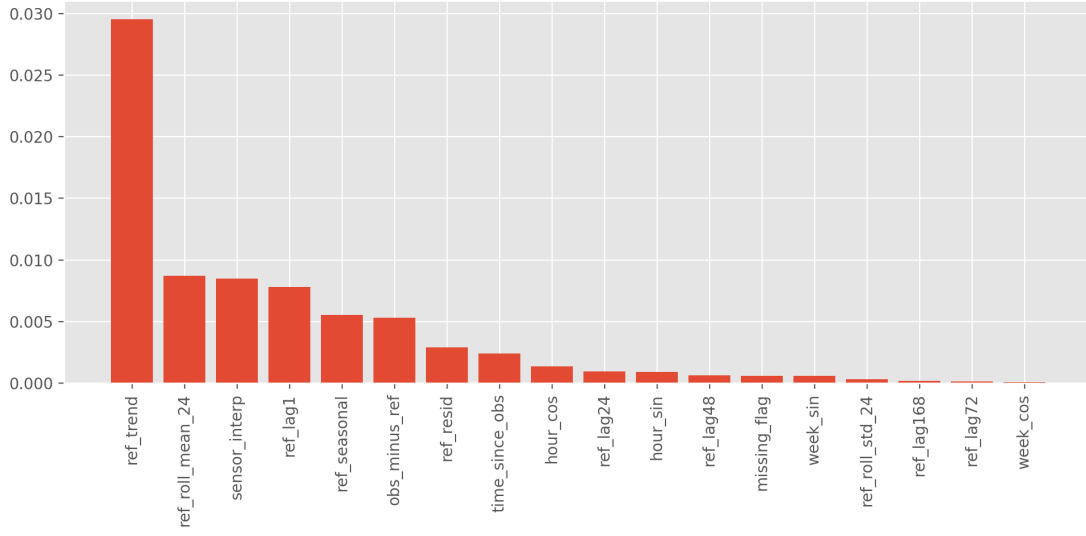


Fig. 13 Global variable-based analysis for the surface layer (C_S).

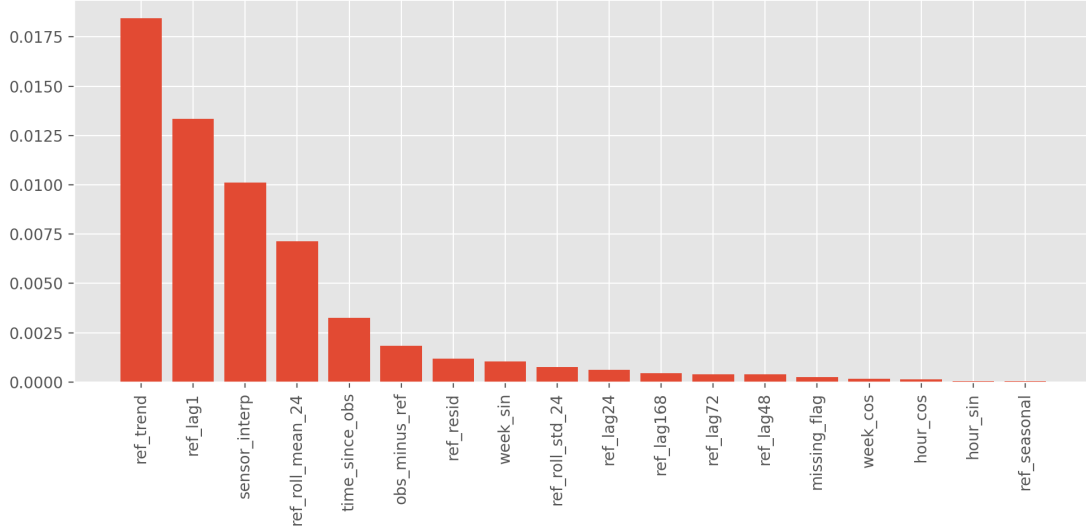


Fig. 14 Global variable-based analysis for the intermediate layer (C_M).

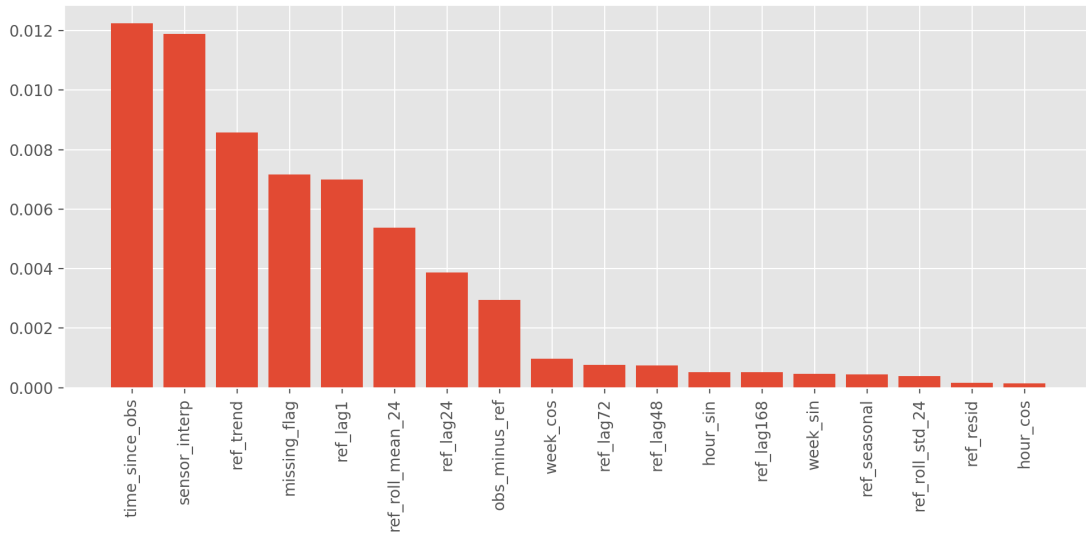


Fig. 15 Global variable-based analysis for the deep layer (C_P).

is not declared, the local explainability section loses epistemological precision and becomes vulnerable to the criticism that it treats diagnostic prediction and imputation as equivalent evidence.

The local results show a clear pattern of asymmetry among the three layers. In the low-error cases, the three depths present practically null residual errors, indicating that the pipeline is capable of reproducing favorable local regimes with high precision. However, this type of successful prediction, when considered in isolation, has limited analytical value and restricted demonstrative strength. In temporal models strongly anchored in recent persistence and highly stable auxiliary attributes, the correct prediction of low-complexity observations does not, by itself, constitute robust evidence of generalization or predictive superiority.

The most relevant evidence emerges in adverse scenarios, particularly in high-error cases and prolonged-gap situations. In these contexts, the intermediate layer presented the most consistent local behavior, with an absolute error of 0.46 in the difficult case and 0.09 in the long-gap scenario. The surface layer, although maintaining operationally acceptable performance in the extended gap, with an error of 0.12, showed more pronounced deterioration in the high-error case, reaching 0.59. The deep layer, in turn, was the most fragile from the local perspective, recording an error of 0.82 in the most difficult case and 0.56 in the prolonged missing-data block scenario. Table 6 presents a summary of the three types of local error cases for each layer.

Table 6 Absolute local error for the three case types across the superficial, intermediate, and deep layers.

Layer	Low-error case	High-error case	Long-gap case
Superficial	0.00	0.59	0.12
Intermediate	0.00	0.46	0.09
Deep	0.00	0.82	0.56

The deep layer presented a qualitatively distinct temporal pattern. In this case, the local explanation was distributed across a larger number of relevant events, with nineteen nonzero contributions and without an absolute collapse at the final time step. The largest event accounted for approximately 18.9% of the absolute explanatory mass; the two largest events accounted for 36.3%; and the four largest events accounted for 65.3%. This represents a more dispersed temporal distribution and, in principle, one that is more compatible with the physics of a damped and delayed thermal regime at greater depth. However, this apparent gain in distribution did not translate into local predictive robustness. On the contrary, the deep layer was precisely the one that presented the largest errors in the critical cases.

The cross-reading of the three cases, stratified by layer, allows a more detailed analytical conclusion to be formulated regarding the behavior of the system. The low-error case indicates that the pipeline achieves high performance when the local structure of the series remains sufficiently aligned with the variables that dominate its predictive dynamics. The high-error case reveals that this alignment is not stable and that, when the real series deviates from the backbone constructed by trend, short lags, and auxiliary interpolation, prediction loses calibration. The long missing-data block case, in turn, is the most important test, as it represents the effective use condition of the method. Under this condition, the intermediate layer was the only one to maintain local performance clearly consistent with its global superiority; the surface layer preserved practical utility, but remained anchored in an excessively reactive mechanism; and the deep layer showed the greatest discrepancy between apparent global stability and local fragility. This contrast among layers is, in itself, relevant to the present study, as it indicates that reconstruction quality depends not only on the LSTM architecture, but also on the dynamic regime of depth, the strength of auxiliary proxies, and the way the model internalizes, or fails to internalize, the available temporal memory.

5 Conclusion

This study demonstrated that LSTM-based soil temperature imputation can achieve high predictive performance, while also showing clearly that this performance is not homogeneous across depths. The global evaluation confirmed that the optimized models achieved very high fitting metrics, with reduced errors and strong agreement between observed and reconstructed values. However, the joint interpretation of the global results, TimeSHAP-based explainability, and performance under prolonged gaps leads to a more stringent conclusion: the pipeline is functionally effective, but its robustness is selective, its internal logic is concentrated on a few dominant predictors, and its reconstruction capacity varies substantially according to the depth analyzed.

The surface layer presented the lowest global mean error, confirming a high fitting capacity in aggregate terms. However, the interpretive analysis showed that this good performance is strongly associated with a short-memory regime, with extreme concentration of importance in the final time steps of the window

and strong dependence on variables such as the reference trend, short lags, and auxiliary interpolation. Consequently, the surface layer proved to be operationally useful, including under long gaps, but supported by an excessively reactive mechanism.

The intermediate layer presented the most consistent behavior throughout the study. It combined high global performance with greater stability in the most difficult local cases and, above all, preserved the best calibration under prolonged gaps. This result is especially relevant because it shows convergence between average quality and robustness in the critical use regime of the method. Unlike the other layers, the intermediate layer not only performed well in aggregate terms, but also sustained this performance when subjected to extended missing-data blocks. This suggests that, at this depth, there was a better balance among the physical dynamics of the process, the representational capacity of the architecture, and the usefulness of the auxiliary variables. Among the three layers, it was the only one whose global superiority found convincing confirmation at the local level and in the scenario of greatest practical demand.

The deep layer, in turn, most clearly exposed the limits of the pipeline. Although it maintained good structural agreement in the global metrics, its local behavior was clearly more fragile, especially in high-error cases and under prolonged gaps. The explainability analysis showed a less temporally collapsed logic than in the shallower layers, which is physically plausible given the greater thermal damping at depth. Even so, this more dispersed temporal distribution did not translate into better reconstruction. On the contrary, the results suggest that the deep layer tends to preserve the general shape of the trajectory, but loses amplitude calibration in critical situations. Analytically, this corresponds to a model that appears stable at the global level, but becomes vulnerable when confronted with locally demanding reconstructions.

From an applied perspective, the results support the potential of LSTM-based imputation for reconstructing soil thermal series, including under prolonged failures, provided that its use is interpreted with caution and according to depth. The intermediate layer emerges as the stratum of greatest operational reliability, the surface layer as an efficient but reactive solution, and the deep layer as the region in which the pipeline requires greater interpretive prudence. Thus, the main contribution of this study is not merely to show that the model “works”, but to delimit more precisely where it works better, where it remains only conditionally reliable, and where it begins to fail. This distinction is methodologically more important than a homogeneous narrative of success, because it brings the evaluation closer to the real behavior of the system under conditions of use.

In summary, this work provides a relevant contribution by showing that high predictive performance and high interpretive quality are not synonymous. The developed pipeline is competent, especially in the intermediate layer, and presents practical value for reconstructing thermal series with gaps. However, the critical analysis revealed that its decision logic remains concentrated, selective, and strongly dependent on auxiliary attributes. Rather than weakening the study, this conclusion strengthens it, because it replaces a simplistic narrative of success with a more rigorous, technically honest, and scientifically defensible characterization of the model’s behavior.

Declarations

Ethics approval: Not applicable.

Consent to participate: Not applicable.

Availability of data and materials: The local soil temperature sensor data can be made available upon request.

Competing interests: The authors declare that they have no competing interests.

Funding - Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), through calls FAPERGS 09/2023 (PqG) and 10/2024 (ARD/ARC), under grant numbers 24/2551-0001485-3 and 25/2551-0000915-4, respectively.

- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Author contributions:

Leticia Barros Dias Soares: Conceptualization, methodology, data processing, model implementation, formal analysis, visualization, and writing – original draft.

Natássia Siqueira Medeiros: Methodological review, interpretation of results, writing – review and editing, and supervision.

João Inácio Moreira Bezerra: Methodological review, computational analysis support, writing – review and editing, and supervision.

Geilson de Almeida Soares: Scientific supervision, interpretation of results, writing – review and editing.

Jairo Valões de Alencar Ramalho: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Projection, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Marlon Maurício Hernandez Cely: Scientific supervision, methodological review, interpretation of results,

and writing – review and editing.

Honório Joaquim Fernando: Scientific supervision, methodological review, and writing – review and editing. All authors read and approved the final manuscript.

Code availability The code used for data preprocessing, model training, hyperparameter optimization, reconstruction, and explainability analysis may be made available upon reasonable request.

Declaration on the use of artificial intelligence The authors declare that generative artificial intelligence tools were used as auxiliary tools during manuscript preparation, specifically for language editing, improvement of textual clarity, support in the structural organization of selected sections, and assistance with the standardization/development of visual elements. The use of these tools did not involve autonomous generation of scientific results, definition of the methodology, independent data analysis, final interpretation of results, or full/major writing of the manuscript. All content suggested or adjusted with artificial intelligence support was critically reviewed, edited, checked, and validated by the authors, who take full responsibility for the authorship, originality, accuracy of information, scientific integrity, and final version of the manuscript. No artificial intelligence tool was considered an author of the manuscript.

Acknowledgements. Soares, L. and Soares, G thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for their scholarships. The authors acknowledge the technical and research staff at the Universidad Tecnológica del Uruguay (UTEC) involved in the deployment, maintenance, and data acquisition of the environmental monitoring stations. This study was supported by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), through calls FAPERGS 09/2023 (PqG) and 10/2024 (ARD/ARC), under grant numbers 24/2551-0001485-3 and 25/2551-0000915-4, respectively.

References

- Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J.M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., Herrera, F.: Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion* **99**, 101805 (2023) <https://doi.org/10.1016/j.inffus.2023.101805>
- Adekanmbi, A.A., Sizmur, T.: Importance of diurnal temperature range (DTR) for predicting the temperature sensitivity of soil respiration. *Frontiers in Soil Science* **2**, 969077 (2022) <https://doi.org/10.3389/fsoil.2022.969077>
- Alejo-Sanchez, L.E., Márquez-Grajales, A., Salas-Martínez, F., Franco-Arcega, A., López-Morales, V., Acevedo-Sandoval, O.A., González-Ramírez, C.A., Villegas-Vega, R.: Missing data imputation of climate time series: A review. *MethodsX* **15**, 103455 (2025) <https://doi.org/10.1016/j.mex.2025.103455>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631. ACM, ??? (2019). <https://doi.org/10.1145/3292500.3330701> . <https://doi.org/10.1145/3292500.3330701>
- Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2623–2631 (2019). <https://doi.org/10.1145/3292500.3330701> . <https://doi.org/10.1145/3292500.3330701>
- Alizamir, M., Wang, M., Ikram, R.M.A., Ahmed, K.O., Heddami, S., Kim, S.: An efficient computational investigation on accurate daily soil temperature prediction using boosting ensemble methods explanation based on SHAP importance analysis. *Results in Engineering* **24**, 103220 (2024) <https://doi.org/10.1016/j.rineng.2024.103220>
- Bodjrenou, S., *et al.*: Evaluation of reanalysis estimates of precipitation, radiation, and temperature in Benin. *Journal of Applied Meteorology and Climatology* **62**(8), 1005–1022 (2023) <https://doi.org/10.1175/JAMC-D-21-0222.1>
- Barboza Pereira, P.G.: Implementación de un sistema de monitoreo iot para caracterizar el perfil térmico del suelo. Monografía de grado, Universidad Tecnológica del Uruguay (UTEC), Uruguay (2024)
- Brazil: Planeta pode ultrapassar 1,5°C nos próximos cinco anos, diz relatório da Organizações Meteorológica Mundial (2024). <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2024/06/>

- Bento, J., Saleiro, P., Antunes, M., Soares, C.: Timeshap: Explaining recurrent models through sequence perturbations. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 2565–2573. ACM, ??? (2021). <https://doi.org/10.1145/3447548.3467166> .
<https://doi.org/10.1145/3447548.3467166>
- Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.J.: Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* **6**(1), 3–33 (1990). Accessed 2026-01-22
- Du, W., *et al.*: Saits: Self-attention-based imputation for time series. *Expert Systems with Applications* (2023) <https://doi.org/10.1016/j.eswa.2023.119619>
- Peña, M.F., *et al.*: Shats: a shapley-based explainability method for time series models. *Future Generation Computer Systems* (2025) <https://doi.org/10.1016/j.future.2025.107801>
- Domingues, A.M.B., Nóbrega, E.S.B., Ramalho, J.V.A., Brum, R.S., Quadros, R.S.: Parameter analysis of earth-air heat exchangers over multi-layered soils in south brazil. *Geothermics* **93**, 102074 (2021) <https://doi.org/10.1016/j.geothermics.2021.102074>
- Domingues, A.M.B., Razera, A.L., Ramalho, J.V.A., Rodrigues, M.K., Fernando, H.J., Isoldi, L.A., dos Santos, E.D.: Performance analysis and geometric evaluation of galvanized t- and y-shapes for earth-air heat exchangers using constructal design. *Geothermics* **136**, 103595 (2026) <https://doi.org/10.1016/j.geothermics.2025.103595>
- Durap, A.: Explainable deep learning techniques for wind speed forecasting in coastal areas: Integrating model configuration, regularization, early stopping, and shap analysis. *Neural Computing and Applications* **37**, 21219–21257 (2025) <https://doi.org/10.1007/s00521-025-11433-w>
- Vries, A.: The growing energy footprint of artificial intelligence. *Joule* **7**(10), 2191–2194 (2023) <https://doi.org/10.1016/j.joule.2023.09.004>
- ECMWF: ECMWF Reanalysis v5 (ERA5) (2026). <https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5> Accessed 2026-01-23
- Feedzai: TimeSHAP. <https://github.com/feedzai/timeshap>. Official repository used to support implementation and experimental adaptation. Accessed 23 mar. 2026 (2026)
- Geng, Q., *et al.*: Soil temperature prediction based on explainable artificial intelligence and LSTM. *Frontiers in Environmental Science* **12**, 1426942 (2024) <https://doi.org/10.3389/fenvs.2024.1426942>
- Hersbach, H., *et al.*: The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* (2020) <https://doi.org/10.1002/qj.3803>
- Hao, J., *et al.*: Improving long-term multivariate time series forecasting using seasonal-trend decomposition-based deep learning. *Scientific Reports* (2024) <https://doi.org/10.1038/s41598-024-52240-y>
- Hyndman, R.J., Athanasopoulos, G.: *Forecasting: Principles and Practice*, 3rd edn. OTexts, Melbourne (2021). <https://otexts.com/fpp3/> Accessed 2026-01-22
- Hanifi, S., Cammarono, A., Zare-Behtash, H.: Advanced hyperparameter optimization of deep learning models for wind power prediction. *Renewable Energy* **221**, 119700 (2024) <https://doi.org/10.1016/j.renene.2023.119700>
- Hashemi, S., Samadianfard, S., Nazemi, A.H.: Sequence-to-sequence and meta lstm techniques enhanced by remote sensing for advanced evapotranspiration estimation. *Earth Science Informatics* **18**, 386 (2025) <https://doi.org/10.1007/s12145-025-01896-9>
- Imanloozadeh, A., Lubitz, W.D.: Characteristics, design, and optimization of earth-air heat exchangers: A review. *Renewable and Sustainable Energy Reviews* **230**, 116677 (2026) <https://doi.org/10.1016/j.rser.2025.116677>

- Jacobs, A., Top, S., Vergauwen, T., Suomi, J., *et al.*: Filling gaps in urban temperature observations by debiasing era5 reanalysis data. *Urban Climate* (2024) <https://doi.org/10.1016/j.uclim.2024.102226>
- Kumar, H.K.Y., *et al.*: Revitalizing temperature records: A novel framework for addressing long continuous gaps in temperature time series. *Agricultural and Forest Meteorology* (2024) <https://doi.org/10.1016/j.agrformet.2024.110322>
- Khosravi, K., Golkarian, A., Barzegar, R., Alami, M.T., López-Vicente, M.: Multi-step ahead soil temperature forecasting at different depths based on meteorological data: Integrating resampling algorithms and machine learning models. *Pedosphere* **33**(3), 510–525 (2023) [https://doi.org/10.1016/S1002-0160\(22\)60062-5](https://doi.org/10.1016/S1002-0160(22)60062-5)
- Kan, J.-C., Passos, M.V., Destouni, G., Barquet, K., Ferreira, C.S.S., Kalantari, Z.: Forecasting heat-related impacts with multivariate multi-step time series models using advanced deep learning. *Sustainable Cities and Society* **137**, 107142 (2026) <https://doi.org/10.1016/j.scs.2026.107142>
- Kumar, H.K.Y., Varija, K.: Revitalizing temperature records: A novel framework towards continuous data reconstruction using univariate and multivariate imputation techniques. *Atmospheric Research* **312**, 107689 (2024) <https://doi.org/10.1016/j.atmosres.2024.107689>
- Lembrechts, J.J., Aalto, J., Ashcroft, M.B., De Frenne, P., Kopecký, M., Lenoir, J., Luoto, M., Maclean, I.M.D., Nijls, I., *et al.*: Global maps of soil temperature. *Global Change Biology* **28**(9), 3110–3144 (2022) <https://doi.org/10.1111/gcb.16060>
- Lim, B., Arik, S.Ö., Loeff, N., Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* **37**(4), 1748–1764 (2021) <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates, Inc., ??? (2017). <https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions>
- Liu, Z., Xie, M., Zhou, Y., He, Y., Zhang, L., Zhang, G., Chen, D.: A state-of-the-art review on shallow geothermal ventilation systems with thermal performance enhancement: System classifications, advanced technologies, and applications. *Energy and Built Environment* **4**(2), 148–168 (2023) <https://doi.org/10.1016/j.enbenv.2021.10.003>
- Lai, X., Yao, Y., Mu, J., Lu, W., Zhang, L.: Ratai: recurrent autoencoder with imputation units and temporal attention for multivariate time series imputation. *Artificial Intelligence Review* **58**, 37 (2025) <https://doi.org/10.1007/s10462-024-11039-z>
- Mampitiya, L.I., Rozumbetov, K., Rathnayake, N., Erkudov, V., Esimbetov, A., Arachchi, S., *et al.*: Artificial intelligence to predict soil temperatures by development of novel model. *Scientific Reports* **14**, 9889 (2024) <https://doi.org/10.1038/s41598-024-60549-x>
- Nasiri Khiavi, A.: Application of remote sensing indicators and deep learning algorithms to model the spatial distribution of soil moisture. *Earth Science Informatics* **18**, 404 (2025) <https://doi.org/10.1007/s12145-025-01842-9>
- Nixtla: NeuralForecast. <https://github.com/nixtla/neuralforecast>. Accessed 21 mar. 2026 (2026)
- Niako, N., Melgarejo, J.D., Maestre, G.E., Vatcheva, K.P.: Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with arima and lstm. *BMC Medical Research Methodology* **24** (2024) <https://doi.org/10.1186/s12874-024-02448-3>
- Optuna Project: Optuna. <https://github.com/optuna/optuna>. Official repository used to support implementation and reproducibility. Accessed 23 mar. 2026 (2026)
- Pastorini, M., *et al.*: Enhancing environmental data imputation: A physically-constrained machine learning framework. *Science of the Total Environment* **926**, 171773 (2024) <https://doi.org/10.1016/j.scitotenv.2024.171773>
- Perlaky, N., Ellenburg, W.L., Nair, U., Klein, N., Solomon, M.B.: Soil temperature, environment, and

- moisture monitoring network—a low-cost sensing network for alabama. *Environmental Monitoring and Assessment* **197**(7), 724 (2025) <https://doi.org/10.1007/s10661-025-14166-4>
- Park, J., Muller, J., Arora, B., Pastorello, G., Varadharajan, C., Sahu, R., Agarwal, D.: Long-term missing value imputation for time series data using deep neural networks. *Neural Computing and Applications* **35**(13), 9687–9700 (2023) <https://doi.org/10.1007/s00521-022-08165-6>
- Ramalho, J.V.A., Fernando, H.J., Brum, R.S., Domingues, A.M.B., Navarro Pastor, N.R., Burlón Olivera, M.R.: Accessing the thermal performance of earth–air heat exchangers surrounded by galvanized structures. *Sustainable Energy Technologies and Assessments* **54**, 102838 (2022) <https://doi.org/10.1016/j.seta.2022.102838>
- Schönauer, M., *et al.*: Topographic indices and era5-land data to describe soil moisture and soil temperature patterns in a temperate forest. *Journal of Hydrology: Regional Studies* **59**, 102336 (2025) <https://doi.org/10.1016/j.ejrh.2025.102336>
- Silva, A.L., *et al.*: Comparing observed and reanalysis data for trends in temperature extremes across brazil. *Theoretical and Applied Climatology* (2025) <https://doi.org/10.1007/s00704-025-05777-5>
- Santamaría-Bonfil, G., *et al.*: Equivalent imputation methodology for handling missing data in geochemical databases. *Chemometrics and Intelligent Laboratory Systems* (2022) <https://doi.org/10.1016/j.chemolab.2022.104644>
- Santamaría-Bonfil, G., Santoyo, E., Díaz-González, L., Arroyo-Figueroa, G.: Equivalent imputation methodology for handling missing data in compositional geochemical databases of geothermal fluids. *Geothermics* **104**, 102440 (2022) <https://doi.org/10.1016/j.geothermics.2022.102440>
- SHAP Contributors: SHAP. <https://github.com/shap/shap>. Official repository used to support implementation and experimental adaptation. Accessed 23 mar. 2026 (2026)
- SHAP Developers: SHAP KernelExplainer Documentation. <https://shap.readthedocs.io/en/latest/generated/shap.KernelExplainer.html>. Accessed 21 mar. 2026 (2026)
- Soares, T.A., Langner, M., Sacht, H., Figueiredo, A., Vicente, R.: Multi-objective optimisation of an earth-to-air heat exchanger for thermal regulation in southern brazilian bioclimatic zones. *Energy and Buildings* **347**, 116243 (2025) <https://doi.org/10.1016/j.enbuild.2025.116243>
- Tian, Z., Yu, X., Feng, G.: Short-term wind speed prediction model based on long short-term memory network with feature extraction. *Earth Science Informatics* **18**, 333 (2025) <https://doi.org/10.1007/s12145-025-01851-8>
- Wu, Y., *et al.*: Effective lstms with seasonal-trend decomposition and ... *Expert Systems with Applications* (2024)
- Yilmaz, M.T.: Accuracy assessment of temperature trends from era5 and era5-land reanalyses using ground station observations in türkiye. *Science of the Total Environment* **858**, 159787 (2023) <https://doi.org/10.1016/j.scitotenv.2022.159787>
- Zhao, J., *et al.*: Multivariate time series data imputation using attention-based generative adversarial networks. *Neurocomputing* (2023) <https://doi.org/10.1016/j.neucom.2023.126669>
- Zhan, Y., Chen, J.: Integrating shap and lime in lstm-based models. *Journal of Information Systems Engineering and Management* **10**(16s), 412–423 (2025) <https://doi.org/10.52783/jisem.v10i16s.2627>
- Zhao, J., Zhao, F., Deng, H.: Evaluation of climate prediction models in yunnan, china: traditional methods and ai approaches. *Scientific Reports* **15**, 43347 (2025) <https://doi.org/10.1038/s41598-025-27326-w>