Fusion of Orbital and Surface Data for Air Temperature Estimation in Belém Using Machine Learning

Leonardo de O. Tamasauskas[1], Waldemiro J. A. G. Negreiros[1], Pedro H. do V. Guimarães[1], Williane G. S. Pereira[1], Jean A. C. Dias[1], Gabriel B. Costa[2], Marcos C. da R. Seruffo[1,2]

[1]Operations Research Laboratory (LPO)
Federal University of Pará (UFPA), Belém-PA, Brazil

[2]Graduate Program in Anthropogenic Studies in the Amazon
Federal University of Pará (UFPA), Castanhal-PA, Brazil

leonardo.tamasauskas@icen.ufpa.br, waldemiro.negreiros@ifpa.edu.br,
pedro.guimaraes@itec.ufpa.br, williane.pereira@icen.ufpa.br,
jean.dias@itec.ufpa.br, gabrielbritocosta@gmail.com,
seruffo@ufpa.br

Abstract. *This study evaluates fusion strategies between ERA5-Land orbital data and INMET surface observations to estimate daily air temperature in Belém, using data from 1994 to 2024. Different Random Forest models were compared, along with SARIMAX, LSTM, and stacking (SARIMAX and RF). The RF model with multivariate fusion achieved the best performance (RMSE = 0.5157 °C; R^2 = 0.708; Skill = 6.19%), reducing bias relative to the orbital baseline, while chronological validation confirmed the temporal stability of RF-based models. The results indicate that, under low equatorial thermal variability, tree-based models are the most robust for satellite–surface fusion and thermal downscaling.*

1. Introduction

Understanding thermal variability is crucial for assessing the impacts of global climate change, which has already resulted in an increase of 1.59 °C in the global mean surface temperature compared to pre-industrial levels [Wei et al. 2025]. This warming has contributed to unprecedented climate extremes, suggesting that intense events may become the new norm even during La Niña phases [McCulloch et al. 2024]. In the context of the eastern Amazon, specifically in humid equatorial climate regions such as the municipality of Belém, rising air temperatures directly impact forest ecology by altering the thermal niches of species and compromising biodiversity [Ismaeel et al. 2024]. Continuous and accurate monitoring of these variables therefore becomes indispensable for mitigating environmental damage and formulating regional adaptation strategies.

Despite its importance, ground-based monitoring faces significant operational challenges. Surface meteorological stations, such as those operated by the Instituto Nacional de Meteorologia (INMET), are essential for *in situ* measurements and local calibration, but they have inherent limitations in spatial coverage [van Tiggelen et al. 2025]. This scarcity of sensors creates "data deserts" that hinder the generalization of observations to larger areas and limit the capture of microclimatic variations [Brousse et al. 2024, Rains et al. 2024]. To address this gap, orbital data and land reanalysis estimates, such as

ERA5-Land, have become vital tools. However, the integration of these sources requires rigorous quality control procedures and temporal regularization to ensure compatibility and homogeneity of daily time series.

To address the complexity of meteorological time series and optimize the integration of these multiple sources, the literature has demonstrated significant advances in the application of statistical methods and Artificial Intelligence techniques [Camps-Valls et al. 2025, Nourani et al. 2025]. Contemporary predictive models, ranging from statistical approaches with seasonal components (SARIMAX) to Machine Learning algorithms based on decision trees (Random Forest) and *Deep Learning* architectures (Long Short-Term Memory - LSTM), offer enhanced capabilities for capturing nonlinear dynamics [Bhatti et al. 2024]. The use of combined strategies, such as *stacking ensembles*, has also been explored to maximize predictive robustness by leveraging the strengths of different algorithms in climate modeling.

Despite the proliferation of these approaches, a critical scientific gap remains regarding the applicability and generalization of complex models in equatorial regions. The intrinsic complexity of recurrent neural networks exacerbates modeling challenges in scenarios with low annual thermal amplitude, often resulting in unstable simulations when these models operate outside their ideal statistical distribution [Lai et al. 2025]. Additionally, multivariate fusion of satellite and surface data at these latitudes lacks robust temporal cross-validation protocols that account for strong dependence on non-thermal covariates and mitigate the risk of data leakage-factors that often mask the true model error [Lopez-Gomez et al. 2025].

In this context, the present study aims to evaluate the performance of different data fusion scenarios and predictive architectures (Random Forest, SARIMAX, LSTM, and *stacking*) for estimating daily air temperature in the municipality of Belém, Pará. Using the temporal integration of INMET observations and ERA5-Land estimates between 1994 and 2024, this study implements rigorous chronological cross-validation. As its main contribution, this work elucidates the behavior of machine learning models in low thermal variability scenarios, demonstrating the superiority and stability of decision tree-based algorithms in capturing nonlinear relationships, in contrast to the generalization limitations exhibited by deep neural network architectures in this specific geographic context.

2. Methodology

The methodology consisted of integrating air temperature data observed by INMET with ERA5-Land estimates for the period 1994–2024, followed by quality control, temporal regularization, and the generation of derived features based on autocorrelation and seasonality. The data were chronologically divided into training, validation, and test sets to avoid temporal leakage. Random Forest models were evaluated under different data fusion scenarios, along with SARIMAX, LSTM, and *stacking* approaches, with optimization performed via temporal cross-validation and comparison based on standardized statistical metrics.

2.1. Characterization of the Study Area

The study was conducted in the municipality of Belém, Pará, located in the eastern Amazon, a region characterized by a humid equatorial climate and low annual thermal am-

plitude [Alvares et al. 2013]. The target variable corresponds to the daily air temperature observed at the automatic meteorological station of the Instituto Nacional de Meteorologia (INMET), which was adopted as the surface reference [Fick and Hijmans 2017]. As the orbital predictor variable, the 2 m air temperature from the ERA5-Land reanalysis product was used, covering the period from 1994 to 2024 [Muñoz Sabater et al. 2021]. This dataset has a spatial resolution of $0.1° \times 0.1°$ (≈ 9 km) and consistent daily temporal resolution, and is widely used in climate studies due to its high spatial coverage and temporal homogeneity [Hersbach et al. 2020].

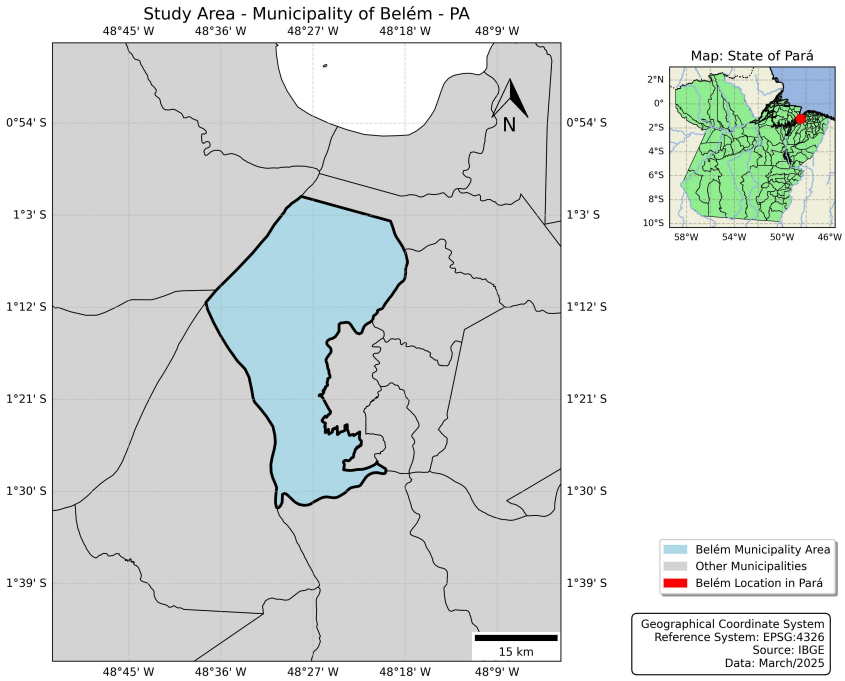

Figure 1. Location of the study area in Belém, Pará.

2.2. Data Collection and Preprocessing

Two main data sources were used: surface meteorological observations from INMET and 2 m air temperature estimates derived from ERA5-Land, both with daily temporal resolution. The hourly INMET records were aggregated into daily averages to ensure temporal compatibility with the orbital data. Quality control procedures included the removal of missing or invalid values, verification of temporal duplication, and aggregation into daily means. Surface observations were considered as *ground truth* for model calibration and validation [Silván-Cárdenas and Wang 2010].

The orbital data were converted from Kelvin to degrees Celsius and temporally indexed. Integration between datasets was performed through daily temporal intersection, retaining only the dates common to both series. Subsequently, the resulting series was regularized to a continuous daily frequency, ensuring compatibility with time series models [Wilks 2011]. Short temporal gaps were filled using time-dependent linear interpolation, while missing values at the boundaries were handled using *forward fill* and *backward fill*, a procedure recommended for meteorological variables with low daily variability [Little and Rubin 2019].

After preprocessing, the final dataset consisted of a continuous daily time series containing observed air temperature (target variable), orbital air temperature (main predictor), and a temporal index used for generating seasonal features and lags [Hubbard and Lin 2006, Hersbach et al. 2020, Wilks 2011].

2.3. Feature Engineering

Derived features were constructed based on physical-climatic knowledge and air temperature autocorrelation, including 7- and 30-day moving averages, thermal anomalies, temporal trends, cyclic seasonal representations using sine and cosine functions, lags from 1 to 30 days, and moving standard deviations of 7, 14, and 30 days. These features capture thermal persistence, intra-seasonal variability, and annual periodicity, which are fundamental characteristics of meteorological time series. The use of lags and moving statistics is grounded in time series theory [Box et al. 2015], while cyclic encoding of seasonality using trigonometric functions is recommended for continuous periodic variables [Hyndman and Athanasopoulos 2018]. The inclusion of derived features allows machine learning models to capture nonlinearities and temporal dependencies, resulting in improved predictive performance [Khosravi et al. 2011].

2.4. Temporal Data Splitting

The data were chronologically divided into three subsets: training (up to 2016), validation (2017–2020), and test (2021–2024). This strategy prevents temporal leakage and enables the evaluation of model generalization on unseen data [Hyndman and Athanasopoulos 2018].

2.5. Evaluated Random Forest Models

Three experimental scenarios were evaluated: (i) **RF_SatOnly**, which uses only the orbital predictor (baseline); (ii) **RF_Fusion**, which combines orbital temperature with temporal lags; and (iii) **RF_Fusion_Features**, which incorporates the complete set of derived climatic features. Hyperparameters were optimized using random search (*RandomizedSearchCV*) with temporal cross-validation (*TimeSeriesSplit*) [Bergstra and Bengio 2012, Hyndman and Athanasopoulos 2018]. Random Forest-based models were employed due to their ability to model nonlinear relationships and their robustness to noise in air temperature estimation from orbital data [Wu and Li 2019].

2.6. SARIMAX Model

The SARIMAX model was implemented with exogenous variables (orbital temperature and seasonal terms), daily frequency, and a three-year rolling window to capture recent dynamics in the series. The parameters $(1, 1, 1)(1, 0, 1)[7]$ were adopted, representing weekly seasonality. SARIMAX models are widely used in meteorological time series because they capture linear temporal dependence and seasonal components [Box et al. 2015, Kothiyal et al. 2025].

2.7. LSTM Model

A recurrent neural network of the LSTM type was implemented with 30-day input sequences, MinMax normalization fitted only on the training set, two LSTM layers

with 64 and 32 units, *dropout* of 0.2, and *early stopping* based on validation error. LSTM networks are suitable for modeling long-term temporal dependencies, although they require long time series and sufficient variability for optimal performance [Hochreiter and Schmidhuber 1997].

2.8. Stacking Ensemble

A *stacking ensemble* approach was adopted, combining Random Forest and SARIMAX as base models and Ridge regression as the meta-model, with regularization defined via cross-validation. Ensemble strategies tend to improve predictive robustness by combining models with different biases and variances [Wolpert 1992, Sagi and Rokach 2018].

2.9. Evaluation Metrics

Model performance was evaluated using the metrics RMSE, MAE, R^2, Bias, NSE (Nash–Sutcliffe Efficiency), and Skill Score relative to the orbital baseline. These metrics are widely used in hydrometeorological and climate modeling studies to quantify error, explanatory power, and relative performance [Willmott and Matsuura 2005, Nash and Sutcliffe 1970, Murphy 1988].

3. Results and Discussion

Table 1 presents the performance of the models in estimating daily air temperature in Belém. The **RF_Fusion** model achieved the best performance (RMSE = 0.5157 °C; $R^2 = 0.708$), with a skill gain of 6.19% compared to the orbital baseline (**RF_SatOnly**). The **RF_Fusion_Features** model showed statistically equivalent results, indicating that the inclusion of additional climatic variables did not produce significant gains, possibly due to collinearity among predictors.

Table 1. Performance of the models in estimating daily air temperature during the test period (2021–2024).

Model	RMSE (°C)	MAE (°C)	R^2	Bias (°C)	NSE	Skill (%)
RF_Fusion	0.5157	0.4039	0.7080	0.0211	0.7080	6.19
RF_Fusion_Features	0.5159	0.4052	0.7078	0.0051	0.7078	6.15
Stacking_Final	0.5165	0.4046	0.7071	0.0441	0.7071	6.05
RF_SatOnly	0.5497	0.4301	0.6682	0.0703	0.6682	0.00
SARIMAX	0.5531	0.4337	0.6640	0.0894	0.6640	-0.63
LSTM	0.6447	0.4976	0.5477	0.0259	0.5477	-17.28

The **Stacking_Final** model did not outperform the standalone Random Forest, suggesting that the ensemble did not add additional predictive information relative to the base model. The **SARIMAX** model showed inferior performance compared to the baseline, highlighting its limitations in representing nonlinear relationships. The **LSTM** model exhibited the highest error (RMSE = 0.6447 °C).

The comparison between **RF_SatOnly** and **RF_Fusion** demonstrates the importance of surface observations for model calibration, with an RMSE reduction of approximately 0.034 °C and lower systematic bias. Figures 2 and 3 show that the fusion models better capture temporal variability and align more closely with the 1:1 line, while SARIMAX exhibits excessive smoothing and LSTM presents greater dispersion.

The temporal stability of the models was evaluated using chronological cross-validation with *TimeSeriesSplit*. The **RF_Fusion_Features** model achieved a mean RMSE of 0.5188 °C (± 0.0104), indicating high consistency across partitions and strong generalization capability. The **SARIMAX** model showed a mean RMSE of 0.5258 °C (± 0.0092), reflecting inferior performance and greater smoothing of thermal variations, although with similar variability across splits. These results reinforce the robustness of tree-based models in capturing nonlinear relationships over time. For the **LSTM**, temporal cross-validation was not applied due to its high computational cost and dependence on continuous sequences; its evaluation was performed only using chronological train–validation–test splitting, a common procedure for recurrent architectures in time series.

Additionally, the feature engineering strategy adopted in the Random Forest models incorporated thermal memory, temporal trends, rolling variability, and cyclical seasonal components, allowing the models to better represent short-term atmospheric persistence and annual climatic behavior. Although the inclusion of advanced derived variables in **RF_Fusion_Features** did not substantially improve predictive accuracy relative to **RF_Fusion**, this result suggests that the baseline fusion structure already captured most of the relevant thermal information, while additional predictors may have introduced redundant information due to collinearity effects. This behavior reinforces the stability and interpretability of tree-based ensemble methods under low thermal variability conditions in equatorial environments.

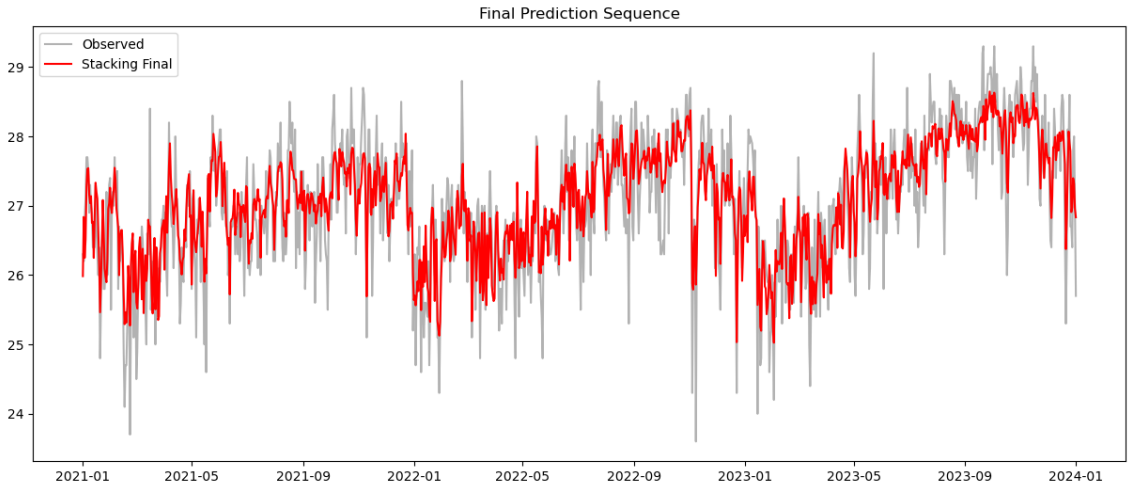

Figure 2. Comparison between observed air temperature (gray) and predictions from the final stacking model (red) over the test period (2021–2024), showing the model's ability to capture temporal variability and general trends while smoothing short-term fluctuations.

The results indicate that the high effectiveness of the LSTM model in purely univariate autoregressive tasks, as reported by [Tamasauskas et al. 2025] with $RMSE = 0.46°C$ and $R^2 = 0.83$, was not maintained in the satellite–surface fusion scenario proposed in this study, which yielded $RMSE = 0.6447°C$ and $R^2 = 0.5477$. This disparity can be attributed to the multivariate nature of the problem, the rigorous chronological split with a temporal *gap*, and the low regional thermal variability, which reduce the advantages of sequential learning. In this context, the robustness of Random Forest, with

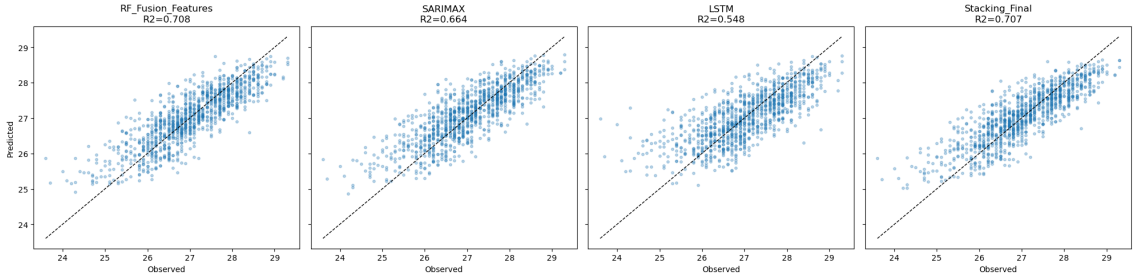

Figure 3. Scatter plots between observed and estimated values for the RF_Fusion_Features, SARIMAX, LSTM, and Stacking_Final models. The dashed line represents the 1:1 relationship.

$RMSE = 0.5157°C$, outperformed the recurrent neural network in capturing nonlinear relationships under equatorial conditions. These results reinforce the suitability of tree-based models for thermal *downscaling* and data fusion in regions with low thermal amplitude.

4. Conclusion

To evaluate the performance of different data fusion strategies combining orbital and surface temperature sources for estimating air temperature in the city of Belém, machine learning, statistical, and hybrid models were compared. Overall, Random Forest-based models achieved superior performance, outperforming models such as LSTM, SARIMAX, and the stacking ensemble combining SARIMAX and RF. Among the RF-based models, it is noteworthy that the use of orbital fusion with temporal lags and the fusion with the complete set of derived climatic features showed similar performance, with improvements of 6.19% and 6.15%, respectively, compared to the orbital baseline model. SARIMAX exhibited limitations in modeling nonlinearities, while LSTM showed the worst performance, despite strong results in univariate autoregressive forecasting. These results demonstrate the suitability of tree-based models in the context of the study area, an equatorial region with low thermal variability. For future work, it will be important to investigate the validity of these findings across different cities in the Amazon region, encompassing both coastal and inland areas, as well as varying levels of urbanization.

References

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. d. M., Sparovek, G., et al. (2013). Köppen's climate classification map for brazil. *Meteorologische zeitschrift*, 22(6):711–728.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).

Bhatti, M. A., Song, Z., Bhatti, U. A., et al. (2024). Aiot-driven multi-source sensor emission monitoring and forecasting using multi-source sensor integration with reduced noise series decomposition. *Journal of Cloud Computing*, 13:65.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brousse, O., Simpson, C. H., Poorthuis, A., et al. (2024). Unequal distributions of crowdsourced weather data in england and wales. *Nature Communications*, 15:4828.

Camps-Valls, G., Fernández-Torres, M. A., Cohrs, K. H., et al. (2025). Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, 16:1919.

Fick, S. E. and Hijmans, R. J. (2017). Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12):4302–4315.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hubbard, K. and Lin, X. (2006). Reexamination of instrument change effects in the us historical climatology network. *Geophysical research letters*, 33(15).

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Ismaeel, A., Tai, A. P. K., Santos, E. G., et al. (2024). Patterns of tropical forest understory temperatures. *Nature Communications*, 15:549.

Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on neural networks*, 22(9):1341–1356.

Kothiyal, S., Goel, S., Singh, R., Chilwal, A., and Ranjan, R. (2025). Sarima-based time series analysis of rainfall and temperature for the tarai region of uttarakhand. *Journal of Agrometeorology*, 27(1):73–76.

Lai, C.-Y., Hassanzadeh, P., Sheshadri, A., Sonnewald, M., Ferrari, R., and Balaji, V. (2025). Machine learning for climate physics and simulations. *Annual Review of Condensed Matter Physics*, 16:343–365.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.

Lopez-Gomez, I., Wan, Z. Y., Zepeda-Núñez, L., Schneider, T., Anderson, J., and Sha, F. (2025). Dynamical-generative downscaling of climate model ensembles. *Proceedings of the National Academy of Sciences*, 122(17):e2420288122.

McCulloch, M. T., Winter, A., Sherman, C. E., et al. (2024). 300 years of sclerosponge thermometry shows global warming has exceeded 1.5 °c. *Nature Climate Change*, 14:171–177.

Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N. (2021). Era5-land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383.

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly weather review*, 116(12):2417–2424.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part i—a discussion of principles. *Journal of hydrology*, 10(3):282–290.

Nourani, V., Tosan, M., Huang, J. J., Gebremichael, M., Kantoush, S. A., and Dastourani, M. (2025). Advances in multi-source data fusion for precipitation estimation: remote sensing and machine learning perspectives. *Earth-Science Reviews*, page 105253.

Rains, D., Trigo, I., Dutra, E., Ermida, S., Ghent, D., Hulsman, P., Gómez-Dans, J., and Miralles, D. G. (2024). High-resolution (1 km) all-sky net radiation over europe enabled by the merging of land surface temperature retrievals from geostationary and polar-orbiting satellites. *Earth System Science Data*, 16:567–593.

Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4):e1249.

Silván-Cárdenas, J. and Wang, L. (2010). Retrieval of subpixel tamarix canopy cover from landsat data along the forgotten river using linear and nonlinear spectral mixture models. *Remote Sensing of Environment*, 114(8):1777–1790.

Tamasauskas, L. d. O., Pereira, W. G., Negreiros, W. J., Guimaraes, P. H. d. V., Dias, J. A., Corrêa, A. B., Costa, G. B., and Seruffo, M. C. d. R. (2025). Comparison of lstm and sarima models for air temperature forecasting in belém, amazônia, pará. In *Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA)*, pages 256–265. SBC.

van Tiggelen, M., Smeets, P. C. J. P., Reijmer, C. H., Kuipers Munneke, P., and van den Broeke, M. R. (2025). Imau antarctic automatic weather station data, including surface radiation balance (1995–2022). *Earth System Science Data*, 17:4933–4955.

Wei, S., Li, Q., Xu, Q., Li, Z., Zhang, H., and Lin, J. (2025). Updates to c-lsat 2.1 and the development of high-resolution land surface air temperature and diurnal temperature range datasets. *Earth System Science Data*, 17:4985–5005.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press.

Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.

Wu, H. and Li, W. (2019). Downscaling land surface temperatures using a random forest regression model with multitype predictor variables. *Ieee Access*, 7:21904–21916.