Scilight

Regional Ecology and Management
https://www.sciltp.com/journals/rem



Article

Check for updates

Evaluating ERA5 Correction Methods in the Brazilian Cerrado Using Spatial Validation and Predictor Similarity

Elis Dener Lima Alves

Goiano Federal Institute of Education, Science and Technology—Ceres Campus, Ceres, GO 76300-000, Brazil; elis.alves@ifgoiano.edu.br

How To Cite: Alves, E.D.L. Evaluating ERA5 Correction Methods in the Brazilian Cerrado Using Spatial Validation and Predictor Similarity. *Regional Ecology and Management* **2026**, *1*(1), 14. https://doi.org/10.53941/rem.2026.100014

Received: 9 October 2026
Revised: 16 October 2026
Accepted: 3 November 2026
Published: 10 November 2026

Abstract: Five methods for correcting daily ERA5 air temperature and relative humidity in the Brazilian Cerrado were evaluated against INMET observations during 2001–2024 using spatial and spatiotemporal validation. The primary assessment included 99 temperature stations and 98 humidity stations. For temperature, monthly linear scaling and linear regression reduced the median station-level root mean square error (RMSE) by 8.08% and 7.77%, respectively. Linear scaling was retained as an interpretable regional monthly adjustment, without a claim of statistical superiority over regression. Performance varied geographically: its median station RMSE changes ranged from a 2.55% deterioration to a 14.36% reduction across five blocks. Spatial-weight sensitivity showed similar positive residual dependence for both methods. Relative humidity gains were small and lacked consistent concordance improvements, so no regional correction was applied. Temperature correction produced 23,510,412 cell-day estimates over 2682 cells, decreasing mean temperature from 25.17 to 24.69 °C. A temperature-and-seasonality similarity screen included 96.19% of these estimates. At held-out stations, median dissimilarity had a near-zero association with RMSE (Spearman ρ = 0.026), limiting its interpretation as a reliability measure. The original constant-width 3.877 °C error interval describes pooled residual dispersion. A separate nested spatial calibration achieved 90.56% overall coverage, but coverage fell to 84.13% in one block and 85.48% in the warmest training-defined temperature stratum. Regional mean improvements therefore coexist with geographic and climatic limits to transferability. Results are based on daily means screened using an 80% hourly-completeness requirement (at least 20 distinct valid hours per station–variable–UTC day).

Keywords: air temperature; relative humidity; atmospheric reanalysis; systematic bias; concordance correlation coefficient

1. Introduction

Atmospheric reanalyzes provide continuous climate time series on regular spatial grids and are used especially in areas where weather station networks are sparse or have gaps in their records. ERA5 combines observations and atmospheric modeling on a global grid and provides variables at high temporal resolution. However, a grid cell represents an area, whereas a station records conditions at a specific point. Differences in resolution, terrain, elevation, and land cover can therefore lead to discrepancies between the two sources.

In the Brazilian Cerrado, this correspondence is affected by the biome's extent and environmental diversity. The region includes grassland, savanna, and forest formations [1], as well as regional differences in terrain, temperature, and precipitation [2]. Annual MapBiomas Collection 10 maps document the mosaic of native vegetation, cropland,

Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

and pasture over 1985–2024 [3]. These land-cover contrasts increase the possibility that an ERA5 grid cell encompasses conditions that differ from those observed at the station.

Air temperature and relative humidity may also respond differently to correction procedures. Temperature discrepancies may reflect relatively stable differences in mean conditions and seasonality. Relative humidity, in contrast, depends on temperature and vapor pressure and can therefore vary without an equivalent change in water vapor content [4]. The effectiveness of a correction method must consequently be assessed separately for each variable.

Recent regional ERA5-family applications emphasize that performance depends on the variable and setting. Dos Santos Ferreira et al. [5] compared ERA5 and ERA5-Land daily maximum and minimum temperatures with stations in Brazil's semiarid region. Zhang et al. [6] used ERA5-Land precipitation and soil saturation in regional landslide warning thresholds in Italy. These applications provide context for assessing regional reanalysis reliability, while addressing different variables and uses from the daily mean temperature and humidity corrections examined here.

Recent comparisons also address monthly ERA5-Land temperature in Pernambuco [7], daily and monthly meteorological variables across Brazil [8], and temperature correction methods in the Central Himalaya [9]. These studies motivate assessing correction performance for the intended variable, temporal scale, and validation setting, rather than assuming a method ranking transfers between regions.

Methods such as linear scaling and linear regression provide relatively simple adjustments, whereas ridge regression, generalized additive models, and random forest can accommodate more flexible relationships and additional predictors. Greater complexity, however, does not guarantee better performance at locations excluded from model fitting. This transferability must be assessed through spatial validation because random splits can yield overly optimistic performance estimates when nearby stations are dependent [10].

Applying a model across the entire grid may involve predictor combinations poorly represented by the stations. Here, the implementation-specific predictor-similarity domain describes support in the predictor space specified for each model, drawing on the applicability concept of Meyer and Pebesma [11]. Passing this screen does not establish local accuracy, nominal interval coverage, or geographic representativeness.

This study aimed to evaluate five methods for regional correction of daily ERA5 air temperature and relative humidity in the Brazilian Cerrado using INMET automatic weather station observations during 2001–2024. Linear scaling, linear regression, ridge regression, a generalized additive model, and random forest were compared using errors, concordance, sensitivity to spatial and spatiotemporal partitions, and residual spatial structure. Regional temperature fields were accompanied by diagnostics of predictor similarity and the dispersion of spatial validation errors. These diagnostics were interpreted separately from evidence of local predictive accuracy.

2. Materials and Methods

2.1. Study Area and Datasets

The study area comprised the Brazilian Cerrado, delineated using the biome boundaries provided by the Brazilian Institute of Geography and Statistics (IBGE) [12]. The biome covers approximately 2 million km^2, predominantly on the Brazilian Central Plateau, and includes grassland, savanna, and forest formations associated with variations in terrain, soils, water availability, and fire regime [1]. Its tropical climate is characterized by alternating wet and dry seasons, with regional differences in air temperature and precipitation arising from the biome's extent and geographical location [2]. Annual land-cover maps represent the heterogeneous mosaic of native vegetation and agricultural cover [3]. These climatic and environmental differences may affect the correspondence between point observations at stations and the mean conditions represented by reanalysis grid cells. The Cerrado boundary, the ERA5 grid, and the distribution of the stations used in this study are shown in Figure 1.

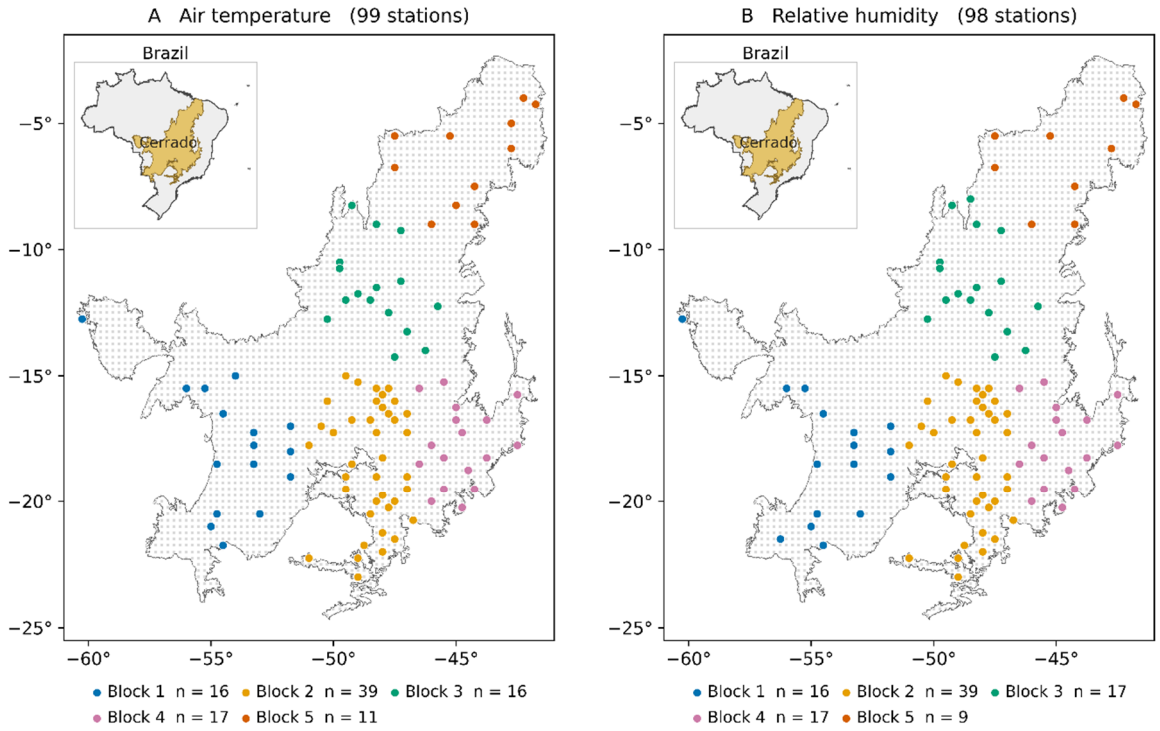


Figure 1. Eligible station networks for air temperature (**A**; 99 stations) and relative humidity (**B**; 98 stations), grouped into the five primary spatial validation blocks. Colored points mark the associated ERA5 grid-cell centers; gray squares show the complete 2682-cell Cerrado grid. Source: Prepared by the author (2026), using IBGE [12] biome boundaries.

Daily air temperature and relative humidity time series were analyzed from 1 January 2001, to 31 December 2024. Reference observations were obtained from the official annual archives of automatic weather stations operated by Brazil's National Institute of Meteorology (INMET) [13]. The hourly dry-bulb air temperature and relative humidity fields were imported as TEM_INS and UMD_INS in UrbanClimaX, and daily means were derived within the workflow rather than taken from a separately supplied daily-mean product. The initial dataset comprised 101 stations linked to their corresponding reanalysis grid cells. After applying the availability and completeness criteria, 99 stations were retained for air temperature and 98 for relative humidity. The primary spatial validation included 491,035 daily pairs for air temperature and 483,257 for relative humidity. These counts include only dates with valid records available simultaneously at the station and its corresponding ERA5 grid cell.

The eligible stations form an uneven observational network rather than a spatially representative probability sample of the Cerrado. Station-wise summaries weight stations equally, without weighting by the area that each location might represent. They therefore describe performance across the available network, not an area-weighted error for the entire biome. Some grid cells lack nearby reference observations, and the range of conditions sampled also depends on the retained station years. Spatial validation evaluates transfer to excluded monitored locations; it does not supply direct validation for every unmonitored cell.

Reanalysis data were obtained from ERA5 at 0.25° × 0.25° resolution [14], through an ARCO ERA5 temporal service with an Open-Meteo fallback. The archived request specifies hourly temperature_2m and relative_humidity_2m, nearest-cell selection, and elevation = nan, which disables the fallback API's elevation downscaling. Each station was linked to the nearest available ERA5 cell center, without spatial interpolation, and paired by UTC date and variable. The complete Cerrado grid contains 2682 cells and the 2001–2024 period contains 8766 days. Retained station identifiers, coordinates, spatial blocks, and paired-day counts are supplied in data/station_network_spatial_2001_2024.csv within the supplementary archive (Section S11). Environmental covariates supported the station-associated spatial models; final monthly scaling required only ERA5 and month.

2.2. Data Preparation and Integration

Inspection of recovered UrbanClimaX source documents implemented import checks, rather than a complete execution history for the original dataset. The reader treats empty and null fields and the sentinels −9999, −9999.0, and −9999.0 as missing, normalizes decimal commas, coerces nonnumeric values to missing, and discards rows

with unparseable timestamps. Acquisition removes repeated station–date–reported-hour keys, keeping the first record. This path implements no additional physical-range or temporal-consistency screening or explicit selection by provider quality flags. Code inspection establishes these implemented operations; it does not verify every historical execution or recover the provider's complete upstream quality-control history. Official provenance alone does not establish that every retained observation passed a separately audited protocol.

In the processing workflow, timestamps were standardized to UTC, and daily temperature and relative humidity were calculated as arithmetic means of available nonmissing hourly observations, separately for each source and variable. Daily station values were retained only after applying the hourly completeness criterion described below. Missing hours were not imputed. Pairing followed daily aggregation, so station and ERA5 values could still represent different contributing hours. Source routines and preprocessing procedures are described in Supplementary Sections S1 and S10.

A strengthened hourly completeness requirement was applied during reanalysis: each station–variable day was required to contain at least 80% of the 24 expected hourly observations within a UTC calendar day, corresponding to at least 20 distinct valid hours. Days with 19 or fewer valid hours were treated as missing and excluded from daily pairings. Temperature and relative humidity were screened separately, and missing hours were not filled. This requirement was applied to the archived sample, and daily means, annual eligibility, model fits, validation diagnostics, and regional products were recomputed using the screened observations.

INMET relative humidity was read from the hourly station humidity field; this is distinct from deriving reanalysis humidity from temperature and dew point. The reanalysis request specifies hourly relative humidity at 2 m before daily averaging. The documented Open-Meteo fallback uses the Magnus relationship below, where T and T_d denote hourly air and dew-point temperatures in °C, with the calculated humidity restricted to 0–100% [15]. The precise ARCO conversion routine, its agreement with this formula, and the conversion used for each original reanalysis record cannot be verified from the recovered products. Equation (1) documents the fallback relationship and does not reconstruct every original record. Averaging hourly humidity also differs from calculating humidity from daily mean temperature and dew point.

$$RH_h = 100 \exp\left[\frac{17.625\, T_{d,h}}{243.04 + T_{d,h}} - \frac{17.625\, T_h}{243.04 + T_h}\right] \quad (1)$$

Daily pairs were organized by station code, grid cell identifier, date, and climate variable. Each record was supplemented with the month, coordinates, grid cell elevation, and annual proportions of native vegetation, cropland, pasture, urban area, water, and other land cover. Grid cell elevation was the elevation at the cell center, derived from the European Space Agency's Copernicus DEM GLO-90 [16], with a resolution of approximately 90 m. Elevation was accessed through the Open-Meteo Elevation API and cached for reproducibility [17]. Elevation values were available for all cells used in the analysis. Station elevations reported by INMET were retained for auditing but were not used as predictors in the regional application because equivalent station information was unavailable for cells without stations. MapBiomas Collection 10 data were matched to the year of each climate record to retain changes in landscape composition over the study period [3].

UrbanClimaX, the author's workflow, implemented data retrieval, station–grid matching, correction, validation and regional application. Supplementary scripts reproduce diagnostics from stored OOF data and refit the stated full-row sensitivities. They do not reconstruct the original capped samples (Supplementary Section S10).

An eligible station–variable year required paired daily values on at least 87% of its calendar days, including the leap-day denominator when applicable. Integration with MapBiomas required at least 90% valid mapped coverage and matched land cover to the same year as the climate record. Each station–variable series then had to contain at least five eligible years and 913 valid daily pairs, and each regional model fit required at least 20 stations. Only records with all model-specific predictors were used. These criteria describe annual coverage and modeling support after the hourly completeness screening described above.

The retained temperature series contain 1820–8706 paired days per station (median 5405), and humidity series contain 1819–8706 (median 5379.5). Both networks have 5–24 eligible years per station, with a median of 15. Paired coverage within those retained calendar years has a median of 99.16% for temperature and 99.14% for humidity; this denominator excludes ineligible years and is not completeness over the entire 24-year period. Station dates, retained-year lists, annual paired-day coverage and blocks are supplied in the supplementary CSVs (Section S11). These describe retained pairs and do not recover excluded-year histories or hourly counts.

2.3. Regional Correction Methods

Five methods were compared for estimating daily values equivalent to INMET observations at ERA5 grid cells. Let $y_{i,t}$ denote the observed value at station i on day t, $x_{i,t}$ the ERA5 value at the associated grid cell, and m_t the month of the record.

Regional linear scaling applied a separate additive adjustment for each month. The mean difference between INMET and ERA5 was estimated using only the calibration data and added to the reanalysis value for the corresponding month, following Teutschbein and Seibert [18] (Equation (2)).

Equation (2) is an unweighted mean over the calibration rows retained for a fit. After subsampling, a station's contribution in a month is its retained row count divided by the total retained count for that month. This differs from equal station weighting. The final scaling fit used 240,000 records from the eligible 99-station temperature network, rather than all 491,035 pairs. The source describes inverse-group-size sampling by station, year, and month, but original sampled-row identifiers were not retained. The final model and twelve monthly coefficients are supplied in the supplementary package and Table S11. The coefficient CSV retains the full precision stored in the original model. Stored coefficients permit application of that fit without reconstructing its unavailable training sample.

$$\hat{y}_{i,t} = x_{i,t} + \overline{d}_{m_t}, \qquad \overline{d}_m = \frac{1}{n_m} \sum_{(i,t)\in m} \left(y_{i,t} - x_{i,t}\right) \tag{2}$$

Here, $\hat{y}_{i,t}$ is the corrected value for station i on day t, $x_{i,t}$ is the original ERA5 value, $\overline{d}_m$ is the mean adjustment estimated for month m, and n_m is the number of valid daily pairs for that month in the calibration dataset.

Month was represented by sine and cosine terms to avoid an artificial discontinuity between December and January (Equation (3)). Regional linear regression estimated observed values from ERA5 and these components of annual seasonality (Equation (4)).

$$s_t = \sin\left[\frac{2\pi(m_t - 1)}{12}\right], \qquad c_t = \cos\left[\frac{2\pi(m_t - 1)}{12}\right] \tag{3}$$

Here, s_t and c_t are the sine and cosine components of seasonality on day t, m_t is the month number, ranging from 1 to 12, and π is the mathematical constant *pi*.

$$\hat{y}_{i,t} = \beta_0 + \beta_1 x_{i,t} + \beta_2 s_t + \beta_3 c_t \tag{4}$$

Here, β_0 is the intercept, β_1 is the coefficient associated with the ERA5 value, and β_2 and β_3 are the coefficients of the cyclic month components.

For the spatial models, $\mathbf{z}_{i,t}$ comprised the ERA5 value, the two cyclic month terms, latitude, longitude, elevation, and the proportions of the six land use and land cover classes. Ridge regression estimated coefficients using an L2 penalty to reduce their instability in the presence of correlated covariates, following Hoerl and Kennard [19] (Equation (5)).

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i,t} \left(y_{i,t} - \beta_0 - \mathbf{z}_{i,t}^{\mathsf{T}}\boldsymbol{\beta}\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\} \tag{5}$$

Here, $\widehat{\boldsymbol{\beta}}$ is the vector of estimated coefficients, $\mathbf{z}_{i,t}$ is the vector containing the record's p predictors, $\boldsymbol{\beta}$ contains their coefficients, and λ controls the strength of the penalty. The intercept β_0 was not penalized.

The spatial generalized additive model (GAM) represented each predictor by a smooth function based on cubic splines (Equation (6)). The sum of these functions allowed nonlinear relationships to be fitted without explicit interactions, following Hastie and Tibshirani [20] and Wood [21].

$$\hat{y}_{i,t} = \beta_0 + \sum_{j=1}^{p} f_j\left(z_{j,i,t}\right) \tag{6}$$

Here, f_j is the smooth function estimated for predictor j, $z_{j,i,t}$ is the value of that predictor at station i on day t, and p is the total number of predictors included in the model.

The spatial random forest combined regression trees constructed from subsets of records and predictors. Predictions were obtained by averaging the B trees in the ensemble, following Breiman [22] (Equation (7)).

$$\hat{y}_{i,t} = \frac{1}{B} \sum_{b=1}^{B} T_b\left(\mathbf{z}_{i,t}\right) \tag{7}$$

Here, B is the number of trees, T_b is the prediction from tree b, and $\mathbf{z}_{i,t}$ is the predictor vector supplied to the model.

The ridge regression penalty was selected from values ranging from 10^{-4} to 10^{4}. For the GAM, bases with 4, 6, 8, and 10 knots were evaluated. Random forests were tested with 100, 200, 300, and 500 trees, a minimum of five observations per terminal leaf, and 80% of the predictors available at each split. Relative humidity predictions were constrained to the range of 0–100%.

Primary model fits used at most 240,000 calibration records. The run metadata described the subsampling as balanced by station, year, and month. In the available source code, this corresponds to sampling without replacement, with record weights inversely proportional to the size of each station–year–month group. This procedure does not impose equal sample sizes across groups or stations. The original sampled row identifiers were not retained, so the exact composition of each training sample could not be verified.

GAM knot counts and random-forest sizes were selected using spatial holdouts within each outer calibration set. These complexity settings were subsequently retained in the repeated-partition and spatiotemporal sensitivity analyses. Regularization selection followed a separate procedure: the exported final ridge and GAM pipelines used RidgeCV with cv = None and scoring = None, which selects the penalty through efficient leave-one-out cross-validation of individual records using mean squared error. This internal procedure does not evaluate spatial transfer. Performance evaluation therefore relied on predictions for stations withheld in the outer spatial validation blocks. The ridge penalties selected within individual primary folds were not exported and could not be independently verified. The final ridge model uses 33 logarithmically spaced penalties from 10^{-4} to 10^{4}. The GAM is an additive cubic B-spline regression followed by scaling and ridge regularization, with 25 penalties over the same range and constant extrapolation beyond boundary knots. Its 4, 6, 8, or 10 knots were compared by inner spatial RMSE, with a 0.5% tolerance favoring a smaller basis. Forests considered 100, 200, 300, and 500 trees and stopped enlargement when the inner spatial RMSE improvement fell below 0.5%, retaining the smaller ensemble. The source master seed is 42. Nine repeated-partition seeds and 44 GAM/RF selection or reuse records are supplied (Supplementary Sections S10 and S11); sampled-row identifiers and fold-specific ridge penalties remain unavailable. Original fitted model states record scikit-learn 1.9.0.

A separate weighting sensitivity analysis refitted monthly scaling within the same five held-out blocks using all available training rows. It compared the ordinary pair mean with equal weights for station-month means, and with equal station weights after first averaging year-month means within each station. These alternatives isolate specified weighting choices on a common full-row estimator; they do not reproduce the original capped sample. The supplementary coefficient and station-metric CSVs distinguish these full-row fits from the capped primary estimator (Supplementary Section S11).

2.4. Spatial Validation and Predictor Similarity

The primary assessment used five-fold spatial cross-validation. K-means grouped stations using their ERA5 grid-cell latitude and longitude after coordinate standardization; longitude was first multiplied by the cosine of mean latitude. Because each coordinate was then standardized separately, this constant longitude factor does not preserve a geographic distance correction. The partition should be interpreted as clustering in standardized coordinate space. In each fold, all records from one geographic block were withheld and the other four blocks supplied calibration. Clustering prioritized geographic grouping without equalizing station counts or enforcing a minimum train–test distance. Station-wise summaries give equal weight to stations, but they do not make the network geographically representative [10]. Actual nearest-training distances and block-specific results were therefore evaluated explicitly.

Performance was calculated separately for each station, variable, and method using bias (Equation (8)), mean absolute error (Equation (9)), root mean square error (Equation (10)), and Lin's concordance correlation coefficient (Equation (11)).

$$\text{Bias} = \frac{1}{n}\sum_{k=1}^{n}(\hat{y}_k - y_k) \tag{8}$$

Here, $\hat{y}_k$ is the estimated value, y_k is the value observed by INMET, and n is the number of valid daily pairs.

$$\text{MAE} = \frac{1}{n}\sum_{k=1}^{n}|\hat{y}_k - y_k| \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\hat{y}_k - y_k)^2} \tag{10}$$

Here, RMSE is the root mean square error, and n is the number of pairs evaluated.

$$\rho_c = \frac{2s_{y\hat{y}}}{s_y^2 + s_{\hat{y}}^2 + \left(\overline{y} - \overline{\hat{y}}\right)^2} \tag{11}$$

Here, ρ_c is Lin's concordance correlation coefficient, $s_{y\hat{y}}$ is the covariance between observed and estimated values, s_y^2 and $s_{\hat{y}}^2$ are their respective variances, and $\overline{y}$ and $\overline{\hat{y}}$ are their means [23].

Percentage RMSE reduction was used to compare each correction with the original ERA5 data (Equation (12)).

$$G_{\mathrm{RMSE}} = 100\left(\frac{\mathrm{RMSE}_{\mathrm{raw}} - \mathrm{RMSE}_{\mathrm{corrected}}}{\mathrm{RMSE}_{\mathrm{raw}}}\right) \tag{12}$$

Here, G_{RMSE} is the percentage reduction, $\mathrm{RMSE}_{\mathrm{raw}}$ is the error of uncorrected ERA5, and $\mathrm{RMSE}_{\mathrm{corrected}}$ is the error after correction. Positive values indicate a reduction in error.

Partition sensitivity used three repetitions for each of 5, 7, and 8 blocks. Empirical 95% semivariance crossing distances of 308.0 km for temperature and 306.9 km for humidity informed the original configuration choice; they were neither identifiable fitted decorrelation ranges nor enforced separation buffers. Repetitions reused the same station network and primary selections of spline knots and forest sizes, so their 5th percentiles are descriptive sensitivity summaries, not confidence limits. Partition diversity was measured using adjusted Rand indices. A separate spatiotemporal analysis coupled each held-out block with one contiguous year group: 2001–2005, 2006–2010, 2011–2015, 2016–2020, or 2021–2024. Training excluded both the spatial block and its associated year group; evaluation used their intersection.

An additional separation sensitivity analysis retained the same five held-out blocks and removed training stations within 150 or 300 km of any held-out station-associated cell. Linear scaling and regression were refitted using all retained daily training pairs, with at least 20 training stations and no complexity tuning. A zero-buffer full-row fit supplied the matched comparator. These refits evaluate geographic separation for the simple methods and differ from the balanced, capped primary estimator. Block summaries are medians of individual station gains; summaries across blocks are explicitly labeled as medians of block medians (Tables S1 and S5).

Spatial dependence of errors was assessed using Moran's *I*, applied separately to the biases of the original ERA5 data and the predictions from each method in the spatial and spatiotemporal validation analyses (Equation (13)).

$$I = \frac{N}{S_0}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(e_i - \bar{e})(e_j - \bar{e})}{\sum_{i=1}^{N}(e_i - \bar{e})2} \tag{13}$$

Here, *I* is Moran's index, *N* is the number of stations, e_i and e_j are the biases at stations *i* and *j*, $\bar{e}$ is the mean bias, w_{ij} is the spatial weight between the stations, and S_0 is the sum of the spatial weights.

The original Moran analysis used row-standardized five-nearest-neighbor weights and 999 permutations, followed by Benjamini–Hochberg adjustment [24,25]. Sensitivity was assessed with k = 3, 5, 8, and 10 directed neighbors, symmetrized k = 5, and 300 and 500 km distance bands, using great-circle distances at both ERA5 cell centers and INMET locations. Within each strategy and variable, 9999 common permutations were generated using PCG64 with seed 20261017 and station-code ordering. The two-sided statistic was centered on $E[I] = -1/(N-1)$; adjustment covered original ERA5 and five corrections within each coordinate and weight configuration, rather than across sensitivity graphs. Distance-band islands retained zero-weight rows and were counted explicitly. Statistical classifications were interpreted with effect sizes; a difference in *q*-value classifications is not a test of a difference between methods or proof of spatial independence.

The spatial structure of station-mean bias was described using an empirical semivariogram (Equation (14)).

$$\gamma(h) = \frac{1}{2N(h)}\sum_{(i,j)\in N(h)}(e_i - e_j)2 \tag{14}$$

Here, $\gamma(h)$ is the semivariance for distance class h, $N(h)$ is the number of station pairs in that class, and e_i and e_j are the biases observed at the stations.

The empirical crossing was the median distance of the first distance class whose mean semivariance reached at least 95% of the population variance of station biases [26]. Sensitivity was examined using 5, 10, 15, and 20

approximately equal-count distance classes. This bin-dependent summary is distinct from a fitted spherical range and was not used as evidence that validation blocks remove dependence.

The dissimilarity index (DI) measured nearest-neighbor distance in model-specific, importance-weighted standardized predictor space against the corresponding outer calibration set. Predictors for linear scaling and linear regression were ERA5 and the sine and cosine of month; the spatial methods also used coordinates, elevation, and land cover. Distances were normalized by the mean positive internal nearest-neighbor distance. The threshold was the internal 95th percentile calculated using comparisons between different stations; reference data were limited to 20,000 rows in validation. Standardization, weighting, reference data, and thresholds were obtained within the calibration data. This adaptation of Meyer and Pebesma's [11] applicability concept uses a different normalization and threshold from their full cross-validation-based procedure; it is termed a predictor-similarity domain here (Equation (15)).

$$\mathrm{DI}_k = \frac{\min_j d(\mathbf{z}_k, \mathbf{z}_j)}{\bar{d}_{\mathrm{cal}}} \tag{15}$$

Here, DI_k is the dissimilarity index for record k, $\mathbf{z}_k$ is its importance-weighted standardized predictor vector, $\mathbf{z}_j$ denotes vectors in the calibration dataset, d is the Euclidean distance, and $\bar{d}_{\mathrm{cal}}$ is the mean of the positive nearest-neighbor distances among the importance-weighted standardized calibration vectors. Record k was considered within the implementation-specific similarity domain when $\mathrm{DI}_k \leq Q_{0.95}(\mathrm{DI}_{\mathrm{cal}})$, where $Q_{0.95}(\mathrm{DI}_{\mathrm{cal}})$ is the internal 95th percentile of calibration dissimilarity to records from a different station. This training-only diagnostic is not itself a validation of predictive accuracy or environmental representativeness.

Held-out DI was compared with absolute daily error, median station DI with station RMSE, and pooled errors inside and outside the fold-specific domain. Spearman correlations and DI/threshold strata were descriptive; no independent-day significance test was used. DI definitions and weights remained those of each trained model. Group contrasts may also reflect differing temperatures, seasons, and observation support, and are not causal effects of dissimilarity (Table S3; Figure S2).

2.5. Regional Application of the Correction and Generation of Climate Fields

The original operational rule combined positive lower summaries of RMSE reduction across the 5-, 7-, and 8-block configurations, improvements in MAE and concordance, and a nonsignificant adjusted Moran test. The revised interpretation does not use the Moran threshold to establish independence or discriminate methods with nearly identical effects. Linear scaling and regression were compared using paired station RMSE differences on identical held-out observations. Exploratory uncertainty was summarized by 10,000 resamples of the five geographic blocks, retaining all stations within each sampled block. These percentile ranges are conditional on the stored predictions and only five observed blocks; they do not establish equivalence or provide a precise population-level confidence interval.

Monthly regional linear scaling was retained for temperature because the primary and partition-sensitivity summaries supported error reduction and because an additive monthly offset has a direct interpretation while preserving ERA5 spatial contrasts and within-month anomalies. Linear regression remained a close alternative, and neither the paired comparison nor the Moran results established statistical superiority of linear scaling. Relative humidity corrections showed small and inconsistent gains across metrics, so no corrected humidity field was produced.

The final temperature model was fitted to eligible station data subject to the 240,000-row balanced sampling limit, and applied to all grid cells and dates. Linear scaling added the corresponding monthly offset to ERA5. Its final similarity screen used ERA5 temperature and the sine and cosine of month, with unit feature weights, a 5000-record reference covering 99 stations, and an internal DI threshold of 2.9415. Latitude, longitude, elevation, and land cover were absent from this final screen. The final coverage percentage therefore describes temperature and seasonal support, rather than geographic or environmental representativeness. Records with missing required predictors were not imputed.

The original regional error layer used the 5th and 95th percentiles of pooled spatial out-of-fold residuals. With e = prediction − observation, observation bounds are [prediction − $Q_{95}(e)$, prediction − $Q_{05}(e)$]. Reapplying these endpoints to the same residual sample is descriptive. Independent coverage was examined separately in two outer-block procedures: first, calibrating bounds from training residuals of each stored outer model; second, fitting linear scaling to all training rows and calibrating its bounds from four inner spatial folds, entirely within the outer training data. The second procedure evaluates a new full-row estimator and does not validate or replace the original constant-width regional layer. Coverage and widths were summarized by block, month, station, predictor domain,

and temperature strata defined by the outer-training ERA5 20th and 80th percentiles. No regional interval map was regenerated from the nested procedure.

Daily time series were aggregated by grid cell to generate fields of mean ERA5 temperature, mean corrected temperature, mean adjustment, dissimilarity, the proportion of records passing the predictor-similarity screen, and empirical error-interval width. The predictor-support indicator records similarity in the specified predictors and does not assign a probability of predictive accuracy.

3. Results and Discussion

3.1. Method Performance in Spatial Validation

The primary spatial validation included 491,035 daily pairs for air temperature at 99 stations and 483,257 pairs for relative humidity at 98 stations. Absolute metrics are reported as medians across stations, whereas improvements are the medians of changes calculated separately at each station. Because these two summaries are obtained through different procedures, the median improvement does not necessarily equal the percentage difference between the median errors of the original ERA5 data and the corrected time series.

For air temperature, regional linear scaling achieved the largest median RMSE reduction, at 8.08%, followed by regional linear regression at 7.77%. MAE reductions were 14.06% and 14.39%, respectively. Median bias decreased from 0.507 °C in the original ERA5 data to 0.039 °C with both methods. Concordance increased from 0.848 to 0.872 with linear scaling and to 0.870 with linear regression.

Ridge regression produced the lowest median absolute errors after correction, with an RMSE of 1.243 °C and an MAE of 0.912 °C, but its median improvements were smaller than those of the regional methods: 4.39% for RMSE, 8.69% for MAE, and 0.0068 for concordance. Random forest reduced RMSE by 0.24% and MAE by 2.64%, with a change in concordance close to zero. The spatial GAM increased RMSE by 16.58% and MAE by 21.72% and reduced concordance. Thus, greater functional complexity and the inclusion of location, elevation, and land cover did not systematically improve performance relative to methods based on ERA5 and seasonality.

For relative humidity, random forest achieved the largest median error reductions in the primary validation, at 1.83% for RMSE and 2.37% for MAE, but reduced concordance by 0.0025. Linear regression reduced RMSE by 1.39% and MAE by 0.49%, while concordance decreased by 0.0039. Linear scaling reduced RMSE by 0.49% and MAE by 0.68%, with a change in concordance close to zero. Ridge regression increased MAE by 1.24%, and the spatial GAM increased RMSE by 4.83% and MAE by 7.36%.

Temperature corrections based on ERA5 and seasonality improved aggregate RMSE, MAE, bias, and concordance. Their ranking depends on the metric and aggregation: linear scaling had the larger median individual RMSE reduction, whereas linear regression had a slightly lower median absolute RMSE. Neither a difference of these medians nor a thresholded Moran test establishes superiority. For relative humidity, the small error reductions and inconsistent concordance changes provided limited support for regional application.

For full-row temperature scaling, median station RMSE gains were 8.16% with pair weighting, 7.77% with equal station-month weighting, and 7.70% after balancing years within stations; the proportions improving were 63.64%, 64.65%, and 64.65%. Equal station-month estimates differed from the published monthly coefficients by at most 0.0293 °C. These results indicate modest sensitivity of aggregate temperature correction to the tested weights, while individual deterioration persists. The original published estimator and its 8.08% gain were retained.

Araújo et al. [7] reported monthly ERA5-Land temperature RMSEs of 0.47–1.98 °C across 13 INMET stations in Pernambuco. This variation among stations is consistent with the heterogeneous benefits observed here, where scaling improved temperature RMSE at 63 of 99 stations. Their absolute errors cannot be ranked directly against the present daily ERA5 errors because the reanalysis product, temporal averaging, observation support, and validation design differ.

Brown et al. [8] ranked ERA5-Land highest overall for air temperature across 11 Brazilian flux-tower sites, while temperature bias varied in sign among sites. Their temperature errors were more strongly associated with correlation at the daily scale and mean bias at the monthly scale. This scale dependence is consistent with the aggregate benefit of a monthly mean adjustment here, but does not show that local daily variability has been corrected. Their humidity assessment concerned specific humidity, so it cannot establish the effectiveness of correcting ERA5 relative humidity.

Temperature block sizes were 16, 39, 16, 17, and 11 stations, contributing 85,623, 205,015, 70,201, 81,224, and 48,972 pairs. Linear scaling median individual station RMSE gains were −2.55%, 11.64%, 3.39%, 14.36%, and 6.49%, respectively (Tables S1 and S2; Figure S1). The largest block contributed 41.8% of daily pairs, and its 39 stations also carry substantial weight in the network-wide station median. Block 1 deteriorated by both the median individual gain (−2.55%) and pooled daily RMSE change (−3.21%). Its slightly lower median absolute

corrected RMSE does not contradict this finding because the order of aggregation differs. Regional performance is therefore geographically heterogeneous.

Across the primary temperature network, scaling improved RMSE at 63 of 99 stations (63.64%); its station-level interquartile range was −6.17% to 17.77%. Regression also improved 63 stations, with an interquartile range of −6.68% to 16.94%. These distributions show that positive aggregate summaries coexist with deterioration at more than one third of stations. Humidity scaling improved 59 of 98 stations (60.20%), with an interquartile range of −2.49% to 3.36% (Figure 2).

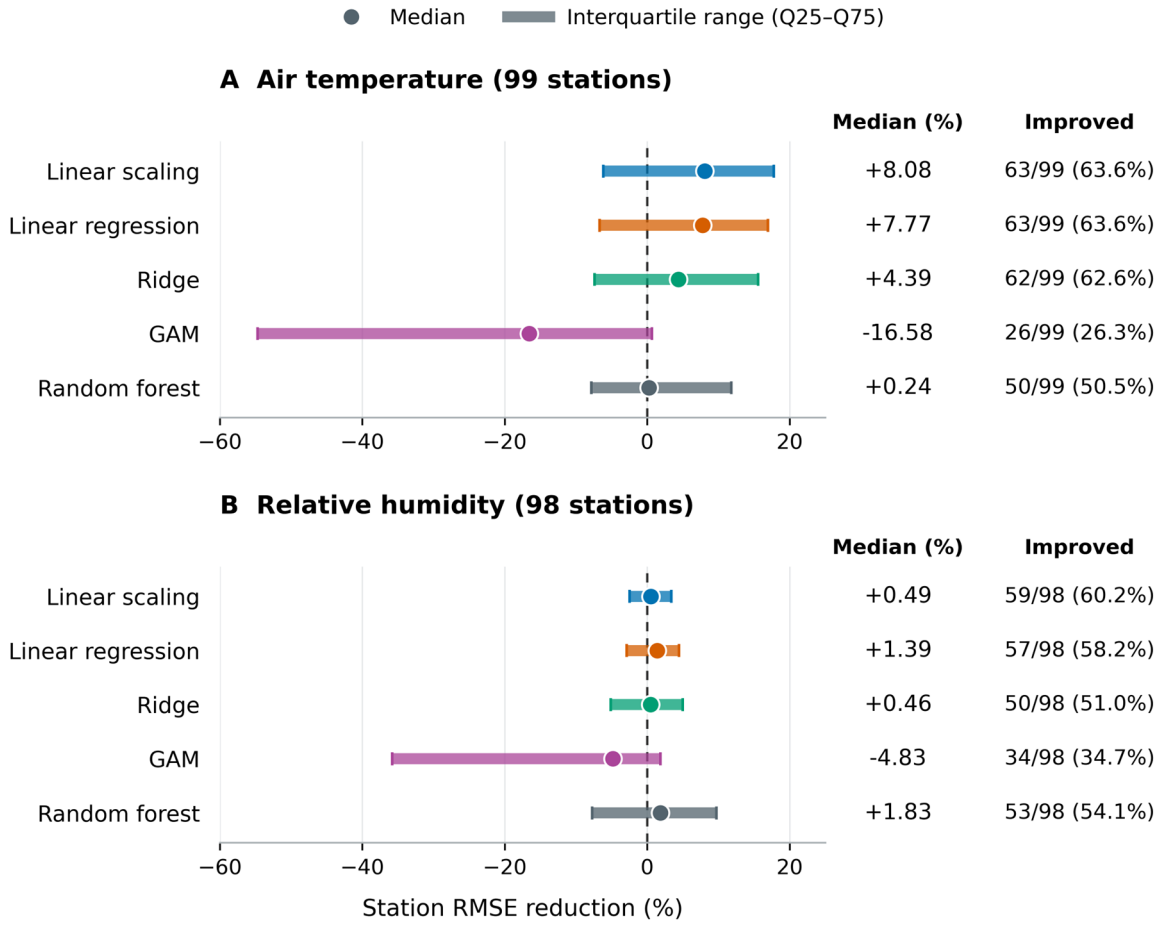


Figure 2. Station-level RMSE reductions in the primary five-block spatial validation for air temperature (**A**; 99 stations) and relative humidity (**B**; 98 stations). Dots show medians and horizontal segments the interquartile range (25th–75th percentiles) across stations. Segments describe the middle 50% of gains rather than confidence intervals or the full range. Source: Prepared by the author (2026).

3.2. Corrections and the Spatial Structure of Errors

Repeated spatial validation gave linear scaling median temperature RMSE reductions of 7.93–8.16% across 5-, 7-, and 8-block configurations (Table 1); the lowest reported 5th percentile was 7.89%. MAE reductions were 13.99–14.38%, and concordance gains were 0.0185–0.0192. Regression achieved RMSE reductions of 7.55–7.82%, a minimum 5th percentile of 7.55%, and MAE reductions of 14.36–14.58%. These summaries describe three repetitions per scale with reused primary spline and forest selections. Two five-block repetitions were identical up to labels (adjusted Rand index 1.000), and their similarity to the third was 0.972. Thus, favorable aggregate sensitivity does not demonstrate independent replication or uniform regional gains (Table S9).

The original five-neighbor Moran values were 0.101 for scaling (q = 0.0555) and 0.106 for regression (q = 0.036). Recomputed shared-permutation tests classified both similarly: centered two-sided q-values at ERA5 coordinates were 0.083 at k = 5, approximately 0.026 at k = 8 and 0.019 at k = 10. Their positive I values remained close (Table S4). The supplied generator, ordering and sequence ensure reproducibility; these results do not establish spatial independence or method superiority.

For context, original ERA5 temperature bias had I = 0.0691 (q = 0.1056) in the original five-neighbor export, compared with I = 0.1007 for scaling and 0.1058 for regression. The new common-permutation sensitivity gave

original ERA5 $I = 0.0690$ and $q = 0.2004$, compared with $I = 0.1006$ and 0.1058 and $q = 0.0825$ for both corrections. Correction reduced aggregate errors without eliminating the spatial organization of station-mean bias. The original BH family likewise comprised ERA5 and all five corrections within each variable and validation strategy (Table S4).

Ridge regression reduced errors in all three configurations, but RMSE reductions remained between 4.06% and 4.18%, below those achieved by the regional methods. Random forest achieved reductions of 0.70–1.38%, with changes in concordance close to zero, whereas the spatial GAM increased RMSE and MAE and reduced concordance at all scales. Greater functional flexibility and the inclusion of elevation and land cover therefore did not improve transferability to held-out stations. This result reinforces the importance of evaluating models at locations excluded from fitting [10] and restricting their application to covariate combinations represented in the training data [11].

The primary GAM temperature RMSE deterioration of 16.58% and the smaller repeated-configuration medians of 3.26–8.54% summarize different fitting and partition runs. In primary block 3, the median individual RMSE gain was −121.40%, and the chosen ten-knot model had a pooled RMSE of 5.003 °C. Repeated medians vary with geographic partitioning and training samples, conditional on reused knot and forest selections. The available results demonstrate GAM instability but do not isolate partitioning from sampling or tuning as a single cause. The primary deterioration is therefore retained explicitly (Table S2a).

Satyapragyan et al. [9] found favorable linear-regression and GAM corrections of ERA5 and ERA5-Land mean temperature at three Himalayan automatic weather stations. This contrasts with the 16.58% median station RMSE increase for the GAM in the present primary spatial assessment. Differences in terrain, predictors, model implementation, calibration samples, and validation design limit direct comparison. The contrast shows why the present GAM result should be interpreted for this implementation and spatial-transfer task, rather than as a universal ranking of correction methods.

For relative humidity, random forest maintained median RMSE reductions of 1.20–1.54% but reduced concordance in every configuration. Linear regression achieved RMSE reductions of 0.98–1.42%, accompanied by MAE reductions ranging from −0.13% to 0.50% and decreased concordance. For linear scaling, RMSE reductions ranged from 0.43% to 0.47%, with small changes in MAE and concordance. Ridge regression had a negative 5th percentile of RMSE reduction and increased MAE, while the GAM increased errors. No method produced simultaneous, stable improvements in RMSE, MAE, and concordance for this variable.

The median paired temperature RMSE difference, scaling minus regression, was +0.00012 °C; 48 stations favored scaling and 51 favored regression. Exploratory five-block resampling gave a 2.5th–97.5th percentile range of −0.00044 to +0.00228 °C for this median difference. This very small conditional range includes zero and is based on only five geographic clusters; it is not evidence of equivalence or a definitive test of generalization. The aggregate median percentage gains favoring scaling and the median paired difference need not yield the same ranking.

Table 1. Variation in performance improvements across repeated spatial validation with 5, 7, and 8 blocks.

Variable	Method	RMSE Reduction (%)	Minimum P5 (%)	MAE Reduction (%)	ΔCCC
Air temperature	Linear scaling	7.93 to 8.16	7.89	13.99 to 14.38	0.0185 to 0.0192
	Linear regression	7.55 to 7.82	7.55	14.36 to 14.58	0.0182 to 0.0184
	Ridge	4.06 to 4.18	3.90	8.95 to 10.45	0.0057 to 0.0065
	GAM	−8.54 to −3.26	−9.56	−10.71 to −5.19	−0.0274 to −0.0124
	Random forest	0.70 to 1.38	0.63	2.18 to 3.66	−0.0007 to 0.0001
Relative humidity	Linear scaling	0.43 to 0.47	0.36	0.10 to 0.68	−0.0005 to 0.0001
	Linear regression	0.98 to 1.42	0.87	−0.13 to 0.50	−0.0040 to −0.0038
	Ridge	0.09 to 0.37	−0.19	−1.50 to −0.99	−0.0060 to −0.0054
	GAM	−6.28 to −2.85	−7.36	−8.89 to −5.45	−0.0153 to −0.0101
	Random forest	1.20 to 1.54	0.73	0.19 to 1.28	−0.0036 to −0.0008

Source: Prepared by the author (2026). Note: Ranges span the lowest and highest median station-level improvements across the three block configurations. Minimum P5 is the smallest 5th percentile calculated across only three repetition-level summaries at a configuration. It is a partition-sensitivity summary, not a confidence limit or a percentile across stations. ΔCCC denotes the change in Lin's concordance correlation coefficient. For three ordered repetition medians $a \le b \le c$, the linearly interpolated 5th percentile is $0.9a + 0.1b$.

The spatiotemporal analysis evaluated 87,900 pairs at 68 temperature stations and 83,536 pairs at 65 humidity stations, rather than the full primary spatial sample (Table S8). Temperature median RMSE reductions were 15.37% for regression and 15.30% for scaling, with MAE reductions of 18.67% and 19.32% and concordance gains of 0.0382 and 0.0384. Ridge reduced RMSE by 12.17% and MAE by 18.53%; random forest reduced them by 2.43% and 5.19%. GAM increased RMSE by 26.07%. Because the station–date support differs, these gains cannot be interpreted as a directly comparable increase over the primary spatial performance.

Spatiotemporal results are consistent with a systematic temperature discrepancy, while their interpretation remains conditional on the smaller, jointly held-out sample and reused primary complexities. For humidity, random forest reduced RMSE by 1.93% and MAE by 2.51%, with a concordance decrease of 0.0048. Regression reduced RMSE by 1.64% and MAE by 1.46%, also with decreased concordance. The different temporal support and sparse early-year blocks further limit generalization.

Figure 3 shows the semivariance of the original ERA5 bias as a function of distance between stations. The analysis included 4851 pairs for air temperature and 4753 for relative humidity. Gray points represent station pairs, and colored points represent means for ten distance classes. For readability, the vertical axis of the semivariogram cloud was limited to the 95th percentile of semivariance; all pairs were retained in the calculations. The spherical curve is shown only for relative humidity, for which the model parameters were identifiable.

For air temperature, the empirical sill was 0.422 °C^2, and the operational threshold-crossing distance was 308.0 km. Only one distance class lay below this distance. The nugget and range of a spherical model could not be identified separately, so the fitted range was not used for interpretation. The value of 308.0 km means only that semivariance reached 95% of the empirical sill in the second distance class; it does not identify a reliable decorrelation distance.

For relative humidity, the population variance of station bias was 12.621 pp^2 (pp denotes percentage points), and the empirical crossing was 306.9 km. The descriptive spherical fit had a nugget of 10.406 pp^2, a sill of 14.610 pp^2, a range of 1763.6 km, and R^2 = 0.646. High short-distance semivariance indicates heterogeneous nearby station biases; the longer fitted range describes additional regional structure. The fitted range is a descriptive point estimate from ten binned class means. No range confidence interval was estimated; shared stations make the pairwise semivariances dependent, and the apparent numeric precision should not be interpreted as a precise physical dependence scale.

Changing the number of distance classes between 5 and 20 yielded temperature crossings ranging from approximately 210 to 308 km and humidity crossings from 244 to 675 km (Table S9). Thus, the approximately 307 km values from ten classes are sensitive empirical summaries. The temperature fitted range remains unidentifiable, and the humidity spherical range describes a different property. These observations support interpreting the partition scales as heuristics and reporting actual separation and buffer sensitivity.

Minimum nearest-training distances across the original temperature blocks were only 76.6–140.0 km (Table S1). Thus, the five blocks did not enforce a separation near 308 km. In the additional full-row refits, linear scaling median block-median gains were 6.51%, 6.50%, and 6.50% under 0, 150, and 300 km buffers; the worst block remained negative at −2.63%, −3.04%, and −3.43%. At 300 km, 36–84 temperature training stations remained. The buffer check supports an aggregate adjustment under greater separation while preserving the evidence of geographic heterogeneity. These block-level summaries use a distinct estimator and aggregation from the primary 8.08% network statistic (Table S5).

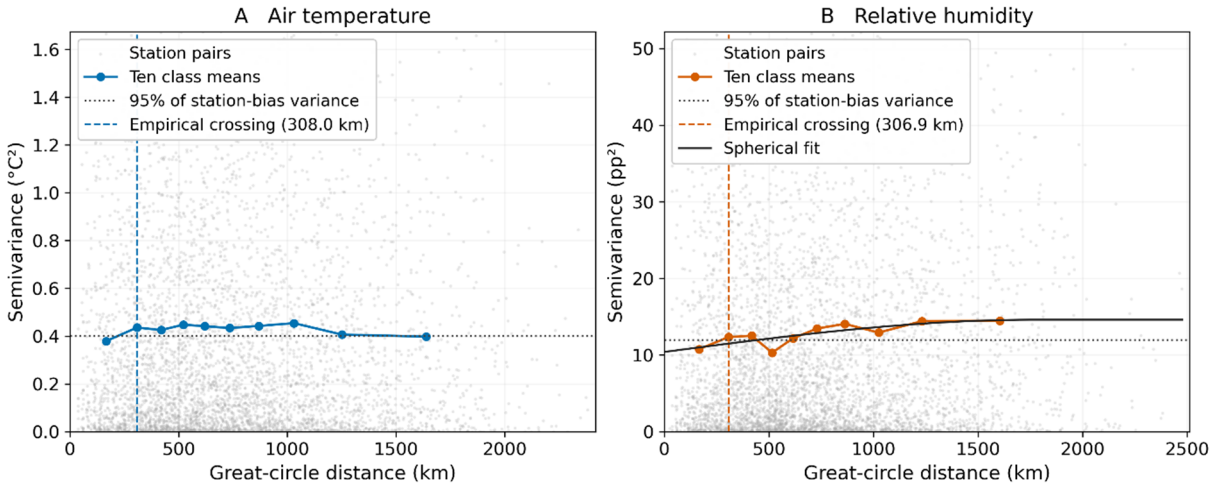


Figure 3. Semivariogram clouds and means for ten approximately equal-count distance classes, based on original ERA5 station-mean bias. Vertical lines show empirical 95% crossings, not fitted decorrelation ranges; horizontal lines show 95% of the station-bias population variance. Source: Prepared by the author (2026).

3.3. Regional Correction and Coverage of the Predictor Similarity Domain

Monthly scaling was retained as a transparent regional temperature adjustment with favorable aggregate performance. The paired comparison and spatial-weight sensitivity did not establish its superiority over regression. The negative gain in one geographic block and the poorer coverage in warm conditions show that this choice cannot imply improvement at every station or location. Humidity was not corrected because its error reductions lacked consistent concordance gains.

The model was applied to all 8766 days for the 2682 ERA5 grid cells within the Cerrado, yielding 23,510,412 cell-day estimates. Mean ERA5 temperature ranged from 18.66 to 29.46 °C across grid cells, with an overall mean of 25.17 °C. After correction, values ranged from 18.18 to 28.97 °C, with an overall mean of 24.69 °C (Figure 4A,B).

The annual mean adjustment was −0.482 °C and was identical across all 2682 grid cells because the selected method applies the same monthly offset throughout the region. Monthly adjustments ranged from −0.325 °C in November to −0.648 °C in June. The correction therefore varies seasonally while fully preserving the relative spatial pattern of ERA5 within each month. The corrected field should be interpreted as a regional monthly adjustment to the mean, rather than a spatial adjustment determined by elevation or land cover.

The uneven station distribution limits the geographic representativeness of these fields. Spatial block validation reduces dependence between calibration and evaluation locations but does not fill observational gaps. In sparsely monitored areas, location-specific bias remains unverified even when ERA5 temperature and seasonality fall within the similarity domain. Equal station weighting also cannot balance the area or environmental diversity represented by densely and sparsely sampled regions. The observed block differences support reporting network performance with explicit limits on transfer to unmonitored locations; further observations should target underrepresented climatic, elevation, and land-cover conditions.

The predictor-similarity screen included 96.19% of all cell-day estimates. The proportion passing this screen had a median of 97.78% per grid cell and ranged from 56.24% to 99.77% (Figure 4C). Of the 2682 cells, 2094 had at least 95% of days passing the screen and 2484 had at least 90%; 43 cells remained below 80%. The lowest proportions occurred mainly at the southernmost edge and in parts of the northernmost Cerrado. These percentages refer to ERA5 temperature and seasonal predictors only. They do not demonstrate that 96.19% of estimates are accurate or that error is spatially uniform.

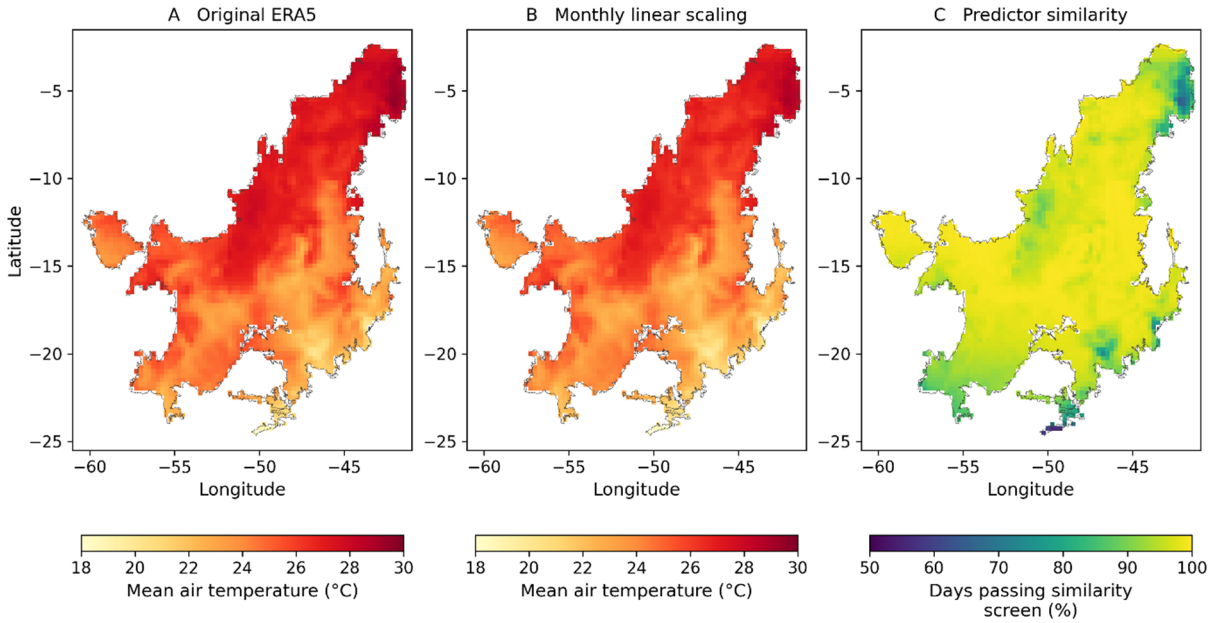


Figure 4. Original (**A**) and corrected (**B**) mean air temperature and percentage of days passing the implementation-specific predictor-similarity screen (**C**) on the ERA5 grid in the Brazilian Cerrado, 2001–2024. Panel **C** describes support in ERA5 temperature and seasonal predictor space. Source: Prepared by the author (2026).

3.4. Dissimilarity and Empirical Interval Coverage

The mean dissimilarity index across grid cells had a median of 0.691 and a mean of 0.960, with values ranging from 0.457 to 14.915. The mean exceeding the median, together with the high maximum, indicates a distribution

with a long upper tail. This tail corresponds to grid cells whose combinations of ERA5 values and seasonality differed most from those represented by the calibration stations.

At held-out temperature stations, scaling DI had only a weak relation to error: the descriptive daily Spearman correlation with absolute error was 0.030, and the correlation of median station DI with RMSE was 0.026 (99 stations). Pooled RMSE was 1.303 °C inside the fold-specific domain (459,449 pairs) and 1.398 °C outside (31,586 pairs); corresponding MAE values were 0.932 and 1.024 °C. Pooled RMSE increased modestly across DI/threshold strata, while station-median RMSE was not monotonic (Table S3; Figure S2). The higher pooled errors outside the domain constitute a group-level contrast, whereas the near-zero correlations and nonmonotonic station medians show that DI did not reliably rank local predictive error. The screen identifies unusual predictor combinations; it does not calibrate a probability of accuracy or establish uniform geographic reliability. The 96.19% final-grid proportion and the held-out proportion use different calibration references and are not interchangeable.

The original pooled error quantiles were $Q_{05}(e) = -1.973$ °C and $Q_{95}(e) = 1.904$ °C for e = prediction − observation, giving observation bounds of prediction − 1.904 °C to prediction + 1.973 °C. Their width, 3.877 °C, is constant over the grid. Coverage was 90.00% when these quantiles were reapplied to their own out-of-fold residual sample, as expected by construction. This is not independent validation of the published grid interval.

The separate training-residual procedure for the stored outer models achieved 90.10% held-out coverage, with a mean width of 3.941 °C, but used in-sample training residuals for calibration. The fully nested spatial procedure for the full-row estimator achieved 90.56% pooled coverage and 89.76% mean station coverage, with a mean width of 4.004 °C (Tables S6 and S7). Its block coverage ranged from 84.13% to 93.59%; coverage was 93.24% in the lower training-defined temperature stratum and 85.48% in the upper stratum. Aggregate coverage near 90% therefore hides geographic and climatic undercoverage. Neither procedure supplies a formal finite-sample coverage guarantee or validates the original uniform grid interval. Figure S3 summarizes monthly coverage; the supporting CSV files remain subject to the availability check in Supplementary Section S11. The nested result cannot be assigned to the original constant-width grid layer.

The regional studies of dos Santos Ferreira et al. [5] and Zhang et al. [6] concern different variables and decision settings. Their relevance here is the need to assess a reanalysis product for its intended regional application. They do not independently validate the present daily mean ERA5 correction, similarity screen, or prediction intervals.

The validation results are conditional on the quality and temporal representativeness of the station-derived daily reference. If missing hours are concentrated in warmer, cooler, drier, or more humid parts of the day, an available-observation mean can differ systematically from a complete daily mean, even when annual daily coverage is high. Such differences could affect estimated ERA5 bias, monthly offsets, and validation errors. The applied 20-h minimum excluded more severely incomplete days, but still allowed up to four missing hours and did not by itself ensure unbiased sampling of the diurnal cycle. Spatial validation, the similarity screen, and interval analyses do not resolve this observational limitation.

The corrected temperature and unchanged humidity were validated separately, so their combination is not a jointly calibrated thermodynamic product. At a fixed RH, replacing T by corrected T changes the vapor pressure implied by RH × $e_s(T)/100$, where e_s is saturation vapor pressure; implied dew point and humidity-dependent indices can therefore change. In addition, nonlinear quantities calculated from daily means generally differ from averages of hourly quantities [4]. Applications involving vapor pressure, evapotranspiration, or heat stress require a consistent humidity calculation and separate validation. Correcting temperature alone does not demonstrate improved joint moisture consistency.

4. Conclusions

Across the available station network, monthly linear scaling reduced median individual station RMSE by 8.08% and MAE by 14.06%, with a concordance gain of 0.0194. The repeated partitions supported aggregate temperature error reduction, while block-specific RMSE changes ranged from a 2.55% deterioration to a 14.36% improvement. Paired comparisons and residual spatial-weight sensitivity did not establish superiority over linear regression. Linear scaling is therefore an operational additive adjustment with geographically variable benefits. RMSE improved at 63 of 99 temperature stations (63.64%), with a station-level interquartile gain range of −6.17% to 17.77%; the positive network summary therefore coexists with deterioration at 36 stations.

For relative humidity, no method simultaneously and consistently improved RMSE, MAE, and concordance. Random forest achieved the largest error reductions in the main validation but decreased concordance. Linear regression and linear scaling maintained small RMSE reductions, also without consistent improvements in the other metrics. The correction was therefore not applied to the regional relative humidity grid.

The temperature correction covered 2682 ERA5 grid cells and 8766 days between 2001 and 2024, yielding 23,510,412 estimates. Mean temperature across grid cells decreased from 25.17 °C in the original ERA5 data to 24.69 °C after correction. The annual mean adjustment was −0.482 °C in every grid cell, with monthly adjustments ranging from −0.325 to −0.648 °C. The final product preserves ERA5 spatial contrasts while seasonally adjusting its mean level; it does not resolve location-specific terrain or land-cover bias.

The similarity screen included 96.19% of corrected cell-day estimates, but this percentage describes support in ERA5 temperature and seasonal predictors, not the proportion of accurate predictions. The near-zero associations of DI with daily absolute error ($\rho = 0.030$) and station RMSE ($\rho = 0.026$), together with nonmonotonic station-median errors, show limited ability to rank local error. Higher pooled errors outside the held-out domain do not overcome this limitation or establish uniform reliability within the domain. Complete regional coverage therefore does not demonstrate accuracy at unmonitored locations or transferability across unrepresented environmental conditions.

The original 3.877 °C layer describes pooled error dispersion. A separate nested spatial calibration achieved 90.56% aggregate coverage, falling to 84.13% in one block and 85.48% in the upper temperature stratum; local uncertainty should therefore not be inferred from a common width. All reported performance estimates and regional fields derive from daily references screened using a minimum of 80% of expected hourly observations, requiring at least 20 distinct valid hours per station–variable–UTC day. Additional observations in underrepresented environments and evaluation of vapor-pressure or specific-humidity correction could strengthen future applications. The present results characterize the available station network and do not establish accuracy at every unmonitored Cerrado cell.

Supplementary Materials

The additional data and information can be downloaded at: https://media.sciltp.com/articles/others/261110085550872 9/REM-26100067-SI.zip. Tables S1–S11 and Figures S1–S3. Materiais_suplementares_Revisor_2.zip provides original OOF data and 18 CSVs, station/block and retained-year indexes, the final monthly coefficients and model, and scripts and derived files for validation diagnostics and full-row sensitivities. The README maps files to reported results. Primary capped-model performance is evaluated from stored predictions; original sampled-row identifiers, fold-specific penalty states, hourly counts and complete upstream quality-control histories remain unavailable (Supplementary Sections S10–S12; Table S10).

Funding

This research received financial support from the Federal Institute of Education, Science and Technology Goiano (IF Goiano).

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting the findings of this study are available from the author upon request at elis.alves@ifgoiano.edu.br. Public input datasets are identified in Section 2.1 and in the References. Additional methodological information is provided in the Supplementary Materials. Limitations on retained records and reproducibility are described in Supplementary Sections S10–S12.

Acknowledgments

The author acknowledges the Federal Institute of Education, Science and Technology Goiano for the financial support provided for this research.

Conflicts of Interest

The author declares no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this manuscript, the author used ChatGPT (OpenAI) solely to assist with language editing and improve the clarity and readability of the text. The author takes full responsibility for the final content of the manuscript.

References

1. Ribeiro, J.F.; Walter, B.M.T. As principais fitofisionomias do bioma Cerrado. In *Cerrado: Ecologia e Flora*; Sano, S.M., Almeida, S.P., Ribeiro, J.F., Eds.; Embrapa Informação Tecnológica: Brasília, DF, Brazil, 2008; pp. 151–212.
2. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; et al. Köppen's climate classification map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. https://doi.org/10.1127/0941-2948/2013/0507.
3. Projeto MapBiomas. *Coleção 10 da Série Anual de Mapas de Cobertura e Uso da Terra do Brasil*; MapBiomas: São Paulo, Brazil, 2025.
4. Lawrence, M.G. The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications. *Bull. Am. Meteorol. Soc.* **2005**, *86*, 225–234. https://doi.org/10.1175/bams-86-2-225.
5. dos Santos Ferreira, L.; Alcantara, C.R.; de Barros Santiago, D.; et al. Validation of ERA5 and ERA5-Land reanalysis data for maximum and minimum temperatures in the Brazilian semiarid region. *Theor. Appl. Climatol.* **2026**, *157*, 198. https://doi.org/10.1007/s00704-026-06057-6.
6. Zhang, S.; Pecoraro, G.; Huang, D.; et al. From hydro-meteorological thresholds towards an operational warning model for landslides at regional scale: A real-case application. *Eng. Geol.* **2026**, *362*, 108542. https://doi.org/10.1016/j.enggeo.2026.108542.
7. de Araújo, C.S.P.; Silva, J.R.I.; de Almeida, C.D.G.C. Total precipitation and air temperature data obtained from ERA5-Land reanalysis: Validation and accuracy. *Rev. Caatinga* **2025**, *38*, e12578. https://doi.org/10.1590/1983-21252025v3812578rc.
8. Brown, J.R.C.; Woods, R.; da Rocha, H.R.; et al. Evaluation of high-resolution meteorological data products using flux tower observations across Brazil. *Hydrol. Earth Syst. Sci.* **2026**, *30*, 141–162. https://doi.org/10.5194/hess-30-141-2026.
9. Satyapragyan, S.; Yadav, J.S.; Bhambri, R. Bias corrections of ERA5 and ERA5-Land temperature using automatic weather station data in the Higher Central Himalaya: Implications for hydro-meteorological and glaciological research. *J. Hydrol. Reg. Stud.* **2026**, *63*, 103079. https://doi.org/10.1016/j.ejrh.2025.103079.
10. Roberts, D.R.; Bahn, V.; Ciuti, S.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. https://doi.org/10.1111/ecog.02881.
11. Meyer, H.; Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **2021**, *12*, 1620–1633. https://doi.org/10.1111/2041-210x.13650.
12. IBGE. *Biomas e Sistema Costeiro-Marinho do Brasil: Compatível Com a Escala 1:250 000;* Instituto Brasileiro de Geografia e Estatística: Rio de Janeiro, Brazil, 2025.
13. INMET. *Dados Históricos Anuais de Estações Automáticas*; Instituto Nacional de Meteorologia: Brasília, Brazil, 2026. Available online: https://portal.inmet.gov.br/dadoshistoricos/ (accessed on 5 November 2026).
14. Hersbach, H.; Bell, B.; Berrisford, P.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. https://doi.org/10.1002/qj.3803.
15. Open-Meteo. Historical Weather API and Temperature–Dew-Point Humidity Calculation. Available online: https://open-meteo.com/en/docs/historical-weather-api and https://github.com/open-meteo/open-meteo/blob/main/Sources/App/Helper/Meteorology.swift (accessed on 17 October 2026).
16. European Space Agency (ESA). *Copernicus Digital Elevation Model*; ESA: Paris, France, 2021. https://doi.org/10.5270/ESA-c5d3d65.
17. Open-Meteo. Elevation API. Available online: https://open-meteo.com/en/docs/elevation-api (accessed on 2 September 2026).
18. Teutschbein, C.; Seibert, J. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.* **2012**, *456–457*, 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052.
19. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **1970**, *12*, 55–67. https://doi.org/10.1080/00401706.1970.10488634.
20. Hastie, T.; Tibshirani, R. Generalized Additive Models. *Stat. Sci.* **1986**, *1*, 297–310. https://doi.org/10.1214/ss/1177013604.
21. Wood, S.N. *Generalized Additive Models: An Introduction with R*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017. https://doi.org/10.1201/9781315370279.
22. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. https://doi.org/10.1023/a:1010933404324.
23. Lin, L.I. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **1989**, *45*, 255–268. https://doi.org/10.2307/2532051.

24. Moran, P.A.P. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23. https://doi.org/10.1093/biomet/37.1-2.17.
25. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1995**, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
26. Cressie, N.A.C. *Statistics for Spatial Data*, revised ed.; Wiley: New York, NY, USA, 1993. https://doi.org/10.1002/9781119115151.