Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Weather-station THI underestimates night-time heat load in free-stall dairy barns: hourly agreement across 172 sensors in Brazil and Uruguay

Diego Azevedo Mota[1]

diego.mota@ufvjm.edu.br

Universidade Federal dos Vales do Jequitinhonha e Mucuri

Thiago Vasconcelos Melo[1]

Universidade Federal dos Vales do Jequitinhonha e Mucuri

Felipe Nogueira Domingues[1]

Universidade Federal dos Vales do Jequitinhonha e Mucuri

Research Article

Keywords: bioclimatology, thermal comfort, heat stress, free stall, embedded sensors, THI

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11193676/v1

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

Additional Declarations: No competing interests reported.

Weather-station THI underestimates night-time heat load in free-stall dairy barns: hourly agreement across 172 sensors in Brazil and Uruguay

Diego Azevedo Mota[1], Thiago Vasconcelos Melo[1], Felipe Nogueira Domingues[1]

[1] Instituto de Ciências Agrárias, Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), Unaí, Minas Gerais, Brazil

Corresponding author: Diego Azevedo Mota, diego.mota@ufvjm.edu.br

ORCID: Diego Azevedo Mota 0000-0001-5959-3646; Thiago Vasconcelos Melo 0000-0002-6992-1639; Felipe Nogueira Domingues 0000-0002-6809-5320

Abstract

The temperature-humidity index (THI) computed from weather-station data underpins much of the heat-stress literature in dairy cattle, although cows live in the barn microenvironment. We assessed agreement between the two sources using 372,633 paired hourly records from 172 sensors installed in free-stall barns across 30 municipalities in Brazil and Uruguay, between September 2024 and September 2025. THI was recomputed for both sources with the National Research Council (1971) equation, because the original database applied different rounding rules to the two sources, which alone created 0.53 THI units of bias. Inference relied on a two-stage bootstrap (cities, then devices within cities), since one city holds 82 sensors. Pooled mean bias was 3.05 THI units (95% CI 2.03 to 3.39) and the across-city mean bias was 1.87 (1.17 to 2.42). The discrepancy is not constant: it reaches 4.69 units when station THI is below 50 and falls to 0.84 between 68 and 72, vanishing above 72. Bias peaked during summer nights and was lowest in the afternoon, driven mainly by air temperature at night and by humidity in the afternoon. Consequently, 11.2% of hours deemed non-stressful by the station were already stressful in the barn, and summer nights averaged 0.73 h below THI 64 inside the barn against 3.54 h at the station. Bias ranged from −3.0 to +3.7 units across cities, ruling out a single correction factor; both negative values came from sites whose sensor recorded a wider diurnal cycle than the station, indicating a pairing failure.

Keywords: bioclimatology; thermal comfort; heat stress; free stall; embedded sensors; THI.

Introduction

The temperature-humidity index (THI) is the most widely used measure of heat load in dairy cattle, even though it ignores radiation and wind. The reason is practical: it requires only air temperature and relative humidity, which official weather networks have recorded for decades and at large scale. Much of what is known about heat-related production losses was built this way.

Ravagnolo et al. (2000) linked 119,337 test-day records from 134 herds in Georgia (USA) to the THI of the nearest station and estimated a loss of about 0.2 kg of milk per unit above 72; Bernabucci et al. (2014) matched records from 484 Italian farms with 35 stations. Neither study measured the air inside the barn, so the substitution was never tested.

The methodological debate has centred on the choice of formula: Bohmanova et al. (2007) favoured, under humid climates, indices that give more weight to humidity, whereas Dikmen and Hansen (2009), using rectal temperature as the response, obtained nearly identical fits with most formulations. Where the data feeding the index come from has received far less attention.

Using station data in place of barn data would be harmless if the two series differed by a constant. There is no physical reason to expect that. The meteorological screen stands over a grassed surface that loses heat by radiation through the night; in a free-stall barn the roof limits that loss, the structure releases heat stored during the day, and the animals release heat and water vapour continuously. Ventilation and evaporative cooling act in the opposite direction, lowering air temperature during the hot hours at the cost of higher humidity. Because THI combines the two variables non-linearly, the net effect changes with the hour of the day and with the season.

The few studies that have tested the assumption agree on the direction of the error and disagree on its size and on the correction proposed. Schüller et al. (2013), across seven farms, counted 64 days with mean THI at or above 72 inside the barn against 4 at the station. Shock et al. (2016), across 48 farms in Ontario, found a daily difference close to one unit, yet so variable between herds that they advised against the use of station data. Ouellet et al. (2019) recorded indoor THI 3.7 to 4.6 units above the value computed from station means. VanderZaag et al. (2023), in turn, accepted stations up to 50 km away for population-level studies, but found a larger excess at

night and a slope below 1 in naturally ventilated barns. An error that changes with the hour and with the level of the index is not removed by a constant.

Management thresholds inherit the problem: 72 and the 65 to 76 range came from station data (Ravagnolo et al. 2000; Bernabucci et al. 2014), whereas 68 and 65 came from responses of high-producing cows (Zimbelman et al. 2009; Pinto et al. 2020). A threshold calibrated at the station, applied to an indoor sensor, triggers the alarm too early; the reverse path delays it.

All of this evidence comes from temperate Northern-Hemisphere climates and from 6 to 48 farms per study. An equivalent assessment is missing for subtropical and tropical South America, where nights with high THI may shorten the relief period that Igono et al. (1992) associated with smaller production losses. Using hourly records from 172 sensors in free-stall barns across 30 municipalities in Brazil and Uruguay, this study quantifies that agreement, describes how it varies with the hour, the season and the level of heat load, measures the misclassification of stressful hours and the shortening of the night-time recovery window, and separates the structural component of the bias from hour-to-hour fluctuation.

Materials and Methods

Monitoring network and data provenance

We analysed hourly records from 172 microenvironment monitoring devices (CowComfort system) installed in free-stall dairy barns across 30 municipalities: 28 in seven federative units of Brazil (Paraná, Rio Grande do Sul, Santa Catarina, São Paulo, Minas Gerais, Goiás and Rondônia) and two in Uruguay (departments of San José and Colonia). The period covered runs from 3 September 2024 to 16 September 2025. According to the manufacturer's technical material, each device is battery powered, transmits through C-Com antennas and takes temperature and humidity readings at 10-minute intervals in the default mode. The hourly record in the database is the mean

of the six readings of that hour. The declared accuracy is ±0.3 °C for temperature and ±2 percentage points for relative humidity. The sensor sits inside the barn of the participating farm, at a height and position that vary between installations, since each barn has its own design.

For each device and hour, air temperature (°C) and relative humidity (%) were recorded in the animals' microenvironment. The same variables for the outdoor condition were supplied by the manufacturer from the weather station nearest to each participating farm, which allowed hourly pairing between the two sources. In both sources the hourly value is the mean of the hour preceding the record, so the pairing compares means over the same time window. Four limitations follow from this provenance and run through the whole study: the choice of reference station was made by the manufacturer, and neither the distance nor the altitude difference between station and farm was reported; the mounting height and position of the sensor inside the barn were neither standardised nor recorded, and the sensor model was not reported by the manufacturer; the cooling systems of each barn (forced ventilation, sprinkling, misting) are absent from the database; and the variables arrive rounded to the nearest integer (1 °C and 1%). In 28 of the 30 municipalities all devices were paired with the same station series; in Carambeí and Castro there are up to six and three distinct series, respectively.

Temperature-humidity index and heat-stress classes

THI was computed for both sources from hourly air temperature (T, °C) and relative humidity (RH, %) using the National Research Council (1971) equation:

$$\mathrm{THI} = (1.8 \times T + 32) - (0.55 - 0.0055 \times RH) \times (1.8 \times T - 26)$$

The choice meets three criteria. It is the same equation, expressed in °F, adopted by Collier et al. (2012) in the class scale used below, which keeps index and thresholds consistent; it is the one used in studies that compared station and barn THI (Ouellet et al. 2019), which makes

magnitudes comparable; and, in subtropical environments, the formulation has little effect on the ability of the index to describe the animal response (Dikmen and Hansen 2009).

Propagated through the equation by a first-order approximation, the declared sensor accuracy corresponds to a combined uncertainty of 0.33 to 0.74 THI unit over the range of 5 to 40 °C and 30 to 100% relative humidity, with a median of 0.50. The sensitivity of the index to temperature, between 1.3 and 1.7 unit per °C, is about ten times its sensitivity to one percentage point of humidity, which anticipates the asymmetry found in the decomposition of the bias.

Recomputing the index was not a formality. The delivered database already carried THI columns, and the audit showed that they follow the NRC equation but with different rounding rules between sources: 97.6% of the microenvironment values correspond to rounding the index, whereas 97.9% of the station values correspond to truncation. That asymmetry inflates the mean difference between sources by 0.53 THI unit. All results below use the recomputed index, with the same rule for both sources.

Hours were classified with the scale of Collier et al. (2012), which starts from the threshold of 68 proposed by Zimbelman et al. (2009) for high-producing cows: no stress (THI < 68), mild ($68 \leq \text{THI} < 72$), moderate ($72 \leq \text{THI} < 80$), severe ($80 \leq \text{THI} < 90$) and extreme ($\text{THI} \geq 90$). As a sensitivity analysis, categorical agreement was recomputed with the single threshold of 72 estimated by Ravagnolo et al. (2000) from station data. For the night-time recovery window we adopted THI = 64, the critical minimum indicated by Igono et al. (1992).

Data processing

The raw database contained 375,635 hourly records. We removed 2,988 records without a matching station measurement and 14 records with physically implausible values (relative humidity equal to zero, or temperature outside the −10 °C to 45 °C range), leaving 372,633 valid

hourly pairs from 172 devices and 30 municipalities. Derived variables included the hour of the day, the Southern-Hemisphere season (summer: December–February; autumn: March–May; winter: June–August; spring: September–November) and the period of the day (late night: 0–5 h; morning: 6–11 h; afternoon: 12–17 h; evening: 18–23 h). Bias was defined as the hour-by-hour difference between microenvironment THI and station THI.

Each municipality was classified under Köppen using the municipal table of Alvares et al. (2013), which accompanies their 1-ha map and reports, for every Brazilian municipality, the class covering the largest area and the mean altitude of the 100-m pixels within its perimeter. The two Uruguayan municipalities fall outside that mapping, and for them the classification does not depend on municipal resolution. The Instituto Uruguayo de Meteorología states that differences between points of the national territory are not large enough to distinguish climate types, and assigns to the whole country the combination temperate (C), humid with no dry season (f) and warmest month above 22 °C (a) (Inumet n.d.). Tarariras and San José de Mayo therefore enter as Cfa, with the caveat that this attribution is national and does not separate internal variation, an asymmetry that does not arise on the Brazilian side, where the table operates municipality by municipality. Grouping by federative unit, used in earlier versions of this analysis, does not support the description: in Minas Gerais, for instance, no class dominates the territory, with Aw over 34.9%, Cwb over 26.0%, Cwa over 25.5% and As over 12.3% of the area (Alvares et al. 2013), so assigning the state class to a highland municipality is a gamble. Because city and device enter as random effects, the climatic grouping organises the description without affecting the estimates.

Statistical analysis

The network is unbalanced: one municipality (Carambeí, PR) holds 82 of the 172 devices, installed on 82 distinct farms, and 56% of the valid hours, and treating the 372 thousand hours as

147 independent observations would produce unrealistic confidence intervals. Two measures
148 addressed this. First, every statistic is reported in two versions: pooled across hours, which
149 describes the database as it is, and as a mean of means, first by device and then by municipality,
150 so that each municipality carries equal weight. Second, all uncertainty comes from a two-stage
151 bootstrap with 2,000 replicates: cities are drawn with replacement and, within each drawn city, its
152 devices are drawn with replacement (Field and Welsh 2007). Because the whole series of a sensor
153 enters or leaves together, the procedure accommodates repeated measures and serial dependence
154 at once, without requiring an explicit model for the correlation structure. Device-level statistics
155 used only sensors with at least 168 hours (seven days).

156 Before the agreement analyses, each sensor passed a physical plausibility check: the mean
157 amplitude of the daily temperature cycle measured inside the barn was divided by the amplitude
158 recorded at the reference station. A roofed barn damps the outdoor cycle, so this ratio should fall
159 below 1; values above it indicate either a sensor that does not represent the interior of the facility
160 or a reference station that is climatically distant. The ratio was computed for sensors with at least
161 168 hours and used in a sensitivity analysis, not as an exclusion criterion in the main analysis.

162 Continuous agreement was assessed by Pearson correlation, by Lin's concordance
163 correlation coefficient (CCC; Lin 1989), by Bland-Altman analysis with mean bias and 95% limits
164 of agreement, and by Deming regression with variance ratio $\lambda = 1$, appropriate when both variables
165 are subject to measurement error. The dependence of the bias on the level of heat load was
166 described across station THI bands. Categorical agreement was quantified by Cohen's weighted
167 kappa (Cohen 1968), with linear and quadratic weights, by the cross-classification matrix and by
168 the proportion of reclassified hours.

The variance of the hourly bias was decomposed into three levels (between cities, between devices within city, and between hours within device) using moment estimators that discount the sampling error of each level; within-device variation was further split into a between-days and a within-day component. Temporal dependence was described by the autocorrelation function of the deviation of the bias from the device mean, computed only between consecutive hours.

Bias by season, period of the day and hour was estimated as a mean of means, with confidence intervals from the same two-stage bootstrap. The origin of the bias was separated into a temperature component and a humidity component by recomputing microenvironment THI with station humidity and vice versa, and averaging the two substitution orders. For summer nights (18 h to 5 h, December to February), we counted, per night and per device, the hours with THI below 64 and the hours with THI at or above 68 in each source. Two sensitivity analyses were run: exclusion of the dominant municipality and use of the 72 threshold in place of 68. Analyses were carried out in Python 3.11. Supplementary tables (Tables S1 to S3) and figure (Fig. S1) are provided in Online Resource 1, and the script that reproduces all results in Online Resource 2.

Results

Network coverage

The 372,633 valid hours are very unevenly distributed across climate types (Table 1). The two subtropical classes account for 20 of the 30 municipalities and 87% of the hours, with Cfb predominating: 8 municipalities at a mean altitude of 890 m, 106 sensors and 71% of the records. Class Cfa contributes 12 municipalities and 26 sensors at lower elevations (370 m on average). At the other extreme, Cwb totals 602 hours, Aw totals 5,449, and Am, represented by a single sensor, 2,495. Temporal coverage is uneven as well: June to September 2025 accounts for about 70% of

the records, whereas December to February holds 10%. This does not preclude the seasonal analysis, but it sets the precision of each cell, made explicit in the confidence intervals.

Table 1. Temperature-humidity index (THI) in the barn microenvironment and at the regional weather station, by Köppen climate type.

Climate type	Municip.	Sensors	Hours	Mean alt. (m)	Micro THI (mean ± SD)	Station THI	Pooled bias	Bias/city	Stress micro (% h)	Stress station (% h)
Cfa (humid subtropical, hot summer)	12	26	58,774	370	66.0 ± 9.6	63.5	2.45	2.07	44.5	34.9
Cfb (humid subtropical, temperate summer)	8	106	264,788	890	61.3 ± 7.3	57.9	3.42	2.88	17.8	10.2
Cwa (dry-winter tropical, hot summer)	4	20	40,525	732	65.0 ± 5.7	63.0	2.05	1.14	32.8	22.1
Cwb (dry-winter tropical, temperate summer)	2	16	602	984	64.9 ± 5.2	61.8	3.09	n.a.	27.4	24.4
Aw (tropical savanna)	3	3	5,449	909	66.3 ± 8.0	67.0	−0.75	−0.94	47.9	44.4
Am (tropical monsoon)	1	1	2,495	166	75.7 ± 5.2	73.4	2.30	2.30	91.3	83.5
Total	30	172	372,633	n.a.	62.6 ± 7.9	59.6	3.05	1.87	24.6	16.4

Municipal Köppen class and mean municipal altitude follow Alvares et al. (2013); the two Uruguayan municipalities enter as Cfa, the class that Inumet (n.d.) assigns to the whole national territory, and have no altitude. Bias = microenvironment THI − station THI. Pooled bias: mean over all hours. Bias per city: mean of means by device and then by municipality, computed only with sensors of at least 168 hours (n.a. when no sensor in the class reaches that minimum). Stress: hours with THI ≥ 68. SD = standard deviation. Municipality-level detail in Table S2 (Online Resource 1).

Agreement between the two sources

Microenvironment and station THI showed strong linear association (r = 0.932; 95% CI 0.918 to 0.952), with substantial but incomplete agreement (Lin's CCC = 0.870; 0.839 to 0.916). Pooled mean bias was 3.05 THI units (2.03 to 3.39), and the 95% limits of agreement ranged from −3.04 to 9.14 (Fig. 1b), a span of 12 units that describes the error of a single isolated hour and not that of the mean. When each municipality carries equal weight, the bias falls to 1.87 (1.17 to 2.42; 139 devices, 23 municipalities). The difference between the two versions is not noise: it measures the weight of the dominant municipality, which has the largest bias in the network (Fig. 3c).

Deming regression gave a slope of 0.912 (0.881 to 1.005) and an intercept of 8.29 (1.75 to 10.22), that is, an offset that shrinks as THI rises, although the confidence interval of the slope reaches unity. The band-by-band description is more direct (Table S3, Online Resource 1): mean bias is 4.69 units when the station records THI below 50, 1.27 between 65 and 68, 0.84 between

68 and 72 and −0.07 between 72 and 76, becoming null above that. In operational terms, the station underestimates heat load exactly in the band where the decision to start mitigation is taken, and converges to the microenvironment when the heat intensifies.

Table 2. Agreement between microenvironment THI and regional weather-station THI, with confidence intervals from a two-stage bootstrap (2,000 replicates).

Statistic	Estimate	95% CI
Mean bias (THI)	3.05	2.03 to 3.39
SD of the differences	3.11	2.64 to 3.36
Lower 95% LoA	−3.04	−4.33 to −2.52
Upper 95% LoA	9.14	7.64 to 9.69
Pearson's r	0.932	0.918 to 0.952
Lin's CCC	0.870	0.839 to 0.916
Deming slope	0.912	0.881 to 1.005
Deming intercept	8.29	1.75 to 10.22
Linear weighted kappa	0.675	0.632 to 0.712
Quadratic weighted kappa	0.784	0.737 to 0.822
Hours classified as no stress at the station already stressful in the microenvironment (%)	11.15	8.52 to 18.88
Hours with THI ≥ 68 in the microenvironment (%)	24.59	17.56 to 42.77
Hours with THI ≥ 68 at the station (%)	16.42	10.94 to 32.89

Statistics computed over the 372,633 paired hours. LoA = limits of agreement; CCC = concordance correlation coefficient; Deming fitted with λ = 1.

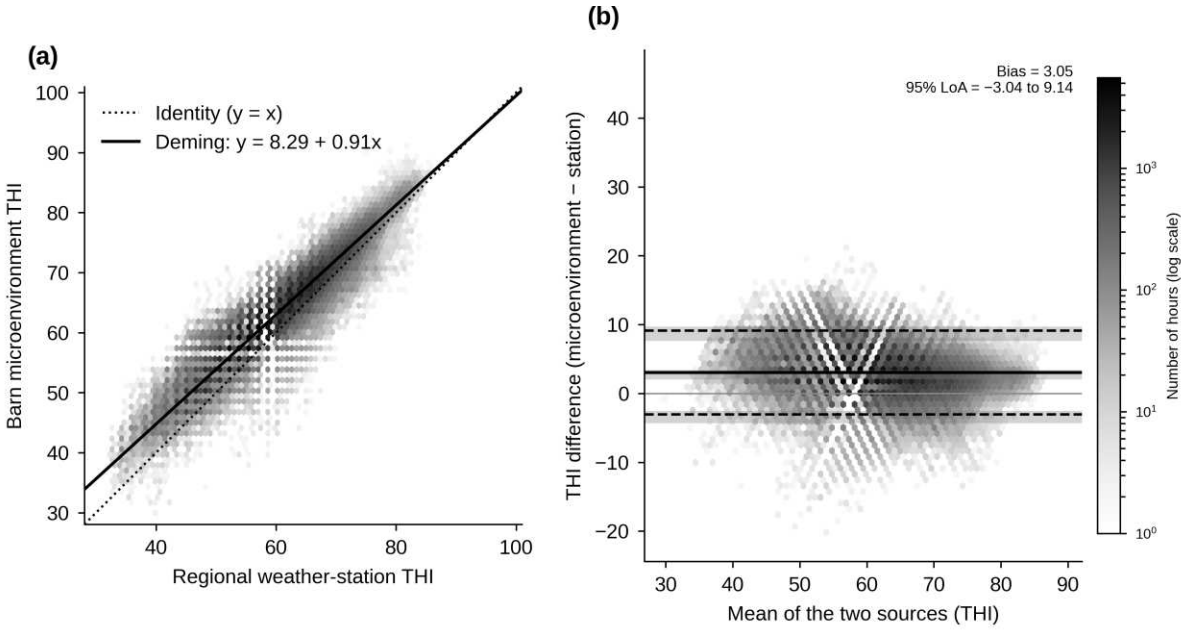


Fig. 1 Agreement between the temperature-humidity index (THI) of the barn microenvironment and that of the regional weather station (372,633 paired hours; 172 devices; 30 municipalities). (a) Scatter plot, with

the identity line (dotted) and the Deming regression (λ = 1). (b) Bland-Altman plot: solid line, mean bias; dashed lines, 95% limits of agreement; grey bands, 95% confidence intervals from the two-stage bootstrap. Hour density is on a logarithmic scale; the diagonal banding pattern arises from temperature and humidity being recorded as integers.

Misclassification of hours

Agreement in categorical classification was moderate to good (linear weighted kappa = 0.675, 0.632 to 0.712; quadratic = 0.784, 0.737 to 0.822). The cross-classification matrix (Fig. 2a) shows where disagreement concentrates: 11.2% of the hours classified as non-stressful by the station were already stressful in the microenvironment (8.5 to 18.9%), and 43% of the hours classified as mild by the station corresponded to moderate stress inside the barn. The reverse path is rare: 1.2% of all hours were classified as stressful by the station and non-stressful in the microenvironment, against 9.3% in the opposite direction. Across the year, the proportion of hours with THI ≥ 68 was higher in the microenvironment in every month (Fig. 2b), with the widest gap in the warm months. With the 72 threshold in place of 68, kappa was 0.669 and the proportion of reclassified hours fell to 5.2%, consistent with the shrinking of the bias in the upper band.

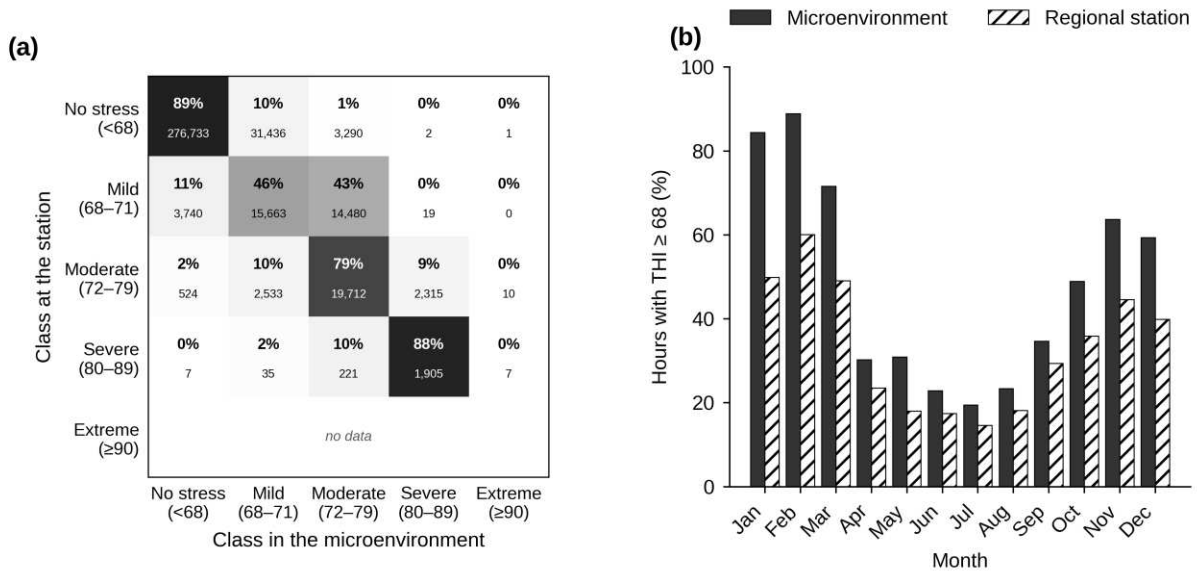


Fig. 2 Classification of hours into heat-stress classes by microenvironment THI and by regional station THI. (a) Cross-classification matrix: row percentage (class assigned by the station) and, below it, the

number of hours. (b) Monthly proportion of hours with THI ≥ 68, computed as the mean of municipal proportions. Classes follow Collier et al. (2012), with the stress threshold at 68 (Zimbelman et al. 2009).

Temporal structure of the bias

Both sources reach a minimum at 6 h and a maximum at 15 h, but with different daily amplitudes: 11.6 THI units in the microenvironment against 13.0 at the station (Fig. 3a). The barn damps the outdoor oscillation and shifts it upward, and the result is a bias with a well-defined daily shape (Fig. 3b): high and stable from the late night to early morning, lowest in mid-afternoon and rising again through the evening. Table 3 organises the pattern by season. The largest bias occurred in the summer late night, 4.43 (95% CI 3.96 to 5.00) THI units, and the smallest in the spring afternoon, 1.35 (95% CI 0.85 to 1.86). In all four seasons the afternoon had the smallest bias of the day. The month-by-hour map (Fig. S1, Online Resource 1) shows the same pattern in more detail: the largest bias occurs at 6 h in January (6.9 units) and the only negative value in the early afternoon of the same month (−1.9).

Table 3. THI bias (microenvironment − station) by season and period of the day, estimated as a mean of means, with 95% confidence interval.

Season	Late night	Morning	Afternoon	Evening	Municipalities / devices
Summer	4.43 (3.96 to 5.00)	2.85 (1.25 to 3.76)	1.78 (1.21 to 2.38)	3.62 (2.77 to 4.55)	8 / 23
Autumn	2.86 (2.10 to 3.53)	2.63 (2.28 to 2.97)	1.76 (1.17 to 2.43)	2.53 (1.75 to 3.18)	10 / 56
Winter	1.52 (−0.18 to 3.04)	2.07 (1.47 to 2.59)	1.49 (0.99 to 1.85)	1.80 (0.65 to 2.93)	18 / 117
Spring	1.95 (0.86 to 2.90)	2.42 (1.93 to 2.91)	1.35 (0.85 to 1.86)	1.98 (1.14 to 2.81)	26 / 143

Late night: 0–5 h; morning: 6–11 h; afternoon: 12–17 h; evening: 18–23 h. Intervals from the two-stage bootstrap. The number of municipalities and devices varies between seasons because sensors entered and left the network during the period.

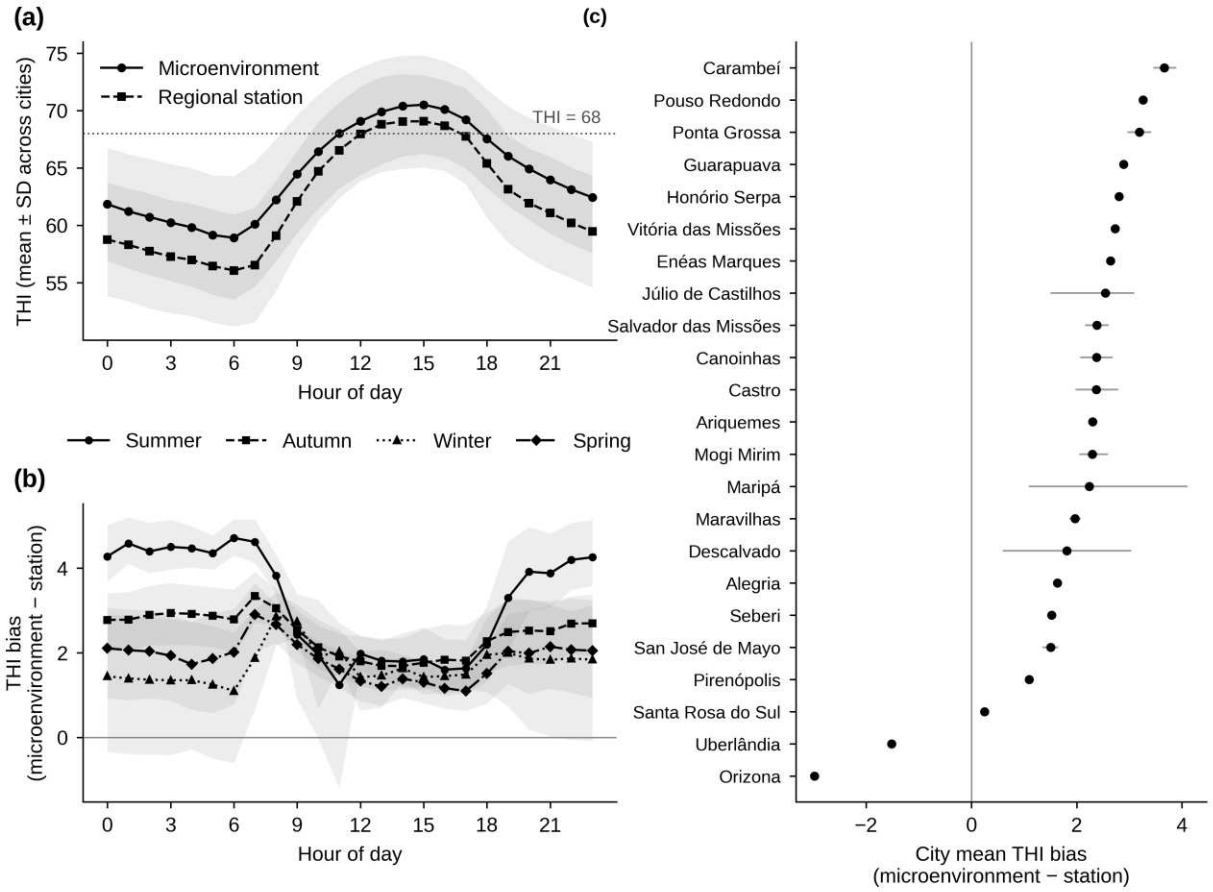


Fig. 3 Temporal and spatial structure of the difference between microenvironment THI and regional station THI. (a) Daily profile of mean THI in both sources (mean of municipal means ± standard deviation across municipalities); dotted line, THI = 68. (b) Bias by hour of the day in each season, with 95% confidence bands from the two-stage bootstrap. (c) Mean bias of each municipality (mean of device means); horizontal bars, 95% confidence interval across devices, absent in municipalities with a single sensor.

Summer nights and the recovery window

Across the 1,469 summer nights with sufficient coverage (24 devices, 9 municipalities), the station indicated 3.54 hours per night with THI below 64 (1.86 to 4.93), against 0.73 hour measured in the microenvironment (0.22 to 1.22). In the opposite direction, the microenvironment accumulated 8.79 hours per night with THI ≥ 68 (7.02 to 10.59), against 4.01 at the station (1.87 to 6.66). The night-time relief window that the station suggests therefore exists, to a large extent, only outside the barn.

Between-site heterogeneity and the origin of the bias

Municipal mean bias ranged from −3.0 units in Orizona (GO) to +3.7 in Carambeí (PR), with 21 of the 23 municipalities above zero (Fig. 3c). Across municipalities, the bias was negatively correlated with mean station THI ($r = -0.32$): the warmer sites showed the smallest differences, and the two negative values occurred in municipalities of Goiás and Minas Gerais with mean THI above 65. Decomposition of the hourly bias variance (139 devices, 23 municipalities) attributed 16.5% to municipality and 23.8% to municipality plus device combined (standard deviations of 1.33 and 0.88 THI units), leaving 2.85 units of hour-to-hour variation within the same sensor (Table S1, Online Resource 1). About a quarter of the discrepancy is therefore structural, and the between-municipality component exceeds the component between barns of the same municipality.

Separating temperature from humidity (Table 4) reveals two regimes. In the evening and late night, the microenvironment was 2.1 to 2.3 °C warmer than the station with almost identical relative humidity, and virtually the whole bias came from temperature. In the afternoon the temperature difference fell to 0.5 °C, while relative humidity inside the barn stood 9.3 percentage points above the outdoor value; the humidity component then exceeded the temperature component, yet total bias was the smallest of the day.

Table 4. Decomposition of THI bias into temperature and relative-humidity components, by period of the day.

Period	Bias (THI)	Temperature component	Humidity component	Temperature difference (°C)	Humidity difference (p.p.)
Late night	2.92	2.84	0.08	1.73	0.10
Morning	2.49	2.30	0.19	1.51	2.96
Afternoon	1.31	0.49	0.82	0.32	7.47
Evening	2.90	2.67	0.24	1.67	0.98

Means of municipal means. Components obtained by recomputing microenvironment THI with station humidity and vice versa, and averaging the two substitution orders. p.p. = percentage points.

Temporal dependence and sensitivity analyses

Autocorrelation of the bias within the same device was high and decayed slowly (0.85, 0.70, 0.57, 0.44, 0.33, 0.23 at lags of 1 to 6 hours), confirming that consecutive hours are not independent observations and justifying resampling whole devices. The plausibility check showed a median amplitude ratio of 0.80 among eligible sensors, consistent with the damping expected from a roof. Two sensors exceeded 1.15: Uberlândia (1.42) and Orizona (1.30), precisely the two municipalities with negative bias (Table S2, Online Resource 1). Excluding them, pooled bias becomes 3.10 and the across-municipality bias rises from 1.87 to 2.26 (21 municipalities). Excluding the dominant municipality (164,435 hours remaining), pooled bias fell to 2.42 units, the Deming slope rose to 0.979, the linear weighted kappa was 0.682 and the proportion of reclassified hours rose to 16.2%. The qualitative conclusions do not change, but the magnitude of the bias depends on which barns dominate the sample.

Discussion

The central result is not the existence of the bias, already reported under temperate climates (Schüller et al. 2013; Shock et al. 2016; Ouellet et al. 2019; VanderZaag et al. 2023), but its shape. The difference between microenvironment and station is not a constant offset that can be removed by addition: it is largest when heat load is low, in the late night, in the evening and in winter, and vanishes above 72 THI units. This reconciles two findings that appear to contradict each other in the literature. The difference of about one unit in daily THI reported by Shock et al. (2016) and the 3.7 to 4.6 units of Ouellet et al. (2019) may describe the same phenomenon measured at different points of the scale and with different mixes of hours; the slope below 1 observed by VanderZaag et al. (2023) in naturally ventilated barns is the same level dependence that appears here, now quantified band by band.

316 The physical explanation is consistent with the decomposition in Table 4. At night the
317 meteorological screen, over an open surface, loses heat by long-wave radiation, while the barn
318 retains heat stored in its structure and receives sensible heat from the animals: the result is 2 °C
319 more indoors with similar relative humidity. In the afternoon, ventilation and evaporative cooling
320 bring indoor temperature close to outdoor temperature and raise humidity by almost 10 percentage
321 points; because the weight of humidity in THI is modest in that band, the bias shrinks. Convergence
322 under heat therefore does not mean that the barn is a benign environment at midday. It means that,
323 during the hot hours, what the animal faces in excess of the open field is radiation and air
324 movement, which THI does not measure (Dikmen and Hansen 2009).

325 The practical consequence concentrates on the threshold. By the Deming line, a station THI
326 of 68 corresponds to 70.3 in the microenvironment, and 72 corresponds to 74.0; in the reverse
327 direction, a threshold of 68 measured inside the barn is equivalent to 65.5 at the station. Two to
328 three THI points may seem little, but this is exactly the order of magnitude that separates the
329 threshold proposals circulating in the literature: 72 estimated from station data by Ravagnolo et al.
330 (2000), 68 proposed by Zimbelman et al. (2009) for high-producing cows, and values close to 65
331 derived from physiological responses by Pinto et al. (2020). Part of the dispersion among these
332 numbers may be metrological rather than biological. Studies that combine thresholds from one
333 source with measurements from another should state the equivalence assumed; the same applies
334 to the rounding rule, which in this dataset alone shifted the difference between sources by half a
335 unit.

336 The night-time window gives that offset a biological meaning. Igono et al. (1992) showed
337 that cool hours at night reduce production losses under heat. On the summer nights of this network
338 the station suggested between three and four hours per night below THI 64, while the

microenvironment delivered less than one; at the same time, the barn accumulated more than twice as many hours above 68. A decision system fed by station data would record recovery nights that, inside the barn, never occurred. Since night-time recovery is precisely what night ventilation and sprinkling aim to provide, the error falls on the most expensive decision.

Between-site heterogeneity is the result that most restricts general solutions. A quarter of the hourly bias variance is structural, and most of that fraction lies between municipalities rather than between barns of the same municipality, which points to the reference station, and not only to the building, as the source of the deviation. By climate type, the largest bias occurred in the Cfb municipalities, at a mean altitude close to 890 m, and the smallest in the Aw ones, but the comparison between classes is confounded with the imbalance of the network, because class Cfb is essentially the dominant municipality. Across municipalities, moreover, altitude did not explain the bias ($r = 0.01$), which rules out the simple reading that barns at higher elevations would accumulate more relative heat. Two municipalities showed negative bias, both in warm regions, and the bias decreased with the mean THI of the site. Those two cases, however, fail a simple test: the daily temperature cycle measured inside the barn was 30% (Orizona) and 42% (Uberlândia) wider than that of the station, whereas the network median is 0.80. A roof does not amplify the outdoor oscillation, and hourly aggregation would not explain the effect either: a one-hour moving average attenuates a 24-h cycle by about 0.3%, three orders of magnitude below the observed excess. Either the sensor does not represent the interior of the facility, or the reference station is far enough away to describe a different microclimate. In both cases, the negative value says more about the pairing than about the barn. Without the distance and the altitude between farm and station, the two hypotheses cannot be separated, and this is the main gap of the design. It also explains why correction equations published for one region (Ouellet et al. 2019; VanderZaag et al.

2023) should not be transported: here, the required correction factor would change sign between Paraná and Goiás.

Four limitations bound the reach of these numbers. Mounting height and position of the sensor were not standardised across farms, which is a plausible candidate for part of the variance between devices of the same municipality, without the database allowing the hypothesis to be tested. The imbalance concentrates 56% of the hours in a single municipality, which is why the bias is also reported as a mean across municipalities (1.87) and the analyses were repeated without the dominant municipality, with stable qualitative conclusions. Summer is under-represented, and the summer late-night estimates rest on 24 devices from 9 municipalities. Recording temperature and humidity as integers adds discretisation noise, visible in the diagonal banding of Fig. 1, which widens the limits of agreement without displacing the mean. To that noise is added the declared sensor accuracy, equivalent to about 0.5 THI unit, which widens the same limits and is six times smaller than the pooled mean bias. None of these limitations affects the daily pattern, which appears in every season, in practically every municipality and in both versions of the statistics.

What remains is what this study cannot say. There are no production, respiration-rate or body-temperature data, so the productive cost of the misclassification remains an inference based on thresholds from the literature. We also did not measure radiation or air velocity, which are precisely what distinguishes the microenvironment from the meteorological screen during the hot hours; a black-globe humidity index (BGHI) or air-velocity measurements would change the afternoon comparison, probably widening the difference that THI does not capture. Combining in-barn sensors with production variables from the same herd is the missing step to convert these THI units into kilograms of milk.

Conclusions

Across 172 free-stall barns in Brazil and Uruguay, THI computed from the nearest weather station underestimated microenvironment THI by 3.05 units on average over hours and by 1.87 on average across municipalities, but this difference is not uniform: it is largest in summer late nights, smallest in the afternoon and practically null above THI 72.

The error falls on the decision band: 11.2% of the hours considered non-stressful by the station were already stressful inside the barn, and summer nights lost about three of the four hours of thermal recovery that the station indicated.

Because a quarter of the discrepancy is structural and ranges from −3.0 to +3.7 units across municipalities, no single correction turns station data into microenvironment data. For night-time management decisions, measurement inside the barn is not replaceable; for large-scale studies that depend on stations, the equivalence between thresholds needs to be stated and, where possible, calibrated per site.

Acknowledgements

The authors thank Cowmed for providing the monitoring records analysed here.

Declarations

Competing interests

The data analysed were provided by Cowmed, manufacturer of the CowComfort system, which also selected the weather stations used as reference. None of the authors holds an employment, ownership or consultancy relationship with the company. The company played no part in the design of the analyses, in the interpretation of the results, in the writing of the manuscript or in the decision to submit it.

Funding

The authors declare that no funds, grants or other support were received during the preparation of this manuscript.

Ethics approval

This study analysed environmental records of air temperature and relative humidity collected by fixed sensors inside commercial barns and by regional weather stations. No animal was handled, restrained or subjected to any procedure, and no animal-derived measurement was taken, so approval by an animal ethics committee was not required. [Confirmar com a instituição.]

Consent to participate

Not applicable. The study involved no human participants.

Consent to publish

Not applicable.

Data availability

The sensor and weather-station records belong to Cowmed and were transferred to the authors under a data transfer agreement. They may be made available upon reasonable request and with the permission of the company. The analysis script is provided in Online Resource 2.

Author contributions

[Confirmar a atribuição antes de submeter.] Conceptualisation: Thiago Vasconcelos Melo and Diego Azevedo Mota. Data curation: Thiago Vasconcelos Melo. Formal analysis: Diego Azevedo Mota. Methodology: Diego Azevedo Mota and Thiago Vasconcelos Melo. Writing, original draft: Diego Azevedo Mota. Writing, review and editing: Thiago Vasconcelos Melo and Felipe Nogueira Domingues. All authors read and approved the final manuscript.

428

References

Alvares CA, Stape JL, Sentelhas PC, Gonçalves JLM, Sparovek G (2013) Köppen's climate classification map for Brazil. Meteorol Z 22:711–728. https://doi.org/10.1127/0941-2948/2013/0507

Bernabucci U, Biffani S, Buggiotti L, Vitali A, Lacetera N, Nardone A (2014) The effects of heat stress in Italian Holstein dairy cattle. J Dairy Sci 97:471–486. https://doi.org/10.3168/jds.2013-6611

Bohmanova J, Misztal I, Cole JB (2007) Temperature-humidity indices as indicators of milk production losses due to heat stress. J Dairy Sci 90:1947–1956. https://doi.org/10.3168/jds.2006-513

Cohen J (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 70:213–220. https://doi.org/10.1037/h0026256

Collier RJ, Hall LW, Rungruang S, Zimbelman RB (2012) Quantifying heat stress and its impact on metabolism and performance. Department of Animal Sciences, University of Arizona, Tucson

Dikmen S, Hansen PJ (2009) Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? J Dairy Sci 92:109–116. https://doi.org/10.3168/jds.2008-1370

Field CA, Welsh AH (2007) Bootstrapping clustered data. J R Stat Soc Ser B Stat Methodol 69:369–390. https://doi.org/10.1111/j.1467-9868.2007.00593.x

Igono MO, Bjotvedt G, Sanford-Crane HT (1992) Environmental profile and critical temperature effects on milk production of Holstein cows in desert climate. Int J Biometeorol 36:77–87. https://doi.org/10.1007/BF01208917

Inumet (Instituto Uruguayo de Meteorología) (n.d.) Clasificación climática. https://www.inumet.gub.uy/clima/estadisticas-climatologicas/clasificacion-climatica. Accessed 17 Sep 2026

Lin LI (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics 45:255–268. https://doi.org/10.2307/2532051

National Research Council (1971) A guide to environmental research on animals. The National Academies Press, Washington, DC. https://doi.org/10.17226/20608

Ouellet V, Bellavance AL, Fournel S, Charbonneau É (2019) Short communication: summer on-farm environmental condition assessments in Québec tiestall farms and adaptation of temperature-humidity index calculated with local meteorological data. J Dairy Sci 102:7503–7508. https://doi.org/10.3168/jds.2018-16159

462 Pinto S, Hoffmann G, Ammon C, Amon T (2020) Critical THI thresholds based on the
463 physiological parameters of lactating dairy cows. J Therm Biol 88:102523.
464 https://doi.org/10.1016/j.jtherbio.2020.102523

465 Ravagnolo O, Misztal I, Hoogenboom G (2000) Genetic component of heat stress in dairy cattle,
466 development of heat index function. J Dairy Sci 83:2120–2125.
467 https://doi.org/10.3168/jds.S0022-0302(00)75094-6

468 Schüller LK, Burfeind O, Heuwieser W (2013) Short communication: comparison of ambient
469 temperature, relative humidity, and temperature-humidity index between on-farm measurements
470 and official meteorological data. J Dairy Sci 96:7731–7738. https://doi.org/10.3168/jds.2013-
471 6736

472 Shock DA, LeBlanc SJ, Leslie KE, Hand K, Godkin MA, Coe JB, Kelton DF (2016) Studying
473 the relationship between on-farm environmental conditions and local meteorological station data
474 during the summer. J Dairy Sci 99:2169–2179. https://doi.org/10.3168/jds.2015-9795

475 VanderZaag A, Le Riche E, Baldé H, Kallil S, Ouellet V, Charbonneau É, Coates T, Wright T,
476 Luimes P, Gordon R (2023) Comparing thermal conditions inside and outside lactating dairy
477 cattle barns in Canada. J Dairy Sci 106:4738–4758. https://doi.org/10.3168/jds.2022-22870

478 Zimbelman RB, Rhoads RP, Rhoads ML, Duff GC, Baumgard LH, Collier RJ (2009) A re-
479 evaluation of the impact of temperature humidity index (THI) and black globe humidity index
480 (BGHI) on milk production in high producing dairy cows. In: Proceedings of the Southwest
481 Nutrition and Management Conference, Tucson, 26–27 February 2009. University of Arizona,
482 Tucson, pp 158–168

483

Supplementary material

The items below are also provided as Online Resource 1.

Table S1 Variance components of the hourly THI bias (microenvironment − station), estimated by moments.

Component	Variance	Standard deviation (THI)	Share (%)
Between cities	1.768	1.33	16.5
Between devices (within city)	0.775	0.88	7.2
Between hours (within device)	8.147	2.85	76.2
between days, within device	2.313	1.52	n.a.
between hours of the same day	5.979	2.45	n.a.

Estimates from 139 devices in 23 municipalities, each with at least 168 hours of record. The last two rows decompose the within-device variation and therefore do not add to the rows above; n.a. = not applicable.

Table S2 Characterisation by municipality.

Municipality	State / Dept.	Köppen	Altitude (m)	Sensors	Hours	Micro THI	Bias	Stress micro (% h)	Stress station (% h)	Indoor/outdoor ampl.
Tarariras	Colonia (UY)	Cfa	n.a.	6	966	62.9	6.44	24.6	0.0	n.a.
Enéas Marques	PR	Cfa	644	1	3,689	62.2	2.64	23.1	14.6	0.92
Maripá	PR	Cfa	398	3	10,745	68.2	1.55	54.7	49.6	1.02
Alegria	RS	Cfa	334	1	3,419	61.1	1.63	18.9	15.1	0.88
Júlio de Castilhos	RS	Cfa	407	4	21,204	65.4	2.98	43.0	30.8	0.75
Nova Boa Vista	RS	Cfa	463	1	3	66.5	7.65	0.0	0.0	n.a.
Salvador das Missões	RS	Cfa	199	2	16,078	68.5	2.38	57.3	45.8	0.78
Seberi	RS	Cfa	456	1	276	61.5	1.52	17.8	16.3	1.00
Vitória das Missões	RS	Cfa	241	1	270	62.5	2.73	16.7	12.6	0.90
Pouso Redondo	SC	Cfa	525	1	367	61.8	3.26	8.4	2.5	0.89
Santa Rosa do Sul	SC	Cfa	33	1	464	63.3	0.25	19.0	24.6	1.02
San José de Mayo	San José (UY)	Cfa	n.a.	4	1,293	53.6	1.51	1.6	0.0	0.89
Arapoti	PR	Cfb	771	1	57	74.4	7.75	100.0	24.6	n.a.
Carambeí	PR	Cfb	974	82	208,198	60.9	3.54	14.9	8.6	0.74
Castro	PR	Cfb	950	11	16,709	59.2	2.62	11.7	7.3	0.84
Guarapuava	PR	Cfb	1072	1	412	62.6	2.89	18.0	15.5	0.78
Honório Serpa	PR	Cfb	852	1	7,885	64.2	2.80	35.5	22.1	0.87
Ponta Grossa	PR	Cfb	879	5	30,752	64.7	3.20	36.1	19.6	0.81
Vitorino	PR	Cfb	789	3	141	65.7	2.80	27.7	34.0	n.a.
Canoinhas	SC	Cfb	831	2	634	59.5	2.39	5.8	6.5	0.67
Maravilhas	MG	Cwa	761	2	3,103	66.4	1.97	39.7	28.6	0.88
Uberlândia	MG	Cwa	810	1	2,288	64.2	−1.51	37.3	35.6	1.42
Descalvado	SP	Cwa	700	2	907	66.5	1.81	40.9	34.8	0.68
Mogi Mirim	SP	Cwa	659	15	34,227	64.9	2.31	31.7	20.2	0.86
Pouso Alto	MG	Cwb	1044	14	310	65.8	3.16	31.9	33.5	n.a.
Santa Rita do Sapucaí	MG	Cwb	924	2	292	64.0	3.01	22.6	14.7	n.a.
Corumbá de Goiás	GO	Aw	1039	1	127	70.0	1.78	56.7	49.6	n.a.
Orizona	GO	Aw	832	1	2,488	63.1	−2.98	38.0	39.2	1.30
Pirenópolis	GO	Aw	857	1	2,834	69.0	1.10	56.2	48.8	1.07
Ariquemes	RO	Am	166	1	2,495	75.7	2.30	91.3	83.5	0.82

Municipalities ordered by climate type. Köppen class and mean altitude follow Alvares et al. (2013); the two Uruguayan municipalities enter as Cfa, the class that Inumet (2026) assigns to the whole national territory, and have no altitude. Station THI is obtained by subtracting the bias from microenvironment THI. Stress: hours with THI ≥ 68. Indoor/outdoor amplitude: ratio between the mean amplitude of the daily temperature cycle in the microenvironment and at the station, computed only with sensors of at least 168 hours; "n.a." marks a municipality with no sensor reaching that minimum. Values above 1 are physically implausible in a roofed barn.

Table S3 Mean THI bias (microenvironment − station) by band of regional station THI.

Station THI band	Mean bias (THI)	Municipalities
< 50	4.69	20
50 to 55	3.35	26
55 to 60	2.91	29
60 to 65	2.35	30
65 to 68	1.27	29
68 to 72	0.84	27
72 to 76	−0.07	25
≥ 76	0.01	16

Mean of municipal means within each band. The number of municipalities varies because not all of them record hours in every band.

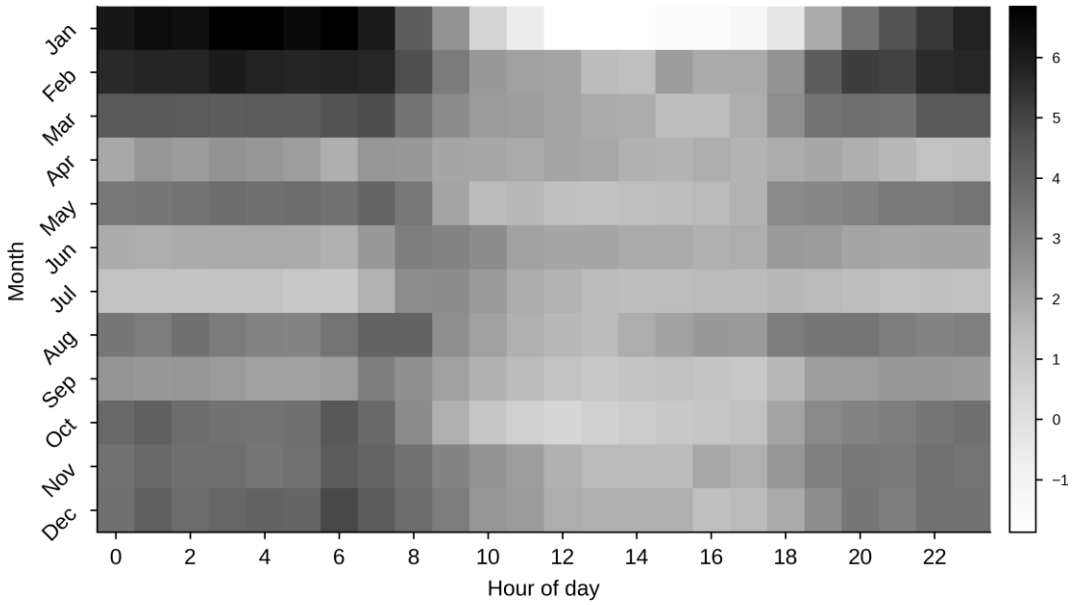


Fig. S1 Mean THI bias (microenvironment − station) by month and hour of the day, computed as the mean of municipal means. Darker shades indicate greater underestimation by the station.