Mitigating Radiometric Bias in Unmanned Aerial Vehicle Thermal Infrared Imagery Using Temperature-Controlled Reference Plates

Gaia Cervini, Jinha Jung, Nicholas A. Roberts, and Konstantina Gkritza

Abstract

Temperature retrieval from unmanned aerial vehicle (UAV) thermal imagery commonly relies on proprietary manufacturer radiometric calibration software. However, the correction procedures used by these workflows are generally undisclosed, and some user-defined parameters, particularly altitude, are restricted to ranges that do not reflect real-world conditions, risking the introduction of bias and higher radiometric uncertainty. Accordingly, this study evaluates the accuracy of proprietary radiometric calibration by examining how its performance changes with flight altitude, while also proposing a practical method to correct residual bias after proprietary processing. Thermal imagery is collected with a DJI Matrice 300 RTK equipped with a Zenmuse H20T sensor at 15, 30, 45, and 60 m. Two temperature-controlled plates (TCPs) spanning cold and hot regimes are used to validate three DJI Thermal SDK workflows: (1) default settings; (2) custom settings using measured emissivity, flight altitude, humidity, and reflected apparent temperature ($T_{reflected}$); and (3) custom settings with per-image tuning of $T_{reflected}$. Residual bias is subsequently corrected using empirical line calibration (ELC) derived from TCP observations. Across all proprietary workflows, temperature error increases with altitude, indicating progressive gain and offset bias not fully corrected by manufacturer processing. Default settings produce the largest error, with mean absolute error increasing from 4.39°C at 15 m to 11.69°C at 60 m. Custom settings reduce error moderately (3.86°C to 9.43°C), while $T_{reflected}$ tuning improves close-range performance but does not eliminate altitude-dependent degradation (0.99°C to 8.57°C). Application of ELC substantially reduces residual bias across workflows, with best performance achieved by combining $T_{reflected}$ tuning with ELC, resulting in sub-1°C accuracy across altitudes. These results demonstrate that proprietary UAV thermal image processing alone is insufficient to ensure robust temperature retrieval but that a simple empirically calibrated postprocessing workflow using TCPs can mitigate altitude-dependent bias and enable high-accuracy temperature estimation in operational remote sensing applications

Introduction

Unmanned aerial vehicle (UAV) thermal infrared (TIR) imagery has become an important tool for applications requiring spatially detailed temperature information, including infrastructure inspection, precision agriculture, and environmental monitoring (Pascucci *et al.* 2008; Turner *et al.* 2014; Messina and Modica 2020; Lin *et al.* 2025). Relative to satellite platforms, UAV systems can acquire thermal data at much finer spatial resolution, can be deployed on demand, and can operate below cloud cover, making them attractive for site-specific and time-sensitive measurements (Cao *et al.* 2019; Sagan *et al.* 2019; Jiang *et al.* 2024).

Despite these advantages, collecting accurate surface temperatures from UAV TIR imagery remains a challenge (Han *et al.* 2020). Thermal cameras infer temperature by measuring the TIR radiance emitted from a surface rather than sensing temperature directly (Yan *et al.* 2012). Consequently, the measured temperature depends on several surface-, path-, and sensor-dependent factors that can change during flight (Maes *et al.* 2017; Guo *et al.* 2019; Daniels *et al.* 2023; Szostak *et al.* 2023; Parisi *et al.* 2024). At the target, the signal is affected by surface emissivity and by reflected apparent temperature from the sky and surrounding environment. Along the path, atmospheric transmittance and upwelling emission vary with humidity, air temperature, and sensor–target distance, so the at-sensor radiance may change even when the surface itself is unchanged (Maes *et al.* 2017; Acorsi *et al.* 2020; Song *et al.* 2025). At the sensor level, uncooled microbolometer systems are vulnerable to warm-up effects, thermal drift, spatial nonuniformity, and vignetting, particularly under variable in-flight conditions (Kelly *et al.* 2019; Aragon *et al.* 2020; Malbéteau *et al.* 2021; Yuan and Hua 2022; Wang *et al.* 2023). These effects can be amplified during UAV operation by rapid changes in ambient conditions, wind exposure, and propeller-induced airflow around the sensor (Acorsi *et al.* 2020; Messina and Modica 2020; Lin *et al.* 2025). Consequently, apparent temperature differences in UAV TIR imagery do not always correspond to true surface temperature variability.

A substantial body of work has sought to mitigate these disturbances and improve UAV thermal accuracy. Previous studies have examined the effects of wind, time of day, modest changes in meteorology, mosaicking strategy, and sensor instability on retrieved temperatures (Maes *et al.* 2017; Malbéteau *et al.* 2021; Szostak *et al.* 2023; Wang *et al.* 2023). Others have proposed corrections for drift, vignetting, and nonuniform camera response (Aragon *et al.* 2020; Yuan and Hua 2022; Wang *et al.* 2023). Several studies have also shown that in-scene references can provide a practical basis for empirical correction (Kelly *et al.* 2019; Han *et al.* 2020; Song *et al.* 2025).

However, an important gap remains in how radiometric accuracy is established in practice. Rather than deriving calibration relationships directly from raw sensor response, many users rely on proprietary manufacturer software to convert thermal image digital numbers into temperature values. These workflows are attractive because they are convenient and embedded in common postprocessing pipelines, but the underlying correction procedures are rarely described in enough detail to permit independent evaluation (Cao *et al.* 2019; Daniels *et al.* 2023; Jiang *et al.* 2024). As a result, radiometric processing often functions as a black box: users may specify parameters such as emissivity, distance, humidity, and reflected apparent temperature but cannot clearly determine how

Gaia Cervini, Jinha Jung, and Konstantina Gkritza are with the Lyles School of Civil and Construction Engineering, Purdue University.

Nicholas A. Roberts is with the Department of Mechanical and Aerospace Engineering, Utah State University.

Corresponding author: Gaia Cervini (gcervini@purdue.edu)

Received February 19, 2026, accepted June 6, 2026.

Photogrammetric Engineering & Remote Sensing
Vol. 92, No. X, MONTH 2026, pp. 1–xx.
0099-1112/26/1–xx

doi: 10.14358/PERS.26-00024R2

those parameters are incorporated internally or why errors emerge under certain conditions (Jiang *et al.* 2024; Zhang *et al.* 2025).

The problem is compounded when proprietary tools impose practical restrictions on user-defined inputs that do not reflect realistic acquisition conditions. For example, the DJI Thermal SDK radiometric conversion tool accepts object distance only within the range 1 to 25 m, even though UAV surveys are commonly performed at greater flight altitudes. When actual acquisition conditions exceed the supported parameter range, inputs are clipped, potentially introducing systematic bias and reducing correction fidelity (Szostak *et al.* 2023; Jiang *et al.* 2024; Parisi *et al.* 2024). This limitation is particularly significant for altitude, as sensor–target distance is directly tied to atmospheric attenuation and therefore to the magnitude of radiometric uncertainty (Maes *et al.* 2017; Acorsi *et al.* 2020; Messina and Modica 2020; Zhu *et al.* 2024).

Taken together, these constraints raise fundamental questions about whether proprietary calibration outputs remain reliable as acquisition conditions diverge from those implicitly assumed by the software. If manufacturer processing introduces errors or leaves bias unresolved, those errors can propagate into downstream thermal products and any temperature-based metrics derived from them (Maes *et al.* 2017; Kapil *et al.* 2023; Wang *et al.* 2023; Lin *et al.* 2025). For both research and operational use, this weakens inference, cross-mission comparability, and confidence in derived products, underscoring the need for transparent, field-deployable validation and correction frameworks.

Accordingly, this study evaluates the performance of proprietary radiometric calibration in UAV TIR imagery under realistic acquisition conditions and develops a practical method to correct the residual bias that remains after manufacturer processing. Specifically, the objectives of this study are to: (1) quantify how the accuracy of proprietary radiometric calibration changes with flight altitude; (2) compare the performance of default and user-parameterized proprietary workflows; and (3) evaluate whether empirical line calibration (ELC) can mitigate the residual bias left by proprietary processing. To address these objectives, thermal imagery of in-scene temperature references is acquired across a range of flight altitudes using a UAV-mounted TIR sensor, and three proprietary processing workflows representing default and user-parameterized configurations are assessed. Residual bias is subsequently corrected using ELC derived from temperature-controlled plate (TCP) observations. In doing so, this study provides a direct assessment of the reliability of proprietary UAV thermal calibration workflows and demonstrates a practical postprocessing strategy for improving quantitative temperature retrieval in operational UAV thermography.

Methods

The UAV TIR image calibration and evaluation was carried out using a three-stage methodology: (1) data acquisition, (2) TIR image radiometric calibration using proprietary software, and (3) TIR image residual bias correction using ELC. Each stage comprised multiple steps, detailed in the following subsections. An overview of the full methodological framework is provided in Figure 1.

Study Site and Materials

Field experiments were conducted at two locations in West Lafayette, IN: Purdue University's Beck Agricultural Center at the Agronomy Center for Research and Education (ACRE) and the Tippecanoe County Amphitheater Park. The ACRE site was selected for its proximity (150 m) to the ACRE weather station, whose measurements were used for radiometric calibration. However, because the area was not consistently free of bystanders, the latter portion of data collection was moved to the Amphitheater Park (9 km from ACRE), which provided a reliably open and safe operating environment. Meteorological observations from the ACRE weather station were obtained via the Purdue Mesonet Data Hub (Indiana State Climate Office 2025), which provides 5-min measurements of air temperature, relative humidity, wind speed, and solar radiation during each flight.

Thermal imagery was collected using a DJI Matrice 300 RTK quadcopter equipped with a Zenmuse H20T multi-sensor payload (SZ DJI Technology Co., Ltd., Shenzhen, China), which integrates a TIR camera, wide and zoom visible cameras, and a laser rangefinder. The thermal sensor measures long-wave infrared radiation over the 8- to 14-μm band at a native resolution of 640×512 pixels. The thermal lens has a diagonal field of view (FOV) of 40.6° and a focal length of 13.5 mm. According to manufacturer specifications, the system supports an object temperature measurement range of −40°C to 150°C in low gain mode, at a stated absolute accuracy of ±2°C or 2%, whichever is greater (DJI 2020).

To support validation and empirical correction of retrieved temperatures, two ElectraCOOL TCP50 thermoelectric cold plates (Advanced Thermoelectric, Melbourne, FL) were placed within the scene during flights (Figure 2). Each plate was controlled using a TLK38-S temperature controller (Ascon Technologic Srl, Pavia, Italy) with an accuracy of ±2°C and powered by a 2300 W portable generator. The active surface of each plate is 10 × 10 cm and consists of black anodized aluminum. Based on values reported in the MODIS UCSB Emissivity Library (Zhang 1999), the emissivity of black anodized aluminum over the 8- to 14-μm band is ~0.96.

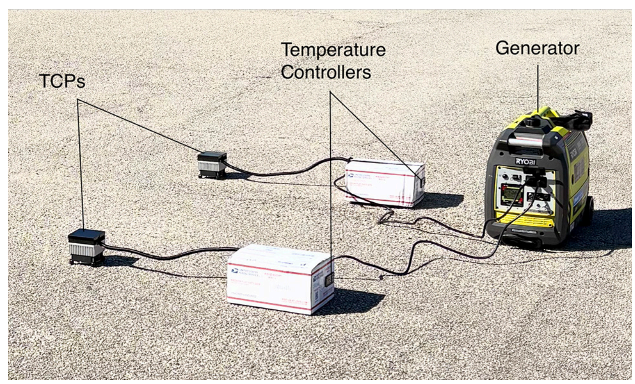

Figure 2. In-scene temperature reference system used for thermal measurement validation.

Data Acquisition

The data were collected on 1 September 2025 (11:30 A.M. to 2:30 P.M. ET) and 2 September 2025 (8:30 to 11:00 A.M. ET) to capture contrasting time-of-day and meteorological conditions. Concurrent observations of air temperature, relative humidity, wind speed, and solar radiation from the ACRE weather station are summarized in Figure 3.

Prior to each flight, the thermal payload was powered on and allowed to warm up for 15 minutes to promote stabilization of the focal

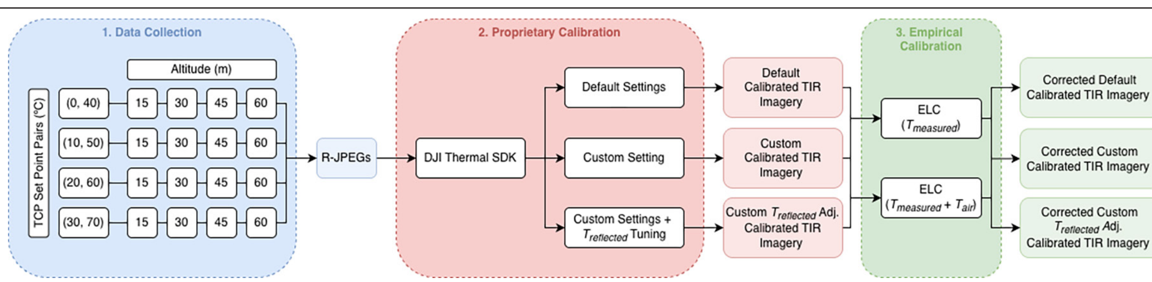

Figure 1. Unmanned aerial vehicle (UAV) thermal infrared (TIR) image calibration framework.

plane array temperature and reduce start-up–related signal drift. The TCPs were also powered on using the portable generator and held at a pair of setpoints until they reached steady state. These setpoints were (0°C, 40°C), (10°C, 50°C), (20°C, 60°C), and (30°C, 70°C), maintaining a constant 40°C separation to (1) span a broad range of reference temperatures and (2) provide "bracketing" references that approximated the lower and upper bounds of temperatures observed in the scene.

For each setpoint pair, the UAV hovered directly above the TCPs at nominal altitudes of 15, 30, 45, and 60 m AGL and captured thermal frames in nadir view by pitching the gimbal to −90°. Example TIR and RGB imagery acquired at each altitude is shown in Figure 4. To minimize vignetting-related bias and ensure consistent sampling geometry, the TCPs were positioned near the center of the thermal frame in all captures. Accordingly, the present experiment was designed to evaluate calibration performance under a controlled central-FOV configuration rather than to test spatial generalization across the full image plane. Throughout the experiment, UAV batteries were hot-swapped as needed, avoiding shutdown of the aircraft and sensor between flight segments.

Proprietary Radiometric Calibration

Once collected, the raw thermal images (R-JPEG) were radiometrically calibrated using thermal_parser, an open-source GitHub repository providing a Python interface to the DJI Thermal SDK to perform radiometric conversion (GitHub n.d.). For each R-JPEG, the workflow outputs a 640×512 temperature array and applies radiometric correction using four user-specified parameters: emissivity ($\epsilon \in [0.10, 1.00]$), object distance ($h \in [1, 25]$ m), relative humidity ($\phi \in [20, 100]$%), and reflected apparent temperature ($T_{reflected} \in [-40, 500]$°C). The temperature array produced by each conversion was written to a GeoTIFF for downstream analysis using GDAL. Because geolocation information is not preserved through DJI Thermal SDK conversion, exiftool was used to extract geographic metadata from each original R-JPEG and write the corresponding georeferencing information to the output GeoTIFF.

Two radiometric calibration modes were evaluated using built-in functions from the thermal_parser package: parse and parse_dirp2. In the first mode (parse), images are calibrated using default parameters read from the image metadata ($\varepsilon = 1.0$, $h = 5$ m, $\phi = 70$%, and $T_{reflected} = 23$°C). In the second mode (parse_dirp2), the user can explicitly specify the parameters: we set $\varepsilon = 0.96$ for the TCP surface and h to the nominal flight altitude for each capture; ϕ was taken from the ACRE weather station and matched to each image using the closest-in-time station observation; following manufacturer guidance, we used ambient air temperature (T_{air}) as a proxy for $T_{reflected}$ (DJI 2022); T_{air} was likewise assigned on an image-by-image basis using the closest-in-time ACRE observation. Applying both calibration modes to the same imagery produced two data sets: a *default*-calibrated data set generated with parse and a custom-calibrated data set generated with parse_dirp2.

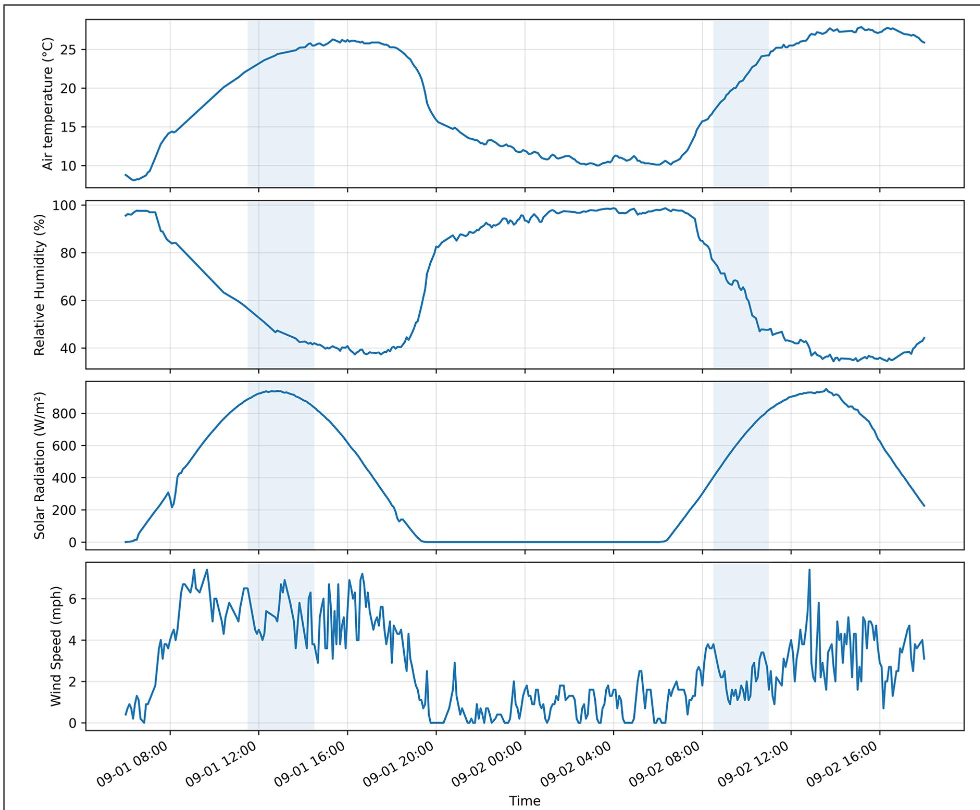

Figure 3. Time series of meteorological conditions during unmanned aerial vehicle (UAV) thermal data collection. Panels show 5-minute observations from the Agronomy Center for Research & Education (ACRE) weather station. Shaded vertical bands indicate data collection windows.

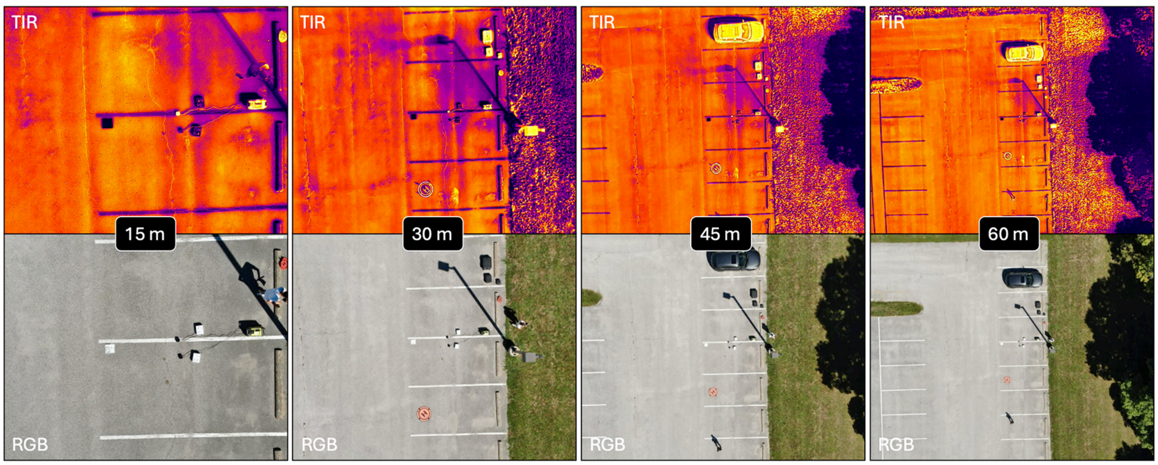

Figure 4. Representative coregistered thermal infrared (TIR) and red–green–blue (RGB) imagery of the same scene acquired at four flight altitudes.

To test whether TCP references could improve accuracy, we applied an additional per-image $T_{reflected}$ tuning step on top of the *custom*-calibrated data set. For each thermal image i, we reran parse_dirp2 over a uniform grid of candidate $T_{reflected}$ values spanning the full range permitted by the SDK. For each candidate $T_{reflected}$ value, we extracted the camera-measured temperature ($T_{measured}$) at the center pixel of the cold and hot TCPs and computed the corresponding errors relative to the TCP reference temperatures ($T_{reference}$). We then selected $\hat{T}_{reflected_i}$ that minimized the sum of absolute errors across both TCPs (L_1 objective) and reprocessed the image using $\hat{T}_{reflected_i}$. The resulting temperatures define the *custom* $\hat{T}_{reflected}$ *adjusted* data set. We chose $T_{reflected}$ as the tuning parameter because it was the only radiometric input that was not directly measured and instead approximated using T_{air}, whereas ε, h, and ϕ were either known a priori or obtained from independent measurements. The resulting $T_{measured}$ from each calibration mode, *default*, *custom*, and *custom* $T_{reflected}$ *adjusted*, were then validated against the corresponding $T_{reference}$.

Empirical Radiometric Calibration

To further improve radiometric accuracy beyond what could be achieved with the proprietary workflow, we developed an empirical line calibration (ELC) model using $T_{reference}$ from the TCPs. ELC is widely used for TIR image radiometric calibration and offers a parsimonious alternative to more complex mappings (e.g., Planck-based exponential fits or neural networks) (Guo *et al.* 2019; Kelly *et al.* 2019; Rakha *et al.* 2022).

A nested-model F-test ($p \leq 8.6 \times 10^{-5}$) of $T_{measured}$ versus $T_{reference}$ revealed that observations from the cold and hot TCPs formed separated clusters with different linear trends, suggesting that a single global slope and intercept was insufficient. To account for this stratification, we modeled the relationship using a first-order correction with an interaction between $T_{measured}$ and TCP type (hot vs. cold):

$$T_{reference} = \beta_0 + \beta_1 T_{measured} + \beta_2 H + \beta_3 (T_{measured} \cdot H) \quad (1)$$

where H is an indicator variable ($H = 1$ for the hot TCP and $H = 0$ for the cold TCP). This specification allows both the intercept and the slope to differ between plates, with β_2 capturing an additive shift between cold and hot regimes and β_3 capturing a differential multiplicative bias between regimes. To examine altitude dependence, we fit the interaction model separately within each nominal flight-altitude bin, which allows for systematic changes in the empirical correction with sensor–target distance. As in standard ELC, the corrected temperature was defined as the model prediction, $T_{corrected} \equiv \hat{T}_{reference}$.

Because environmental conditions and sensor–target distance can influence the apparent radiance received by the sensor, we additionally tested whether incorporating meteorological covariates reduced residual error. We augmented the calibration data set with time-matched meteorological variables and screened candidate predictors for multicollinearity using pairwise Pearson correlations, prioritizing variables that were associated with $T_{measured}$ while remaining noncollinear. Air temperature (T_{air}) emerged as the most informative variable under these criteria. We therefore extended Equation 1 to include both a main effect of T_{air} and an interaction between T_{air} and TCP type:

$$T_{reference} = \beta_0 + \beta_1 T_{measured} + \beta_2 H + \beta_3 (T_{measured} \cdot H) + \beta_4 T_{air} + \beta_5 (T_{air} \cdot H) \quad (2)$$

This formulation permits the influence of T_{air} on the calibration to differ between the cold and hot TCP regimes.

ELC models were fit separately for each proprietary calibration approach and flight-altitude bin. Performance was assessed by comparing $T_{measured}$ and $T_{corrected}$ against $T_{reference}$ using robust error metrics, such as the mean absolute error (MAE), that are less sensitive to occasional outlier images than mean-squared measures.

Sample Efficiency Analysis

In addition to improving accuracy, we sought to minimize the amount of calibration data required to achieve near-asymptotic performance. Specifically, we quantified how error decreases as the number of training images increases and identified the smallest training-set size beyond which additional calibration images yield negligible improvements. This analysis was performed separately for each proprietary calibration approach and flight-altitude using Equation 2.

For each group (calibration approach × altitude), let $\mathcal{I}$ be the set of usable images containing TCP references, with $M = |\mathcal{I}|$. For each training-set size $k \in \{1, \ldots, M-1\}$, we repeatedly sampled k images for calibration, fitting the model using all TCP observations in those images, and evaluating on the remaining $M - k$ held-out images. Performance was measured as the MAE on held-out TCP observations. Repeating this procedure yields a distribution of MAE values for each k; we report the median as the learning curve and the P_{10} to P_{90} interval as a measure of sensitivity to training-set composition.

To obtain a minimum-data recommendation, we identify the point of diminishing returns. For each training-set size k, we summarize performance by the median held-out MAE across random splits and compare it to a near-full-data baseline defined as the median held-out MAE when training on $M - 1$ images (leaving one image for testing). We then define the minimum sufficient training size, k^*, as the smallest k for which the median held-out MAE is no more than ϵ above this baseline, where ϵ=0.05°C. In other words, k^* is the smallest number of calibration images needed to achieve essentially the same typical held-out accuracy as using nearly all available images.

Results

Proprietary Calibration

Across all proprietary calibration configurations, temperature accuracy degraded with increasing flight altitude, consistent with an increasing atmospheric path length and greater sensitivity to parameter misspecification in the radiative transfer terms (Figure 5a). This altitude dependence appeared as both decreasing precision (R^2) and increasing systematic gain/offset changes in $T_{measured}$ versus $T_{reference}$ fits (decreasing slope and increasing intercept), indicating a progressively larger range-dependent bias with altitude.

The *default* workflow yielded the largest errors and the strongest dependence on flight altitude (Figure 5b). The overall MAE increased approximately linearly with altitude (slope ~ +1.62°C/10 m), rising from 4.39°C at 15 m to 11.69°C at 60 m. This altitude sensitivity was driven primarily by the hot TCP: hot TCP errors (slope ~ +2.88°C/10 m) increased much more rapidly with altitude than cold TCP errors (slope ~ +0.37°C/10 m), with MAE climbing from 6.66°C at 15 m to 19.78°C at 60 m.

The *custom* workflow reduced overall error and partially mitigated altitude sensitivity relative to *default* calibration. The overall MAE increased more moderately with altitude (slope ~ +1.22°C/10 m), from 3.86°C at 15 m to 9.43°C at 60 m. However, this improvement was not uniform across temperature regimes: the cold TCP exhibited a higher baseline MAE than under default calibration, increasing by 2°C from 15 to 30 m and then plateauing at higher altitudes. Conversely, the hot TCP remained substantially more accurate than *default* calibration at all altitudes, although its MAE still increased with altitude.

The *custom* $T_{reflected}$ *adjusted* workflow achieved the lowest overall error (MAE = 0.99°C to 8.57°C at 15 to 60 m) but still degraded with altitude (slope ~ +1.68°C/10 m). The most distinctive outcome of $T_{reflected}$ tuning was the near elimination of cold-plate error: the cold TCP MAE was effectively 0°C across all altitudes. In contrast, the hot TCP exhibited the steepest altitude sensitivity of all three modes, despite relatively consistent low-variance error within each altitude bin. This divergence between plates is expected because within a given frame, the cold and hot TCP biases typically occur with opposite sign (*i*.e., colder targets are overestimated while hotter targets are underestimated). Since changing $T_{reflected}$ produces a largely uniform shift in the retrieved temperature field, only one regime can be improved at the expense of the other, so a single $T_{reflected}$ value cannot simultaneously remove error for both TCPs.

Overall, altitude-dependent error was driven primarily by the hot TCP. $T_{reflected}$ tuning nearly eliminated cold TCP error but introduced the strongest altitude sensitivity for the hot TCP, whereas the *custom* workflow reduced hot TCP degradation relative to *default* workflow but retained a higher baseline error for the cold TCP.

Empirical Calibration

Applying an altitude-specific ELC substantially reduced gain and offset biases across all proprietary-based calibration modes (Figure 6). Improvements were most pronounced for cases with the largest proprietary calibration error, consistent with ELC acting primarily as a bias-correction layer. Across all proprietary calibration approaches and altitudes, MAE decreased markedly after ELC, and the inclusion of T_{air} generally provided additional error reduction that became more substantial at higher altitudes (Figure 7). In parallel, the linear association

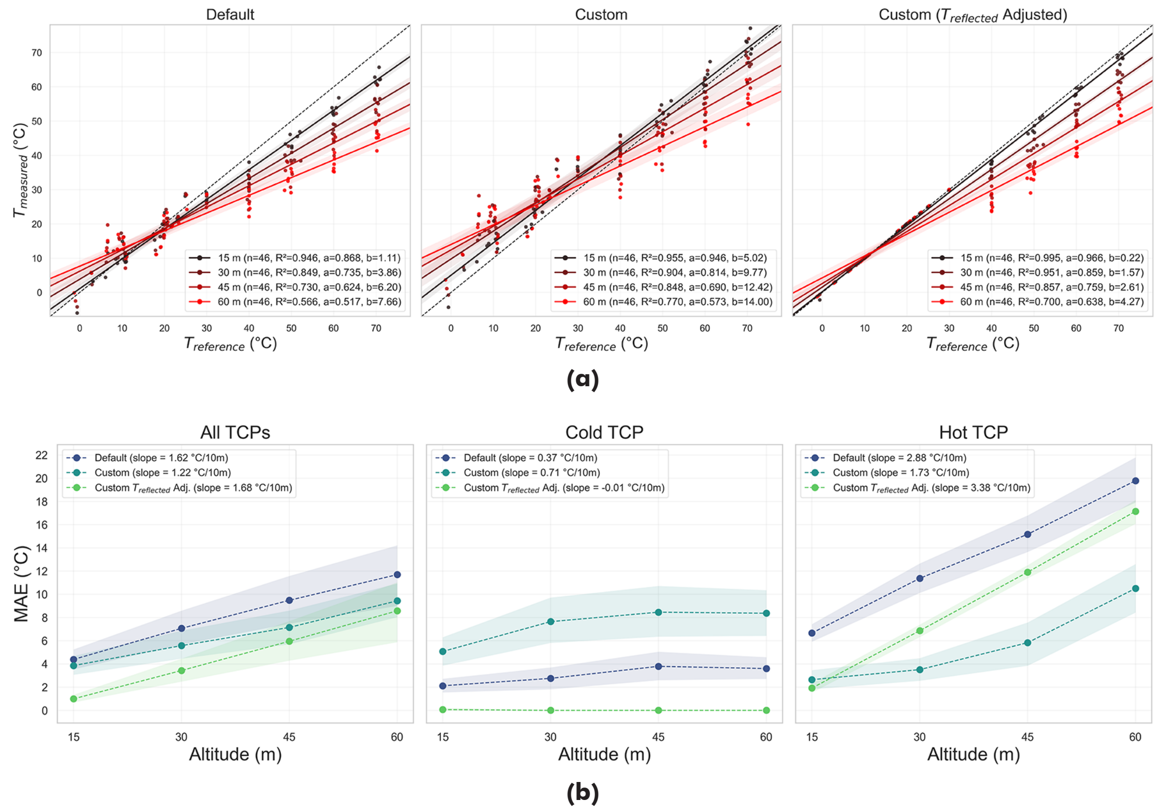

Figure 5. (a) Increasing gain-offset bias and degraded accuracy with increasing altitude for each proprietary calibration approach. (b) Mean absolute error (MAE) versus flight altitude for proprietary calibration configurations and temperature-controlled plate (TCP) temperature. The shaded bands show 95% confidence intervals.

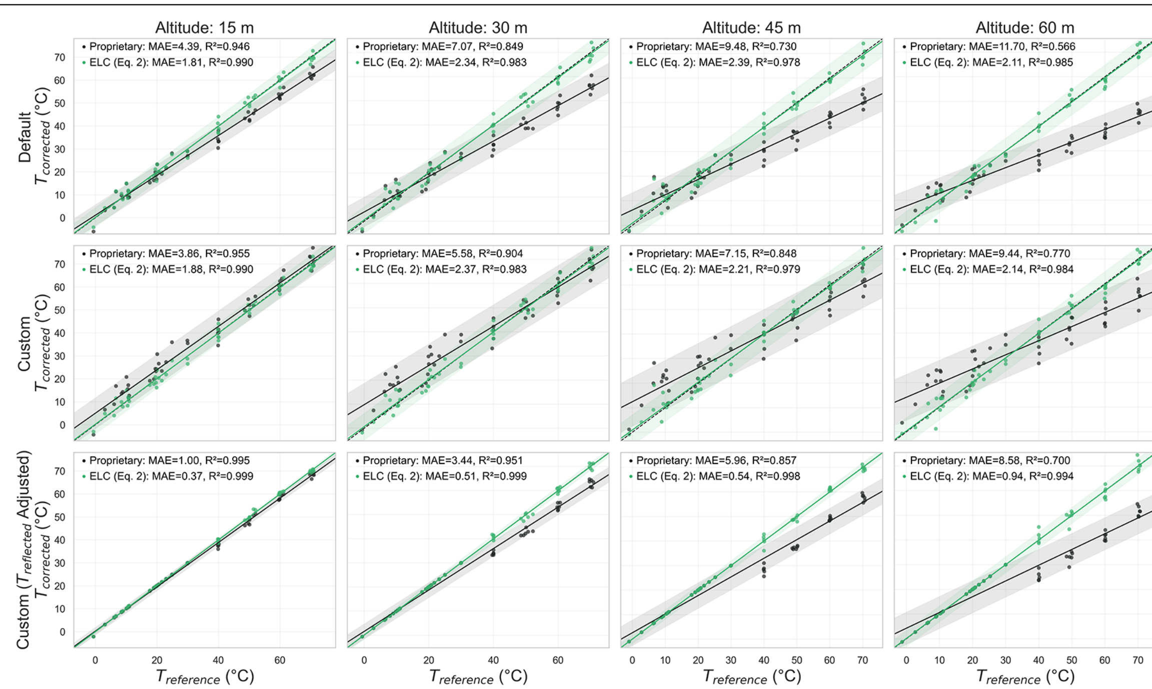

Figure 6. Empirical calibration results relative to proprietary calibration. Solid lines denote least-squares fits, shaded bands show 95% confidence intervals, and the dashed line indicates the 1:1 relationship.

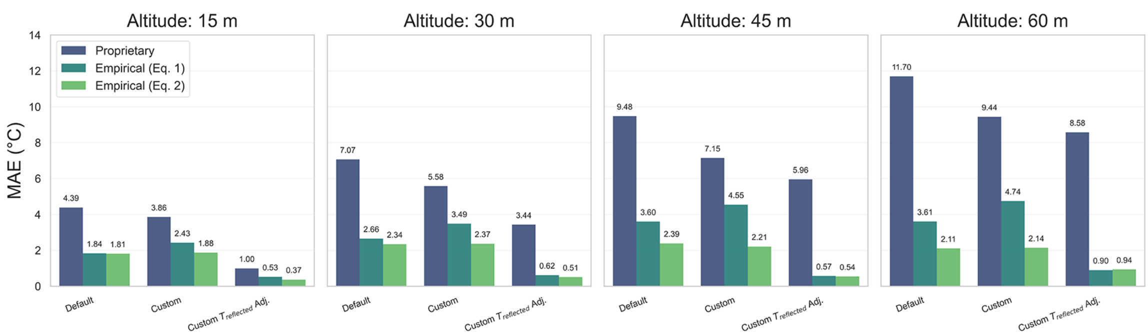

Figure 7. Incremental mean absolute error (MAE) reduction across altitudes and calibration configurations.

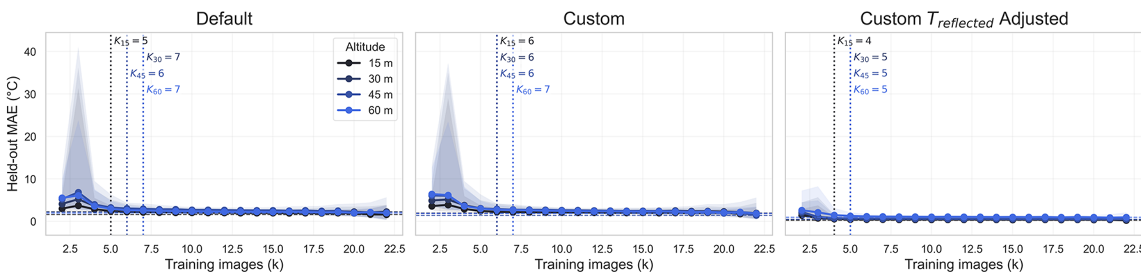

Figure 8. Equation 2 learning curve example for the custom Treflected-adjusted approach quantifying how error changes as the number of training images increases. The solid line shows the median held-out mean absolute error (MAE) over repeated random splits as the number of k images increases, while the shaded band indicates the P10 to P90 range. Horizontal dashed lines mark the baseline performance.

between $T_{corrected}$ and $T_{reference}$ strengthened, as evidenced by R^2 values increasing toward unity.

Under the *default* configuration, ELC reduced MAE by up to 69%. Augmenting the interaction model with T_{air} yielded a modest benefit at low altitude but a progressively larger benefit with increasing altitude, culminating in an additional MAE reduction of 1.50°C at 60 m. This pattern indicates that while a first-order correction removes a large fraction of the systematic bias present in the *default* SDK output, temperature-dependent or environment-mediated residuals become increasingly important as sensor–target distance increases. In practical terms, T_{air} appears to capture part of the altitude-linked error component, whether due to atmospheric effects or unmodeled radiometric terms, that grows more influential at larger distances.

For the *custom* configuration, ELC produced more moderate reductions when using $T_{measured}$ alone, with improvements ranging from ~36% at 15 m to 50% at 60 m. However, incorporating T_{air} provided substantially larger incremental gains than in the *default* calibration case, particularly at higher altitudes where the MAE decreased by an additional 2.60°C at 60 m. Overall, *custom* radiometric parameterization mitigates some systematic errors relative to the *default* workflow but leaves a residual component that is both altitude-linked and strongly coupled to ambient thermal conditions.

In contrast, the *custom* $T_{reflected}$ *adjusted* configuration exhibited the greatest overall improvements with ELC. MAE reduced dramatically at all altitudes (up to 90% reduction), falling below 1°C across the full altitude range. In this case, adding T_{air} provided little to no additional improvement and, at the highest altitude, produced a negligible degradation. This behavior is consistent with $T_{reflected}$ adjustment already removing much of the structured radiometric mismatch that T_{air} would otherwise help explain; the remaining residuals are largely captured by the TCP-interaction correction based on $T_{measured}$ alone. From an operational standpoint, when $T_{reflected}$ has been tuned per image, the simpler Equation 1 ELC model is typically sufficient, and the added complexity of including T_{air} is unlikely to be justified.

Overall, *custom* $T_{reflected}$ *adjusted* achieved the lowest errors at every altitude (MAE ~ 0.37°C to 0.94°C), substantially outperforming *default* and *custom* even after empirical correction. For the *default* and *custom* SDK outputs, inclusion of T_{air} is valuable for controlling altitude-linked residuals, whereas for $T_{reflected}$-adjusted outputs, the temperature field becomes sufficiently "bias linear" that T_{air} contributes minimal additional predictive information.

Sample Efficiency

Across all calibration approaches and altitudes, the learning curves in Figure 8 indicate that empirical calibration is highly sample efficient as near-asymptotic median MAE is typically achieved with only four to seven calibration images. Nonetheless, the approaches differ markedly in the error level at which they stabilize and in how rapidly the split-to-split variability collapses. In general, the *default* and *custom* workflows exhibit higher stabilized error and broader variability bands at small *k*, indicating stronger dependence on which images are included in the calibration subset. In contrast, the *custom* $T_{reflected}$ *adjusted* workflow stabilizes at a substantially lower median MAE and exhibits narrower uncertainty bands, consistent with $T_{reflected}$ tuning reducing structured radiometric mismatch prior to empirical correction and making the residual error more uniform across images.

Discussion

The results show that proprietary radiometric calibration alone leaves substantial residual error in camera-derived temperatures and that a small number of well-controlled reference observations are sufficient to correct both multiplicative and additive bias. In the linear model, the slope term corrects systematic gain error, while the intercept corrects additive shifts driven by imperfect radiometric inputs. Incorporating T_{air} further reduces residual error at higher altitudes, consistent with first-order environmental modulation of apparent radiance and atmospheric path effects increasing with sensor–target distance.

The significance of the interaction term in the ELC models highlights that a single reference target is generally insufficient for robust thermal calibration. Cold and hot targets cluster differently and exhibit distinct error structure, indicating that the camera does not apply a uniform bias across the full temperature range within a frame. Practically, at least two references spanning low and high temperatures are needed to identify both slope and offset reliably and to ensure the correction remains valid across the temperature distribution encountered in the scene. This also motivates selecting target setpoints that bracket or exceed the expected scene temperatures, as this can improve identifiability of the calibration parameters and reduce extrapolation error when the scene contains temperatures near the tails of the distribution.

Altitude also emerges as a primary driver of uncorrected error and calibration reliability. For any of the tested radiometric calibration approaches, error and sensitivity to environmental conditions increase with altitude, reflecting reduced effective spatial resolution, increased mixed-pixel contamination, and greater opportunity for atmospheric and viewing-geometry effects to perturb the measured radiance. Consequently, altitude should be chosen based on required thermal accuracy as well as coverage. Applications with stringent temperature tolerances benefit from lower-altitude acquisition, whereas higher-altitude missions trade accuracy for efficiency and therefore require greater reliance on post-hoc correction and in-scene validation.

Although full operational deployment is beyond the direct scope of this study, the results suggest a four-stage workflow for field implementation. First, a limited subset of TCP-bearing images should be acquired periodically throughout the mission, with the sample-efficiency analysis suggesting that at least four to seven such images are typically sufficient to estimate a stable correction model. Second, the manufacturer SDK radiometric conversion using the best available radiometric inputs (default or custom) should be applied to obtain physically interpretable apparent temperatures. Third, when higher accuracy is required, $T_{\text{reflected}}$ should be optimized on the TCP-bearing calibration subset, and those values should be transferred to neighboring non-TCP images using a temporally matched strategy, such as nearest-neighbor assignment within a stable flight block or linear interpolation between successive calibration frames. Fourth, the ELC model should be fit using the TCP-bearing calibration subset only, and the resulting coefficients should be applied to all remaining survey images acquired within the same flight. Under this workflow, TCP-bearing frames function as intermittent calibration anchors, allowing the full survey to be corrected through parameter transfer rather than by requiring visible references in every image.

This workflow is particularly valuable in realistic operating environments where radiometric inputs may be missing, uncertain, or spatially unrepresentative (e.g., no nearby meteorological stations, rapidly evolving weather, or a poorly constrained $T_{\text{reflected}}$ due to complex surrounding radiative environments). In those cases, a modest number of colocated reference measurements provide a robust mechanism to correct temperature accuracy while also delivering a transparent and repeatable quality-control check that can be carried across surveys, altitudes, and operating environments. Overall, the results motivate three practical implications for accurate UAV-based TIR imaging: (1) calibration must address both gain and offset error, (2) reference targets must span the operational temperature range, and (3) acquisition altitude should be treated as an explicit design variable for radiometric accuracy, not solely a mapping-efficiency parameter.

Several limitations of this study should be noted. First, the TCPs were small relative to the ground sampling distance at higher altitudes, which constrained validation to ≤60 m. At the upper end of the tested range, the TCPs occupied few pixels, increasing sensitivity to mixed-pixel contamination and thereby confounding radiometric error with spot-size limitations. Accordingly, the observed altitude-related bias at higher flight levels should be interpreted as an operational temperature-error signal for the tested target-size/altitude configuration rather than as a clean mechanistic separation of atmospheric and optical-spatial error sources. For the same reason, the reported correction performance should not yet be assumed to transfer unchanged to large, spatially homogeneous targets such as rooftops or water bodies, where mixed-pixel effects are less pronounced and the balance of error sources may differ. Dedicated validation using larger, well-resolved reference surfaces is needed to assess transferability in those settings. Second, meteorological inputs were obtained from a nearby weather station rather than measured at the flight site, which can introduce mismatch in air temperature and humidity, thereby degrading both the proprietary parameterization and the empirical covariate relationships. Deploying a compact on-site sensor package synchronized to the flight timeline would reduce this uncertainty. Third, data collection was limited to daytime conditions over two field dates, so the T_{air}-dependent ELC model (Equation 2) should be interpreted as validated only for the environmental regime sampled in this study, rather than as a season-independent or nighttime-ready correction. Conditions with substantially different radiative forcing, sky temperature, humidity structure, or sensor thermal behavior may alter both the magnitude and form of the correction. Fourth, the $T_{\text{reflected}}$ adjustment should be interpreted as an empirical upstream bias-reduction step within the tested imaging regime rather than as proof that the remaining error is purely radiometric in origin. Fifth, since optimization and validation were both performed using TCPs intentionally located near the center of the FOV, the study does not provide an independent spatial generalization test across the full image plane. Thus, the results should not be interpreted as demonstrating correction of edge-of-frame vignetting or broader spatial nonuniformity across the sensor; instead, they quantify the accuracy achievable under a controlled central-FOV geometry. Finally, although the study proposes a practical parameter-transfer strategy for extending the workflow to larger surveys in which most images do not contain TCPs, the operational implementation of this approach was not directly evaluated in the present study.

Future work should (1) extend validation to higher altitudes using larger targets and explicitly separate path-length atmospheric effects from mixed-pixel limitations; (2) incorporate synchronized on-site meteorological measurements to better resolve environment-induced biases and improve transferability; (3) evaluate performance across a broader diurnal range, including low-solar and nighttime conditions; (4) include larger, well-resolved homogeneous targets so that the transferability of the T_{air}-dependent ELC model can be evaluated beyond the daytime, small-target regime studied here; (5) include separate calibration and check targets distributed from the image center to the edges in order to assess residual vignetting, spatial nonuniformity, and the transferability of the proposed correction beyond the central-FOV configuration used here; and (6) evaluate the proposed parameter-transfer workflow under continuous mapping conditions in which only a limited subset of images contains TCPs, including assessment of how $T_{\text{reflected}}$ and ELC coefficients can be most reliably propagated to neighboring nonreference frames. Finally, benchmarking additional proprietary processing software under identical field conditions would help quantify the radiometric accuracy of other commonly used commercial workflows, enabling clearer assessment of how reliably practitioners can expect these tools to perform in operational temperature mapping.

Conclusions

This study evaluated the radiometric reliability of proprietary UAV thermal calibration workflows under realistic flight conditions and demonstrated that despite their convenience, such workflows leave substantial residual bias that increases systematically with flight altitude. Across all tested configurations, temperature accuracy degraded with increasing sensor–target distance, manifesting as coupled gain and offset errors that distort the dynamic range of retrieved temperatures. While user-specified radiometric inputs improved performance relative to default processing and per-image tuning of $T_{\text{reflected}}$ further reduced bias, none of the proprietary workflows alone produced consistently accurate temperature estimates across the tested altitude range.

ELC provided an effective and practical solution to this limitation. By leveraging a small number of in-scene temperature references, ELC removed both multiplicative and additive bias, substantially improving agreement with ground-truth measurements. The proposed interaction-based formulation further demonstrated that calibration behavior differs across temperature regimes, reinforcing the need for bracketing reference targets. Incorporating air temperature as a covariate yielded additional improvements for workflows with larger residual bias, particularly at higher altitudes, indicating that environment-linked error components become increasingly important as atmospheric path length increases. When combined with per-image $T_{\text{reflected}}$ adjustment, ELC achieved subdegree accuracy across all tested altitudes, effectively transforming proprietary outputs into quantitatively reliable temperature measurements.

Taken together, these findings clarify the role of proprietary radiometric processing in UAV thermography: rather than serving as a complete solution, it should be viewed as an initial approximation that benefits from reference-based correction. The study therefore bridges a critical gap between convenient but opaque manufacturer workflows and the accuracy requirements of quantitative thermal remote sensing. By demonstrating how residual bias can be diagnosed, modeled, and efficiently corrected under field conditions, this work provides both a validation framework and a deployable strategy for improving the reliability of UAV-derived temperature measurements in research and operational applications.

Acknowledgments

This study was supported by the Advancing Self-Sufficiency through Powered Infrastructure for Roadway Electrification (ASPIRE) Engineering Research Center funded by the National Science Foundation under Grant No. 1941524. The authors gratefully acknowledge Taeseok Lim for assistance with data collection.

References

Acorsi, M. G., L. M. Gimenez and M. Martello. 2020. Assessing the performance of a low-cost thermal camera in proximal and aerial conditions. *Remote Sensing* 12(21):3591. https://doi.org/10.3390/rs12213591.

Aragon, B., K. Johansen, S. Parkes, Y. Malbeteau, S. Al-Mashharawi, T. Al-Amoudi, C. F. Andrade, D. Turner, A. Lucieer and M. F. McCabe. 2020. A calibration procedure for field and UAV-based uncooled thermal infrared instruments. *Sensors* 20(11):3316. https://doi.org/10.3390/s20113316.

Cao, S., B. Danielson, S. Clare, S. Koenig, C. Campos-Vargas and A. Sanchez-Azofeifa. 2019. Radiometric calibration assessments for UAS-borne multispectral cameras: Laboratory and field protocols. *ISPRS Journal of Photogrammetry and Remote Sensing* 149:132–145. https://doi.org/10.1016/j.isprsjprs.2019.01.016.

Daniels, L., E. Eeckhout, J. Wieme, Y. Dejaegher, K. Audenaert and W. H. Maes. 2023. Identifying the optimal radiometric calibration method for UAV-based multispectral imaging. *Remote Sensing* 15(11):2909. https://doi.org/10.3390/rs15112909.

DJI. 2020. *Zenmuse H20 Series User Manual*, version 1.0. <https://dl.djicdn.com/downloads/Zenmuse_H20_Series/Zenmuse_H20_Series_User_Manual-EN.pdf> (last date accessed: 2026-02-07).

DJI. 2022. *DJI Thermal Analysis Tool 3 User Guide*, version 1.2. <https://www.dji.com/downloads/softwares/dji-dtat3> (last date accessed: 9 January 2026).

GitHub. n.d. SanNianYiSi/thermal_parser, thermal image parsing utilities. <https://github.com/SanNianYiSi/thermal_parser> (last date accessed: 9 January 2026).

Guo, Y., J. Senthilnath, W. Wu, X. Zhang, Z. Zeng and H. Huang. 2019. Radiometric calibration for multispectral camera of different imaging conditions mounted on a UAV platform. *Sustainability* 11(4):978. https://doi.org/10.3390/su11040978.

Han, X., J. A. Thomasson, V. Swaminathan, T. Wang, J. Siegfried, R. Raman, N. Rajan and H. Neely. 2020. Field-based calibration of unmanned aerial vehicle thermal infrared imagery with temperature-controlled references. *Sensors* 20(24):7098. https://doi.org/10.3390/s20247098.

Indiana State Climate Office. 2025. Purdue Mesonet data hub. <https://ag.purdue.edu/indiana-state-climate/purdue-mesonet/purdue-mesonet-data-hub/> (last date accessed: 9 January 2026).

Jiang, L., H. Zhao, B. Cao, W. He, Z. Yun and C. Cheng. 2024. A UAV thermal imaging format conversion system and its application in mosaic surface microthermal environment analysis. *Sensors* 24(19):6267. https://doi.org/10.3390/s24196267.

Kapil, R., G. Castilla, S. M. Marvasti-Zadeh, D. Goodsman, N. Erbilgin and N. Ray. 2023. Orthomosaicking thermal drone images of forests via simultaneously acquired RGB images. *Remote Sensing* 15(10):2653. https://doi.org/10.3390/rs15102653.

Kelly, J., N. Kljun, P.-O. Olsson, L. Mihai, B. Liljeblad, P. Weslien, L. Klemedtsson and L. Eklundh. 2019. Challenges and best practices for deriving temperature data from an uncalibrated UAV thermal infrared camera. *Remote Sensing* 11(5):567. https://doi.org/10.3390/rs11050567.

Lin, D., N. Yang, Q. Miao, X. Cui and D. Xu. 2025. True 3D thermal inspection of buildings using multimodal UAV images. *Journal of Building Engineering* 100:111806. https://doi.org/10.1016/j.jobe.2025.111806.

Maes, W. H., A. R. Huete and K. Steppe. 2017. Optimizing the processing of UAV-based thermal imagery. *Remote Sensing* 9(5):476. https://doi.org/10.3390/rs9050476.

Malbéteau, Y., K. Johansen, B. Aragon, S. K. Al-Mashhawari and M. F. McCabe. 2021. Overcoming the challenges of thermal infrared orthomosaics using a swath-based approach to correct for dynamic temperature and wind effects. *Remote Sensing* 13(16):3255. https://doi.org/10.3390/rs13163255.

Messina, G. and G. Modica. 2020. Applications of UAV thermal imagery in precision agriculture: State of the art and future research outlook. *Remote Sensing* 12(9):1491. https://doi.org/10.3390/rs12091491.

Parisi, E. I., F. Di Ciaccio, A. Masiero, A. Guarnieri and G. Tucci. 2024. Validation and error characterization in UAV-based thermal mapping, *IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, 29–31 October 2024, Padua, Italy, pp. 574–579. https://doi.org/10.1109/MetroAgriFor63043.2024.10948859.

Pascucci, S., C. Bassani, A. Palombo, M. Poscolieri and R. Cavalli. 2008. Road asphalt pavements analyzed by airborne thermal remote sensing: Preliminary results of the Venice highway. *Sensors* 8(2);1278–1296. https://doi.org/10.3390/s8021278.

Rakha, T., Y. El Masri, K. Chen, E. Panagoulia and P. De Wilde. 2022. Building envelope anomaly characterization and simulation using drone time-lapse thermography. *Energy and Buildings* 259:111754. https://doi.org/10.1016/j.enbuild.2021.111754.

Sagan, V., M. Maimaitijiang, P. Sidike, K. Eblimit, K. T. Peterson, S. Hartling, F. Esposito, K. Khanal, M. Newcomb, D. Pauli, R. Ward, F. Fritschi, N. Shakoor and T. Mockler. 2019. UAV-based high resolution thermal imaging for vegetation monitoring and plant phenotyping using ICI 8640 P, FLIR Vue Pro R 640 and thermoMap cameras. *Remote Sensing* 11(3):330. https://doi.org/10.3390/rs11030330.

Song, M., H. Zhao, S. Zou, X. Cai and Y. Cheng. 2025. A UAV thermal imaging temperature correction method based on block adjustment with temperature control point constraints. *Remote Sensing Letters* 16(9):936–946. https://doi.org/10.1080/2150704X.2025.251355.

Szostak, R., M. Zimnoch, P. Wachniew and A. Jasek-Kamińska. 2023. Self-calibration of UAV thermal imagery using gradient descent algorithm. *Drones* 7(11):683. https://doi.org/10.3390/drones7110683.

Turner, D., A. Lucieer, Z. Malenovský, D. H. King and S. A. Robinson. 2014. Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over Antarctic moss beds. *Remote Sensing* 6(5):4003–4024. https://doi.org/10.3390/rs6054003.

Wang, Z., J. Zhou, J. Ma, Y. Wang, S. Liu, L. Ding, W. Tang, N. Pakezhamu and L. Meng. 2023. Removing temperature drift and temporal variation in thermal infrared images of a UAV uncooled thermal infrared imager. *ISPRS Journal of Photogrammetry and Remote Sensing* 203:392–411. https://doi.org/10.1016/j.isprsjprs.2023.08.011.

Yan, W. Y., A. Shaker, A. Habib and A. P. Kersting. 2012. Improving classification accuracy of airborne LiDAR intensity data by geometric calibration and radiometric correction. *ISPRS Journal of Photogrammetry and Remote Sensing* 67:35–44. https://doi.org/10.1016/j.isprsjprs.2011.10.005.

Yuan, W. and W. Hua. 2022. A case study of vignetting nonuniformity in UAV-based uncooled thermal cameras. *Drones* 6(12):394. https://doi.org/10.3390/drones6120394.

Zhang, S., N. Chu, S. Zheng, L. Li, C. Cai, A. Mohammad-Djafari and F. Zhao. 2025. Temperature calibration of surface emissivities with an improved thermal image enhancement network. *SSRN*. https://doi.org/10.2139/ssrn.5194761.

Zhang, Y. 1999. Modis UCSB emissivity library. Institute for Computational Earth System Science, University of California, Santa Barbara, 1999. <https://www.icess.ucsb.edu/modis/EMIS/html/em.html> (last date accessed: 9 January 2026).

Zhu, H., Y. Huang, Z. An, H. Zhang, Y. Han, Z. Zhao, F. Li, C. Zhang and C. Hou. 2024. Assessing radiometric calibration methods for multispectral UAV imagery and the influence of illumination, flight altitude and flight time on reflectance, vegetation index and inversion of winter wheat AGB and LAI. *Computers and Electronics in Agriculture* 219:108821. http://doi.org/10.1016/j.compag.2024.108821.