Harmonizing satellite thermal data with ground-based observations for climate long-term monitoring

Ewa Głowienka[1], Eva Savina Malinverni[2], Marsia Sanità[2], Krystyna Michałowska[1], Marcin Kucza[1]

[1]Dept. of Photogrammetry, Remote Sensing, and Spatial Engineering, Faculty of Geo-Data Science, Geodesy, and Environmental Engineering AGH University of Krakow, 30-059 Krakow, Poland - eglo@agh.edu.pl, michalowska@agh.edu.pl, mkucza@agh.edu.pl

[2]Dept. of Civil, Building and Architecture, Marche Polytechnic University, 60131 Ancona, Italy - e.s.malinverni@univpm.it, m.sanita@pm.univpm.it

Keywords: Landsat 8, ECOSTRESS, MODIS, Sentinel-3, Ground-based thermal sensors, Environmental monitoring, Climate change monitoring.

Abstract

Estimating land surface temperature (LST) in urban environments remains a complex task due to high surface heterogeneity and variations in material emissivity. This study compares LST products from four satellite sensors (Landsat 8 TIRS, ECOSTRESS, MODIS Terra, and Sentinel-3 SLSTR) with in-situ temperature measurements recorded by 27 ground-based sensors placed across diverse urban surface types in Kraków, Poland.

To minimize cross-platform discrepancies, a harmonization workflow was applied, including spatial reprojection, temporal alignment within a ±15-minute window, and emissivity normalization. Residuals were assessed using standard validation metrics: RMSE, MAE, bias, and Pearson's correlation coefficient. Among the examined datasets, Landsat 8 delivered the most consistent results (RMSE ≈ 7.7°C), while ECOSTRESS exhibited the highest positive bias and spatial variability. Residual errors were notably larger over vegetated areas, reflecting the effects of variable emissivity and canopy shading.

Despite the observed errors, the harmonized dataset enables consistent comparisons between sensors and an effective thermal assessment of urban areas. The results highlight the methodological relevance of harmonization in thermal remote sensing, particularly in heterogeneous urban areas.

1. Introduction

Land surface temperature (LST) is a fundamental variable in the study of urban climate, surface energy balance, and thermal vulnerability assessment. In the context of rapidly urbanizing regions and intensifying urban heat island (UHI) effects, accurate and consistent LST measurements are essential for monitoring anthropogenic thermal dynamics, supporting sustainable urban planning, and mitigating health-related risks associated with heat stress (Weng et al., 2014; Zhou et al., 2018).

Satellite-based thermal infrared (TIR) sensors provide spatially extensive and temporally consistent measurements of LST. Platforms such as Landsat 8 TIRS, ECOSTRESS, MODIS Terra, and Sentinel-3 SLSTR have been widely used in both scientific and operational applications (Li et al., 2013; Martin et al., 2019). However, these datasets vary in terms of spatial resolution (70 m to 1 km), temporal revisit cycles, acquisition time, and emissivity correction methods. These inconsistencies pose challenges for inter-platform comparison and hinder the integration of multi-sensor thermal data into coherent long-term climate records. Several studies have reported that urban heterogeneity, including surface materials, built-up geometry, and vegetation cover, significantly affect the accuracy of satellite-derived LST products. Niclòs et al. (2023) showed that the variability of urban emissivity and atmospheric conditions may cause deviations exceeding ±5°C in ECOSTRESS retrievals, even under clear-sky conditions. Similarly, Weng et al. (2014) emphasized that retrieval geometry and land cover class influence LST performance, particularly in densely constructed areas.

Ground-based temperature sensors provide critical reference data for validating LST products. However, global networks such as GSN, CRN, and FLUXNET are often spatially sparse in urban environments. Local sensor networks can capture the fine-scale thermal structure of cities and support the evaluation of satellite LST under realistic surface conditions (Zhou et al., 2018). Direct comparison between satellite and ground measurements remains problematic due to differences in scale, observation geometry, and timing (Martin et al., 2019). Harmonization techniques have emerged as a necessary methodological bridge between satellite platforms and in-situ validation datasets. These include spatial resampling, temporal synchronization, and surface emissivity normalization. Such processes are indispensable for ensuring the comparability and usability of multi-sensor LST datasets in modelling, downscaling, and climate monitoring applications (Li et al., 2013; Zha et al., 2024).

This study aims to evaluate the harmonization of LST products from multiple satellite platforms through comparison with a dense network of ground-based observations in Kraków, Poland. The analysis aims to quantify cross-platform residuals, assess the impact of local surface characteristics, and evaluate the readiness of harmonized LST data for use in further predictive modelling and urban heat risk analysis.

2. Study area

The survey was conducted in Krakow (population ~800,000), a central European city located in the southern part of Poland. Krakow is characterized by a heterogeneous urban fabric, including densely developed historic centre, modern residential zones, urban green spaces and industrial districts. The climate is classified as humid continental (Köppen classification - Dfb), characterized by warm summers and cold winters, and is increasingly subject to anthropogenically induced thermal anomalies (Peel et al., 2007). Recent observations indicate that Kraków experiences frequent summer heatwaves and elevated nighttime temperatures, contributing to local urban heat island (UHI) effects. Due to its compact urban morphology and pronounced land cover contrasts, Krakow was used as a representative case study for urban thermal heterogeneity and harmonization of satellite data with ground-based data.
The observation date selected for this study was August 19, 2023, chosen based on synoptic meteorological conditions indicating a cloudless sky scenario, negligible atmospheric perturbations (e.g., aerosol loading or convective cloud formation).

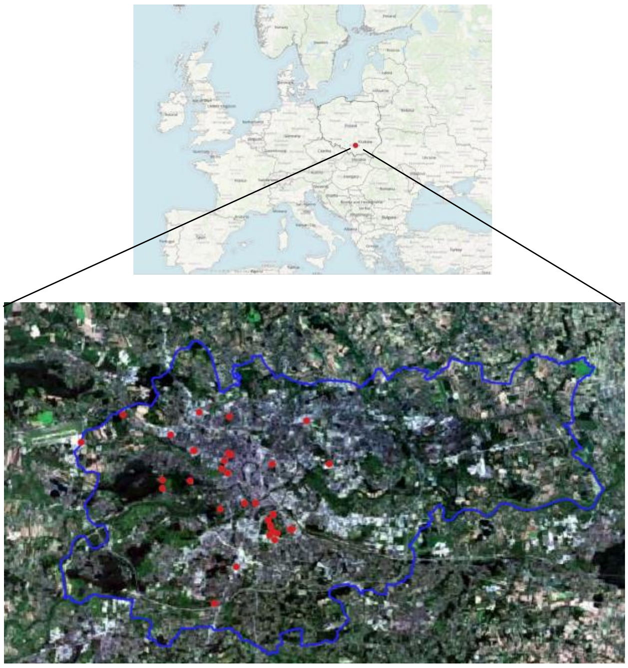

Figure 1. Location of the study area. High-resolution satellite imagery of Kraków showing the distribution of in-situ temperature sensors (red points) used for validation of satellite-derived LST data. The blue boundary denotes the extent of the urban study area.

3. Data and methodology

3.1 Satellite datasets

To enable multi-resolution and multi-temporal LST analysis, four satellite platforms were utilized: Landsat 8 (TIRS), MODIS Terra, Sentinel-3 (SLSTR), and ECOSTRESS. The selected satellite sensors offered complementary spatiotemporal characteristics, allowing for a cross-platform comparison of thermal retrieval performance. Key technical specifications, including spectral range, spatial resolution, acquisition time, product level, and atmospheric correction methods, are summarized in Table 1.

Landsat 8 Level-2 data were pre-processed using the LaSRC (Land Surface Reflectance Code) module, with land surface emissivity estimated via NDVI-based classification (Vermote et al., 2016). ECOSTRESS Level-2 products were obtained from NASA's LP DAAC and include emissivity corrections based on ASTER global surface emissivity datasets (Hulley& Freepartner, 2019; Meerdink et al., 2019). These products are designed for frequent thermal observations and are suitable for urban-scale monitoring. MODIS LST data were sourced from the MOD11A1 daily product and include pre-processed emissivity corrections (Pérez-Díaz et al., 2017). Sentinel-3 SLSTR data were accessed through the Copernicus Open Access Hub and processed using dual-view geometry calibration (Li et al., 2023).

Platform	Sensor	Spectral range (µm)	Spatial resolution	Temporal Frequency	Acquisition Time (Local)	Product Level	Atmospheric correction
Landsat 8	TIRS	10.60 – 11.19	100 m	16 days	9:32	L-2	Yes
ECOSTRESS (ISS)	ECO_L2_LST_URG	8.00–12.00	70 m	Variable (ISS orbit)	9:16	L-2	Yes
MODIS (Terra)	MOD11A1 v6	10.78 – 11.28	1km	Daily	10:30	L-3	Yes
Sentinel-3	SLSTR	10.85 – 12.00	1km	1–2 days	8:50	L-2	Yes

Table 1. Overview of satellite-based LST data used in the study.

3.2 In-Situ ground-based observations

To validate satellite-derived LST products, a network of 27 Thermochron DS1921G-F5 iButton sensors across urban Kraków were deployed. These sensors recorded surface-contact temperature at 10-minute intervals with a stated accuracy of ±1.0°C and resolution of 0.5°C. Sensor locations were selected to reflect the dominant urban surface types, classified into three simplified Local Climate Zone (LCZ) categories:

- vegetated surfaces (urban parks, lawns),
- impervious surfaces (asphalt roads, concrete sidewalks),
- mixed or transitional areas (vacant lots, gravel surfaces).

Each sensor was positioned directly on the surface, shielded from direct solar radiation and reflective interference. Data quality was verified through post-processing to remove outliers and detect measurement dropouts.

3.3 Harmonization workflow

The harmonization process aimed to align spatial, temporal, and physical discrepancies between satellite and in-situ measurements. A three-step framework was implemented.

In the first step all satellite LST raster datasets were reprojected to the UTM Zone 34N coordinate system (EPSG:32634), ensuring precise geospatial alignment. Pixel values corresponding to sensor locations were extracted using a nearest-neighbor technique for coarse-resolution products (MODIS, Sentinel-3), and bilinear interpolation for higher-resolution data (Landsat, ECOSTRESS).

In the next step, temporal alignment was achieved by matching satellite acquisition times with the closest in-situ readings. A ±15-minute tolerance window was applied to ensure consistency in atmospheric conditions. Finally, the residuals were computed as the difference between satellite-derived LST and ground measurements:

$$Residual = T_{sat} - T_{ground} \tag{1}$$

Statistical comparisons were performed using standard validation metrics. The agreement between satellite-derived and ground-based LST measurements was assessed using the following metrics (Pires et al., 2019):

- Root Mean Square Error (RMSE) – measures overall deviation,
- Mean Absolute Error (MAE) – indicates average magnitude of error,
- Pearson's correlation coefficient (R) – correlation strength between platform and in-situ data
- Systematic bias (ΔT) – reflects the mean signed difference between satellite and ground measurements.

To assess the influence of land surface type, residuals were grouped by LCZ category and analyzed via boxplots and mean RMSE per class. This allowed comparison of platform performance in response to surface emissivity variability and urban material heterogeneity.

4. Results and discussion

The spatial variability of retrieved LST across the study area is shown in Figure 2. Significant differences in spatial resolution and thermal contrast can be observed, particularly between Landsat/ECOSTRESS and MODIS/Sentinel-3 products. In-situ sensor locations, marked as black circles, were used to validate each raster product. The maps highlight intra-urban thermal variability, particularly between vegetated and impervious zones.

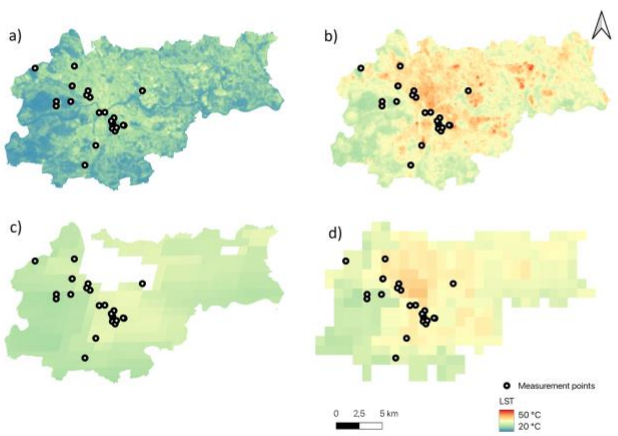

Figure 2. Land surface temperature (LST) retrieved from each satellite platform. a) Landsat 8 (100 m), b) ECOSTRESS (70 m), c) MODIS Terra (1 km), d) Sentinel-3 (1 km). Black dots indicate in-situ sensor locations (n = 27).

4.1 Validation of Satellite-Derived LST

The evaluation of satellite-based LST retrievals revealed considerable differences in accuracy across platforms. While some sensors performed within acceptable error thresholds, others displayed systematic biases and high spatial variability. These results reflect both sensor-specific characteristics and the challenges associated with measuring heterogeneous urban surfaces using orbital thermal data.

To assess the accuracy of satellite-derived land surface temperature (LST), residuals between in-situ sensor data and satellite estimates were computed for all 27 locations. Summary statistics for each platform, including RMSE, MAE, Bias, and correlation coefficient (R), are presented in Table 2.

Validation Metric	Landsat 8	ECOSTRESS	MODIS Terra	Sentinel-3
RMSE	7.72	11.05	7.79	9.71
MAE	6.04	9.61	6.80	8.33
BIAS	-2.04	7.58	2.77	6.20
R	0.17	0.10	0.18	0.06

Table 2. Summary of validation metrics for LST products across four satellite platforms.

Among the tested platforms, Landsat 8 exhibited the lowest overall error metrics, with RMSE = 7.72°C and MAE = 6.04°C. It also demonstrated a slight underestimation trend (bias = −2.04°C), consistent with findings reported in earlier literature (Li et al., 2013; Wan et al., 2004). In contrast, ECOSTRESS exhibited the highest residual error and a systematic warm bias of +7.58°C. While promising in terms of spatial resolution, the sensor exhibited high spatial inconsistency across locations. MODIS provided moderate agreement with in-situ data and a moderate positive bias. Sentinel-3 presented low correlation and scattered residuals, likely resulting from subpixel heterogeneity and coarse spatial resolution in the urban context.

These results confirm that spatial resolution alone does not guarantee accuracy in urban LST retrievals, particularly without local emissivity calibration or contextual correction.

4.2 Platform-specific residual distributions

To assess the distribution of residual errors, Figure 3 presents boxplots of LST residuals ($T_{sat} - T_{ground}$) for each platform. Landsat 8 residuals are strongly clustered around zero, indicating stable and consistent performance across the sensor network. ECOSTRESS and Sentinel-3 display broad distributions and frequent outliers, confirming their higher variability. These differences underscore the sensitivity of thermal retrievals to both sensor design and urban surface composition.

https://doi.org/10.5194/isprs-archives-XLVIII-M-7-2025-127-2025

Residuals were also analyzed in relation to surface cover types, grouped into simplified Local Climate Zone (LCZ) classes: Vegetated, Impervious, and Mixed. The results reveal that (Table 3, Figure 4.):

- Over vegetated surfaces, overestimation was more evident in ECOSTRESS and Sentinel-3, likely due to complex shading and variable canopy emissivity.
- Impervious surfaces exhibited relatively consistent performance, with Landsat 8 showing the lowest residuals across all LCZ types.
- Mixed zones presented the highest variability, as expected due to their heterogeneous thermal behavior. These observations suggest that satellite-derived LST accuracy is influenced not only by sensor specifications but also by the spatial context of the surface being observed.

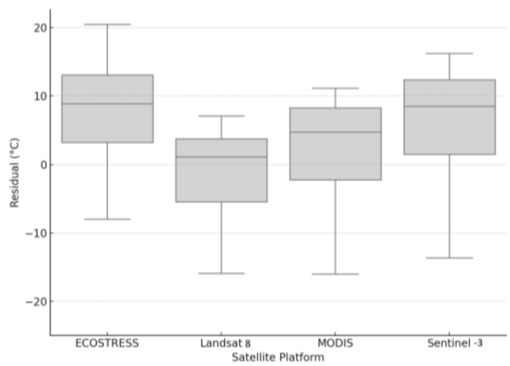

Figure 3. Distribution of residuals ($T_{sat} - T_{ground}$) for each satellite platform.

Figure 5 illustrates the correlation between ground temperature and satellite-derived LST residuals. A slight positive trend is observed for ECOSTRESS and Sentinel-3, indicating that higher ground temperatures are associated with larger overestimations, particularly under clear-sky conditions. Landsat 8 and MODIS show flatter trends, consistent with more stable thermal retrieval algorithms.

LCZ	Platform	RMSE (°C)	MAE (°C)	Bias (°C)
Impervious	ECOSTRESS	10.27	8.84	5.62
	Landsat 8	9.61	7.59	-3.71
	MODIS	8.93	7.92	2.10
	Sentinel-3	10.29	8.54	5.39
Mixed	ECOSTRESS	11.89	10.97	8.30
	Landsat 8	7.00	4.80	-1.03
	MODIS	7.47	6.75	2.28
	Sentinel-3	9.54	9.42	5.71
Vegetated	ECOSTRESS	11.26	9.58	8.96
	Landsat 8	5.96	5.32	-1.08
	MODIS	6.80	5.82	3.63
	Sentinel-3	9.25	7.56	7.20

Table 3. Residual distributions by LCZ type and platform.

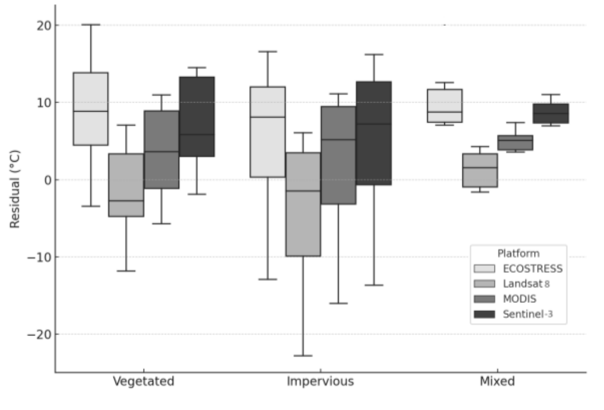

Figure 4. Boxplots of residuals grouped by LCZ and satellite platform.

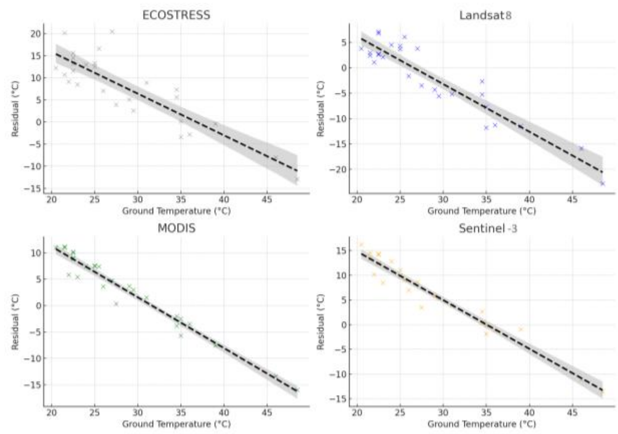

Figure 5. Scatterplots of residual vs ground temperature for each platform.

Despite the relatively small number of ground sensors, the results provide a robust indication of platform-specific error patterns. Landsat 8 emerges as the most reliable source for high-resolution urban LST mapping, while MODIS and Sentinel-3, due to their coarser resolution, are better suited for regional-scale applications unless further corrected. ECOSTRESS, while promising due to its spatial resolution and temporal flexibility, requires careful local calibration due to its high bias variability.

Future work should include multi-temporal analysis, sensor fusion, and integration of surface morphology data to improve LST accuracy in urban environments.

4.3 Harmonization insights and implications

While residuals may appear high in absolute terms, they are consistent with known inter-platform variability, particularly in complex urban environments. These discrepancies reflect the combined influence of spatial resolution, viewing geometry, emissivity assumptions, and the absence of localized atmospheric or emissivity corrections. Similar performance patterns have been reported in earlier LST validation studies, where even after preprocessing, systematic biases remained evident across platforms (Pires et al., 2019; Niclòs et al., 2023).

In this study, harmonization included reprojection to a shared coordinate system (UTM Zone 34N), resampling to match the spatial resolution of ground sensors, and temporal alignment using a ±15-minute window. These steps ensured that each satellite-derived LST product was directly comparable with in-situ data under similar atmospheric conditions and spatial referencing. Residuals were evaluated across simplified LCZ categories, revealing platform-specific biases that were most pronounced in vegetated and transitional areas—consistent with findings by (Weng, 2009), who emphasized the influence of surface heterogeneity on thermal retrieval accuracy.

Harmonization is not a mere preprocessing step, but rather the foundation for robust cross-platform comparison and integration of satellite thermal observations. It enables the identification of systematic retrieval biases, supports validation protocols, and prepares the data for more advanced applications, such as urban heat modeling or machine learning–based downscaling (Zha et al., 2024; Weng et al., 2004).

As urban thermal monitoring becomes increasingly reliant on multi-source satellite observations, standardized harmonization workflows will remain essential for ensuring comparability, reducing uncertainty, and enhancing the usability of LST data in climate-sensitive urban planning. This approach follows recommendations proposed by (Pires et al., 2019; Niclòs et al., 2023), emphasizing the need for temporal and spatial standardization prior to any inter-sensor analysis.

5. Conclusion

The results confirm that harmonization of multi-platform satellite LST data is achievable and methodologically beneficial for urban thermal applications. Four thermal datasets (Landsat 8, ECOSTRESS, MODIS Terra, and Sentinel-3 SLSTR) were evaluated using in-situ ground measurements across diverse urban surface types in Kraków, Poland. The harmonized dataset showed moderate to high RMSE values, ranging from 7.7 to 11.1 °C. Despite these discrepancies, it ensured consistent spatial alignment, temporal synchronization, and emissivity adjustment, allowing for meaningful and comparable cross-platform validation. Among the evaluated platforms, Landsat 8 showed the most stable thermal retrievals, whereas ECOSTRESS, despite its finer spatial resolution, exhibited the largest systematic overestimation. Observed differences across platforms reflect the underlying influence of spatial resolution, sensor geometry, and emissivity correction methods—factors previously identified as key determinants of LST retrieval accuracy (Weng et al., 2004; Niclòs et al. 2023). However, the limited number of ground-based sensors ($n = 27$) might slightly constrain the generalization of the results. Future studies should include broader sensor networks and multi-seasonal observations to enhance representativeness.

Despite its limited direct measurement accuracy, the dataset highlights the methodological importance of cross-sensor harmonization workflows. These results support future efforts in LST calibration, urban thermal modelling, and data-driven planning. In the context of rising heat risks, harmonized thermal products offer a foundation for real-time monitoring, heat alert systems, and adaptive urban climate strategies.

Acknowledgements

Research project supported by program „Excellence initiative – research university” for the AGH University”.

References

Martin, M. A., Ghent, D., Pires, A. C., Göttsche, F.-M., Cermak, J., & Remedios, J. J. (2019). Comprehensive In Situ Validation of Five Satellite Land Surface Temperature Data Sets over Multiple Stations and Years. *Remote Sensing*, *11*(5), 479. https://doi.org/10.3390/rs11050479

Guha, S., Govil, H., Dey, A., & Gill, N. (2018). Analytical study of land surface temperature with NDVI and NDBI using Landsat 8 OLI and TIRS data in Florence and Naples city, Italy. *European Journal of Remote Sensing, 51*, 387–399. https://doi.org/10.1080/22797254.2018.1474494

Hulley, G.C., & Freepartner, R. (2019). ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) Level 2 Product User Guide. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. https://lpdaac.usgs.gov/documents/423/ECO2_User_Guide_V1.pdf

Li, R., Li, H., Hu, T., Bian, Z., Liu, F., Cao, B. (2023). Land surface temperature retrieval from Sentinel-3A SLSTR data: Comparison among split-window, dual-window, three-channel, and dual-angle algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 61, https://doi.org/10.1109/TGRS.2023.3288584

Li, Z.-L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., ... & Zhang, X. (2009). A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. *Sensors,* *9*(5), 3801–3853. https://doi.org/10.3390/s90503801

Meerdink, S. K., Hook, S. J., Roberts, D. A., & Abbott, E. A. (2019). The ECOSTRESS spectral library version 1.0. *Remote Sensing of Environment, 230*, 111196. https://doi.org/10.1016/j.rse.2019.05.015

Niclòs, R., Perelló, M., Puchades, J., Coll, C., Valor, E. (2023). Evaluating Landsat-9 TIRS-2 calibrations and land surface temperature retrievals against ground measurements using multi-instrument spatial and temporal sampling along transects. *International Journal of Applied Earth Observation and Geoinformation, 125*, 103576. https://doi.org/10.1016/j.jag.2023.103576

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society, 108*(455), 1–24. https://doi.org/10.1002/qj.49710845502

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen–Geiger climate classification. *Hydrology and Earth System Sciences, 11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Pérez-Díaz, C. L., Lakhankar, T., Romanov, P., Muñoz, J., Khanbilvardi, R., & Yu, Y. (2017). Evaluation of MODIS land surface temperature with *in-situ* snow surface temperature from CREST-SAFE. *International Journal of Remote Sensing*, *38*(16), 4722–4740, https://doi.org/10.1080/01431161.2017.1331055

Pires, A. C., Göttsche, F. M., Cermak, J., Trigo, I. F., Martins, M. A., & Olesen, F. S. (2019). Comprehensive in situ validation of five satellite land surface temperature data sets over multiple stations and years. *Remote Sensing, 11*(5), 479. https://doi.org/10.3390/rs11050479

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment, 185*, 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Wan, Z., Zhang, Y., Zhang, Q., & Li, Z.-L. (2004). Quality assessment and validation of the MODIS global land surface temperature. *International Journal of Remote Sensing, 25*(1), 261–274. https://doi.org/10.1080/0143116031000116417

Weng, Q. (2009). Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4), 335–344. https://doi.org/10.1016/j.isprsjprs.2009.03.007

Weng, Q., Lu, D., & Schubring, J. (2004). Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment, 89*(4), 467–483. https://doi.org/10.1016/j.rse.2003.11.005

Zha, F., Lu, L., Wang, R., Zhang, S., Cao, S., Baqa, M. F., Li, Q., & Chen, F. (2024). Understanding fine-scale heat health risks and the role of green infrastructure based on remote sensing and socioeconomic data in the megacity of Beijing, China. *Ecological Indicators*. https://doi.org/10.1016/j.ecolind.2024.111847

Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Frolking, S., Yao, R., Qiao, Z., & Sobrino, J. A. (2019). Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives. *Remote Sensing*, *11*(1), 48. https://doi.org/10.3390/rs11010048

Zhou, D., Zhao, S., Liu, S., Zhang, L., & Zhu, C. (2014). Surface urban heat island in China's 32 major cities: Spatial patterns and drivers. *Remote Sensing of Environment, 152*, 51–61. https://doi.org/10.1016/j.rse.2014.05.017