University of Szeged

Doctoral School of Geosciences

Department of Atmospheric and Geospatial Data Sciences

Multi-Scale Model Development for Estimating Urban Air Temperature Patterns Using Land Surface Temperature and Auxiliary Data

Summary of PhD Dissertation

Guo Yuchen

Supervisors:

Dr. János Unger professor emeritus

Dr. Tamás Gál head of department associate professor

Department of Atmospheric and Geospatial Data Sciences

Szeged

2025

1. Introduction

As urbanization increases and global climate change becomes more pronounced, the impact of the urban thermal environment on the lives of residents, urban planning, and public health becomes more significant. In the urban climate research field, near-surface air temperature (Tair) is one of the most important variables. Therefore, studying and accurately estimating urban Tair not only helps to improve the understanding of the urban thermal environment but also provides support to policymakers for more effective urban planning and climate adaptation strategies.

Traditionally, urban Tair is measured mainly by in-situ meteorological observation stations. Although these stations can provide highly accurate data, the high costs of their construction and maintenance have resulted in sparse distribution in many cities, making it difficult to reflect the complex temperature variations within urban areas. Meanwhile, the development of mobile observation (Unger et al. 2010) and crowdsourced data (Venter et al. 2021) has brought new opportunities for urban Tair data collection, but it also faces the problems of complicated operation as well as insufficient data quality and spatial representativeness.

Utilizing satellite-based land surface temperature (LST) data to estimate urban Tair is one of the common methods to obtain high spatial resolution Tair data (Wang et al. 2023). However, the method also faces many challenges. Although there is a proven correlation between LST and Tair, this correlation has complex temporal and spatial variability characteristics (Guo et al. 2024b). For example, during the daytime, the correlation between LST and Tair tends to be weak due to solar radiation. This leads to difficulties in fully reflecting the temporal and spatial variability of Tair using only LST to estimate Tair.

Using auxiliary variables can enhance the accuracy of the urban Tair estimation. However, current studies often rely too heavily on traditional surface cover data (e.g., NDVI and building height), while incorporating meteorological variables remains rare (Guo et al. 2024a). In addition, Local Climate Zone (LCZ) data, which is closely linked to the urban thermal environment, has developed rapidly in recent years and has great potential to be applied in various climate models (Demuzere et al. 2022). However, it also remains seldom used in urban Tair estimation studies.

Furthermore, existing methods for assessing model accuracy are insufficient, and a systematic analytical framework is lacking to thoroughly investigate the sources of model errors and the factors influencing accuracy.

This fact has even resulted in conflicting conclusions in some studies regarding the factors influencing model accuracy and hinders further optimization of Tair estimation models in future research (Guo et al. 2024a).

In summary, current urban Tair estimation research faces numerous challenges, highlighting the urgent need for a high-resolution model that captures both temporal and spatial variations of urban Tair.

2. Research aims

The primary objective of this dissertation is to develop an urban Tair estimation model with high spatial and temporal resolution. During this process, this study aims to address several key challenges in this field by focusing on the following three questions:

What is the temporal and spatial relationship between Tair and LST? LST is the primary predictor variable in LST-based Tair estimation. Satellite-derived LST has been widely used in urban Tair studies due to its demonstrated correlation with near-surface Tair, but the complex spatiotemporal correlation between the two temperatures remains insufficiently explored. A deeper understanding of this relationship–particularly its temporal variability–is crucial for improving model accuracy. However, many existing studies assume a direct correlation between LST and Tair, often overlooking the need for in-depth analysis. This simplification can limit insights during model evaluation and hinder further improvements in estimation accuracy.

How to select appropriate predictor variables for Tair estimation? In addition to LST, other predictor variables play a crucial role in Tair estimation. Although often referred to as auxiliary variables, their importance is not necessarily secondary to LST. Carefully selected predictors can significantly enhance model performance and capture aspects of Tair variability that LST alone cannot. However, existing studies rely too heavily on land cover-related parameters while underutilizing meteorological data. The emergence of new datasets in recent years provides an opportunity to expand the selection of predictor variables, potentially improving the accuracy and robustness of Tair estimation models.

What factors influence model performance? Evaluating model performance is essential, yet current assessment methods vary widely. Some studies rely on quantitative metrics, while others provide only qualitative analyses, making it difficult to identify sources of error and influencing factors systematically. Additionally, conflicting conclusions across studies further complicate the understanding of model accuracy. A more systematic approach is needed to analyze the key factors affecting model performance, which will help refine estimation models and enhance their reliability in future research.

To address these challenges, this dissertation aims to develop a new urban Tair estimation model that integrates satellite-based LST with various auxiliary variables. To ensure accuracy and reliability, the study thoroughly examines the spatiotemporal relationship between Tair and LST. Additionally, it systematically explores key predictor variables, investigates factors influencing model performance, and identifies potential influencing factors on estimation errors. By optimizing model inputs and refining error analysis, this research delivers high-resolution Tair estimations, provides a valuable resource for urban climate studies, and contributes to more precise assessments of urban thermal environments.

3. Data and methods

Tair, as a predicted variable, originates from two sources. In the study conducted in Szeged, the data were obtained from the local urban meteorological station network. Specifically, Tair data were collected from a meteorological monitoring network established as part of the European Union project URBAN-PATH (2019). In China, Tair data were sourced from the Integrated Surface Database (ISD), provided by the National Centers for Environmental Information (NCEI) under the National Oceanic and Atmospheric Administration (NOAA) (Gawuc and Struzewska 2016).

This study utilized two types of satellite data to obtain Land Surface Temperature (LST): Landsat 8 and MODIS. However, Landsat 8 data were used only for comparing LST with Tair and were not included in model development. All MODIS data were obtained from the Google Earth Engine (GEE) database (Kumar and Mutanga 2018). The MODIS data used in this study were selected under clear-sky conditions, and low-quality pixels were removed using the quality control bands (QC_Day and QC_Night) provided in the MODIS dataset.

The auxiliary variables used in this study were categorized into two groups: time-related variables and space-related variables. The former includes atmospheric indices (10 atmospheric variables) and time cyclical variables (day of the year and hour of the day). The latter includes geographic data (longitude, latitude, elevation, and distance to the coastline) and land cover data (LCZ and LCZ-based largest patch index).

In the Szeged study, fewer auxiliary variables were used, including six ERA5-based atmospheric variables and LCZ. All used atmospheric indices were derived from the ERA5 reanalysis dataset (Hersbach et al. 2023). The elevation data were obtained from the Shuttle Radar Topography Mission (SRTM) digital elevation dataset (Jarvis et al. 2008). The LCZ data in Szeged were sourced from Bechtel et al. (2015), while the LCZ data for China were obtained from Demuzere et al. (2022) and are available in the GEE database. Data processing in this study was conducted using GEE, QGIS, and R.

For model development, the Random Forest (RF) regression algorithm was used, as it has been demonstrated to achieve high accuracy in estimating urban Tair (Wang et al. 2023). A tenfold cross-validation approach was applied to assess the performance of the estimation models. During RF model training, two variable importance measures (VIMs) were employed, the percentage increase in mean square error (%incMSE) and impurity-corrected VIM, to evaluate the importance of predictor variables in the model.

To quantify the accuracy of the estimation models, this study used two commonly applied statistical metrics: Root Mean Square Error (RMSE) and the coefficient of determination (R^2). Additionally, Pearson's correlation coefficient (r) (Lee Rodgers and Nicewander 1988; Schober and Schwarte 2018) was used to quantify the impact of predictor variables on model errors.

The models were trained at different spatial scales (urban and national) and temporal scales (diurnal and seasonal). The LST-Tair relationship and the factors influencing model performance were analyzed across multiple temporal scales (e.g., hourly, diurnal, seasonal, and annual). Notably, model training was conducted separately for warm and cold seasons in China based on given weather conditions. Both model training and statistical analyses were carried out using R statistical software. The "randomForest" and "ranger" packages were used for RF training.

4. Key findings

This detection systematically explored the estimation of urban near-surface Tair by integrating satellite-derived LST with multi-scale spatiotemporal auxiliary variables. Through a comprehensive analysis across different temporal and spatial scales, this study developed and validated a high-resolution (1 km, sub-daily) Tair estimation model and examined its applicability under diverse environmental conditions. The major findings and contributions of this study can be summarized as follows:

Thesis 1: The spatiotemporal variation characteristics and mechanistic linkages between LST and Tair were systematically compared with the attribution analysis of dominant drivers.

This study systematically analyzed the spatiotemporal variations and mechanistic linkages between LST and Tair by comparing their diurnal, seasonal, and spatial differences and identifying dominant driving factors. Two complementary case studies were conducted: one focused on daily variations in Szeged, while the other examined annual and latitudinal patterns across China. In Szeged, we compared LST and Tair at different times of the day and seasons, revealing distinct diurnal and seasonal variation patterns. LST was higher than Tair during the daytime due to solar heating but lower at night due to radiative cooling and anthropogenic heat influences. Spatially, the LST-Tair relationship varied across LCZs, with built-up areas showing greater differences than vegetated areas. Additionally, cross-platform discrepancies (MODIS and Landsat 8) in daytime temperature values highlighted challenges in integrating multi-source satellite-based LST data. At the national scale, the study revealed annual cycles of LST and Tair, with a two-peak, two-trough ΔT pattern influenced by seasonal variations in radiation, vegetation activity, and urban surface properties. Urban-rural differences in the LST-Tair relationship also exhibited seasonal variability. These findings enhance our understanding of urban thermal dynamics, providing valuable insights for urban climate studies. Moreover, the comparative analysis of LST and Tair serves as a critical foundation for developing accurate Tair estimation models in the subsequent sections of this dissertation.

Thesis 2: A multi-scale Tair estimation model was developed by integrating satellite-derived LST with time-related (e.g., ERA5 atmospheric profiles) and space-related (e.g., LCZ) variables, achieving high spatiotemporal resolution (1 km, sub-daily).

The model used LST as a primary predictor and incorporated time-related variables (e.g., ERA5 atmospheric profiles) and space-related variables (LCZ classifications) to enhance estimation accuracy. A comparative analysis of four model schemes demonstrated that integrating time-related data significantly improved performance across all seasons. The inclusion of space-related data further refined the model, though its overall contribution was smaller than that of meteorological variables. The final proposed model with both time-related and space-related variables outperformed all other schemes, exhibiting the lowest RMSE and highest R^2 values in both daytime and nighttime conditions. Seasonal evaluations were conducted, indicating the highest accuracy in summer and lower performance in transitional seasons (e.g., spring and autumn). The results demonstrate the model's reliability for estimating Tair with high spatiotemporal resolution (1 km, sub-daily). Overall, this study offers new insights into the selection of predictor variables in the field of urban Tair modeling, providing a valuable tool for urban climate research.

Thesis 3: This study demonstrated the role of LCZ data in urban Tair estimation and analyzed the temporal and spatial patterns of variability in its importance.

This research assessed the significance of LCZ data in Tair estimation by analyzing its impact across different seasons and times of the day. The findings revealed that LCZ data improved model accuracy, particularly at night, when surface characteristics exhibit a stronger influence on near-surface Tair. The importance of LCZ varies by season, with its highest impact observed during summer nights. A detailed importance analysis showed that while space-related variables had a smaller overall impact than time-related variables, they play a crucial role in capturing spatial heterogeneity in urban thermal environments. When evaluated individually in an urban scale in Szeged, LCZ outperformed any single meteorological variable at night, which highlighted its value in improving model accuracy. These results emphasize the importance of incorporating LCZ classifications in urban Tair estimation.

Thesis 4: Revealed the potential of meteorological data in urban Tair studies, which has been neglected, and this finding provides guidance for future Tair estimation studies.

This study highlights the critical role of meteorological variables in improving urban Tair estimation, addressing a gap in previous research that primarily relied on surface-related factors such as NDVI and other land cover parameters. The results demonstrate that time-related variables, particularly those derived from the ERA5 reanalysis dataset, significantly enhance model performance, especially during daytime when atmospheric activity is stronger. Specific humidity (SH) and temperature gradient (TG) are the most influential among these variables. The study also reveals a seasonal dependency in the contribution of meteorological data, which is that time-related variables have the greatest impact in summer when atmospheric mixing is more active. These findings suggest that incorporating meteorological data can substantially enhance urban Tair estimation models, making them more adaptable to different weather conditions. Therefore, this research provides a framework for integrating atmospheric parameters in future Tair modeling efforts, improving the accuracy and applicability of urban climate assessments.

Thesis 5: The applicability of the Tair estimation model was examined under varying temporal, weather, and climatic conditions by generating the Tair maps in different cities.

The applicability of the Tair estimation model was examined across various temporal, weather, and climatic conditions by generating Tair maps for multiple cities, including Szeged, Beijing and the Pearl River Delta (PRD) regions. In Szeged, the proposed model consistently achieved outstanding performance under all seasonal and diurnal conditions. Model validation indicated robust performance, particularly during summer nights coinciding with peak UHI effects. Spatial analysis confirmed Model 4's effectiveness in capturing the UHI with minimal errors, particularly under summer conditions, which validated the model's capability to capture precise urban thermal patterns. In China, the final model also demonstrated strong performance, achieving the lowest errors under warm night and cold day conditions. The model application in PRD and Beijing proved that the proposed model could accurately map UHI dynamics and spatiotemporal variations in Tair and effectively reflect local climatic effects such as sea breezes and nocturnal heat retention. These regional validations indicated the model's adaptability and precision in diverse urban contexts, highlighting its utility as a tool for delineating urban thermal environments and influencing urban climate studies and policy decisions.

Thesis 6: A systematic investigation of the spatiotemporal factors influencing the model's errors was conducted. The results can revise existing perspectives on Tair estimation.

A comprehensive analysis was conducted to investigate the spatiotemporal factors influencing the accuracy of a Tair estimation model. The study in Szeged quantitatively assessed the impact of temporal and spatial variables on model performance. The analysis of temporal variables provided insights into how atmospheric conditions affect model accuracy and revealed that the model performs better under more active atmosphere conditions, such as during the daytime or in summer. The analysis of spatial variables demonstrated that incorporating LCZ effectively differentiates the distinct microclimatic characteristics within urban areas. The study in China indicated that station density had a limited impact on the spatial variability of model performance. Instead, meteorological variables were identified as a key factor contributing to these spatial differences. Furthermore, the methods used for model evaluation, such as cross-validation strategies, also play a crucial role in determining final model performance. Therefore, when comparing different studies, it is essential to account for the influence of validation methods to ensure a fair and accurate assessment of model effectiveness.

5. Scientific significance

Estimating near-surface Tair is essential for urban climate research, environmental monitoring, and public health assessments. However, existing studies face challenges related to spatial and temporal resolution, data accessibility, and model applicability. Traditional approaches often rely on low-resolution meteorological station data or remote sensing products, which fail to capture the fine-scale spatial and temporal variations in urban temperature. Additionally, many models lack a systematic analysis of the key factors influencing Tair estimation accuracy, limiting their applicability and transferability.

To address these limitations, this dissertation developed a high-resolution and open-source-based Tair estimation model, which has both acceptable accuracy and broader applicability. This proposed model provides a valuable tool for studying urban heat, which will support climate adaptation strategies in various urban regions and offer a valuable approach to sub-daily Tair estimation using universally accessible data sources. Additionally, from the theoretical perspective, this study analyzed the impact of LST, atmospheric factors, and surface characteristics on urban Tair during model training. This helps clarify the mechanisms driving urban Tair variations and improves our understanding of urban climate processes. The findings also provide a theoretical basis for refining temperature estimation models and optimizing their application in different urban environments.

List of publishions

I. Publications used in the dissertation

MTMT ID:**10069598**

<u>Guo, Y.</u>, Gál, T., Tian, G., Li, H., & Unger, J. (2020). Model development for the estimation of urban air temperature based on surface temperature and NDVI–a case study in Szeged. Acta Climatologica, 54, 29-40. **DOI:**10.14232/acta.clim.2020.54.3

<u>Guo, Y.</u>, Unger, J., Khabibolla, A., Tian, G., He, R., Li, H., & Gál, T. (2024). Modeling urban air temperature using satellite-derived surface temperature, meteorological data, and local climate zone pattern—a case study in Szeged, Hungary. Theoretical and Applied Climatology, 155(5), 3841-3859. DOI:10.1007/s00704-024-04852-7 **IF: 2.8**

<u>Guo, Y.</u>, Unger, J., & Gál, T. (2024). Model development for estimating sub-daily urban air temperature patterns in China using land surface temperature and auxiliary data from 2013 to 2023. Remote Sensing, 16(24), 4675. **DOI:**10.3390/rs16244675 **IF: 4.2**

II. **Other publications**

<u>Guo, Y.</u> (2019). The influence of urban scale spatial morphology on thermal environment ind Zhengzou, China–A case study. Meteorological Notes of Universities 31, 23-31. **DOI:**10.31852/EMF.31.2019.023.031

III. **Conference Abstracts**

<u>Guo, Y.</u>, Unger, J., & Gál, T. (2020) Evaluation of the connection between urban surface and air temperature – case study in szeged. Meteorológiai Tudományos Nap 2020. november 19. Tudomány és tradíció a meteorológiában, **DOI:** 10.21404/46.MTN.2020

<u>Guo, Y.</u>, Unger, J., & Gál, T. (2022). Model development for the estimation of seasonal urban air temperature patterns using MODIS satellite data, a case study in Szeged (No. EMS2022-567). Copernicus Meetings. **DOI:**10.5194/ems2022-567

Co-authors' Declaration of used Publication 1:

We, the co-author of the publication: "**Guo, Y.**, Gál, T., Tian, G., Li, H., & Unger, J. (2020). *Model development for the estimation of urban air temperature based on surface temperature and NDVI–a case study in Szeged*. Acta Climatologica, 54, 29-40. https://doi.org/10.14232/acta.clim.2020.54.3" officially declare that the jointly published results in the thesis and the publication are greatly contributed by the candidate and was not or will not be used in the past or in the future, respectively, for the purpose of acquiring an academic degree or title.

Date: Szeged (Hungary), 28th March, 2025

Signature of the co-authors:

Gál Tamás:
(University of Szeged, Szeged, Hungary)

Tian Guohang: Tian Guohang
(Henan Agricultural University, Zhangzhou, China)

Li Huawei:
(Henan Agricultural University, Zhangzhou, China)

Unger János:
(University of Szeged, Szeged, Hungary)

Co-authors' Declaration of used Publication 2:

We, the co-author of the publication: "**Guo, Y.**, Unger, J., Khabibolla, A., Tian, G., He, R., Li, H., & Gál, T. (2024). Modeling urban air temperature using satellite-derived surface temperature, meteorological data, and local climate zone pattern—a case study in Szeged, Hungary. Theoretical and Applied Climatology, 155(5), 3841-3859. https://doi.org/10.1007/s00704-024-04852-7" officially declare that the jointly published results in the thesis and the publication are greatly contributed by the candidate and was not or will not be used in the past or in the future, respectively, for the purpose of acquiring an academic degree or title.

Date: Szeged (Hungary), 28th March, 2025

Signature of the co-authors:

Unger János:
(University of Szeged, Szeged, Hungary)

Khabibolla Almaskhan:
(University of Szeged, Szeged, Hungary)

Tian Guohang:
(Henan Agricultural University, Zhangzhou, China)

He Ruizhen:
(Henan Agricultural University, Zhangzhou, China)

Li Huawei:
(Henan Agricultural University, Zhangzhou, China)

Gál Tamás:
(University of Szeged, Szeged, Hungary)

Co-authors' Declaration of used Publication 3:

We, the co-author of the publication: "**<u>Guo, Y.</u>**, Unger, J., & Gál, T. (2024). Model development for estimating sub-daily urban air temperature patterns in China using land surface temperature and auxiliary data from 2013 to 2023. Remote Sensing, 16(24), 4675. https://doi.org/10.3390/rs16244675" officially declare that the jointly published results in the thesis and the publication are greatly contributed by the candidate and was not or will not be used in the past or in the future, respectively, for the purpose of acquiring an academic degree or title.

Date: Szeged (Hungary), 28th March, 2025

Signature of the co-authors:

Unger János:
(University of Szeged, Szeged, Hungary)

Gál Tamás:
(University of Szeged, Szeged, Hungary)

Declaration of Supervisors

I, **Unger János,** hereby confirm that the content of the dissertation is based on the independent work of the doctoral candidate (**GuoYuchen**) and that he has contributed decisively to the results through his independent creative activity. I consider the entire dissertation to be eligible for support from a professional and academic point of view and recommend its acceptance.

Dr. Unger János:
(Department of Atmospheric and Geospatial Data Sciences, University of Szeged, Szeged, Hungary)
Szeged, 28 March, 2025

I, **Gál Tamás**, hereby confirm that the content of the dissertation is based on the independent work of the doctoral candidate (**GuoYuchen**) and that he has contributed decisively to the results through his independent creative activity. I consider the entire dissertation to be eligible for support from a professional and academic point of view and recommend its acceptance.

Dr. Gál Tamás:
(Department of Atmospheric and Geospatial Data Sciences, University of Szeged, Szeged, Hungary)
Szeged, 28 March, 2025

Reference

Bechtel B, Alexander PJ, Böhner J, et al. (2015) Mapping local climate zones for a worldwide database of the form and function of cities. ISPRS Int J Geoinf 4:199–219. https://doi.org/10.3390/ijgi4010199

Demuzere M, Kittner J, Martilli A, et al. (2022) A global map of local climate zones to support earth system modelling and urban-scale environmental science. Earth Syst Sci Data 14:3835–3873. https://doi.org/10.5194/essd-14-3835-2022

Gawuc L, Struzewska J (2016) Impact of modis quality control on temporally aggregated urban surface temperature and long-term surface urban heat island intensity. Remote Sens (Basel) 8:374. https://doi.org/10.3390/RS8050374

Guo Y, Unger J, Gál T (2024a) Model development for estimating sub-daily urban air temperature patterns in China using land surface temperature and auxiliary data from 2013 to 2023. Remote Sens 16(24), 4675. https://doi.org/10.3390/rs16244675

Guo Y, Unger J, Khabibolla A, et al. (2024b) Modeling urban air temperature using satellite-derived surface temperature, meteorological data, and local climate zone pattern—a case study in Szeged, Hungary. Theor Appl Climatol 155:3841–3859. https://doi.org/10.1007/s00704-024-04852-7

Hersbach H, Bell B, Berrisford P, et al. (2023) ERA5 hourly data on pressure levels from 1940 to present [Dataset]. In: Copernicus Climate Change Service (C3S) Climate Data Store (CDS).

Jarvis, A., Guevara, E., Reuter, H. I., & Nelson, A. D. (2008). Hole-filled SRTM for the globe : version 4: data grid. Web publication/site, CGIAR Consortium for Spatial Information. http://srtm.csi.cgiar.org/ Lee Rodgers J, Nicewander WA (1988) Thirteen Ways to Look at the Correlation Coefficient. Am Stat 42:59–66

Kumar L, Mutanga O (2018) Google Earth Engine applications since inception: usage, trends, and potential. Remote Sens 10(10):1509. https://doi.org/10.3390/rs110 50591

Lee Rodgers J, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. Am Stat 42:59–66

Schober P, Schwarte LA (2018) Correlation coefficients: Appropriate use and interpretation. Anesth Analg 126:1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Unger J, Gál T, Rakonczai J, et al. (2010) Modeling of the urban heat island pattern based on the relationship between surface and air temperatures. Időjárás 114:287–302

URBAN-PATH Project (2019) Evaluations and public display of urban patterns of human thermal conditions. Available at: http:// urban-path. hu/, Accessed date: 20 May 2019

Venter ZS, Chakraborty T, Lee X (2021) Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms. Sci Adv 7:1–10. https://doi.org/10.1126/sciadv.abb9569

Wang H, Yang J, Chen G, et al. (2023) Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011–2022. Urban Clim 49:101499. https://doi.org/10.1016/j.uclim.2023.101499