Modeling thermal comfort indicators in an urban environment using remote sensing, open geodata, and artificial intelligence

vorgelegt von

M.Sc.

Stenka Valentinova Vulova

ORCID: 0000-0001-9083-1163

an der Fakultät VI – Planen Bauen Umwelt –
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften
- Dr.rer.nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr.-Ing. Reinhard Hinkelmann
Gutachter: Prof. Dr. Patrick Hostert
Gutachterin: Prof. Dr. Birgit Kleinschmit

Tag der wissenschaftlichen Aussprache: 11.10.2022

Berlin 2022

Acknowledgements

Summary

The world's expanding cities are increasingly at risk from the effects of climate and global change. Heat extremes pose one of the greatest threats to urban residents, causing mortalities, detrimental health effects, economic losses, increased energy consumption, and exacerbated air pollution. Various frameworks, such as the Sponge City, have been proposed to increase the resilience of cities to multiple environmental threats, mainly relying on the ecosystem services provided by urban greenery. Sustainability initiatives in cities need to be implemented at a neighborhood scale capturing socio-demographic vulnerability to heat stress. However, high-resolution, spatially distributed data to identify hotspots of heat exposure are largely unavailable. In particular, two major indicators of thermal comfort in cities - air temperature and evapotranspiration (ET) - urgently need to be characterized. However, their estimation remains challenging, as the urban environment is highly horizontally and vertically heterogeneous. The majority of existing process-based models cannot handle the urban complexity or be upscaled to the city-level. To address these challenges, state-of-the-art technologies, in particular, artificial intelligence (AI), remote sensing, and crowdsourcing, have emerged as a solution to producing spatially-explicit and accurate information on parameters relevant to climate change mitigation. Thus, the overall goal of this thesis was to characterize thermal comfort indicators in an urban environment at a high spatial resolution using remote sensing, open geodata, and AI. Berlin, Germany was selected as a test site due to its high risk of heat-related mortalities and established infrastructure of atmospheric measurements. This thesis had two main objectives: (1) investigate how accurately thermal comfort indicators in urban areas can be modeled at a high spatial resolution with remote sensing imagery, geodata, and meteorological data using machine learning (ML) regression algorithms; and (2) assess the spatial variation of the modeled ET and air temperature across Berlin to provide insights for urban planning and policy focused on heat risk mitigation.

Air temperature, especially at night, is one of the best indicators of heat stress. However, high-resolution air temperature maps have until recently remained unavailable due to the limited number of weather stations in cities. Thus, the first part of this thesis focused on how open source remote sensing, GIS and crowdsourced weather data can be used to predict the spatial distribution of nocturnal air temperature one day in advance at a 30-m resolution for the entire city using ML. Three predictor scenarios were tested: (1) only remotely sensed predictors, (2) only the spatially interpolated crowdsourced air temperature from the previous day, and (3) both remotely sensed and crowdsourced data from the previous day. As crowdsourced weather data is associated with more uncertainties than traditional weather data, validation was conducted with both crowdsourced and traditional air temperature data using "leave-one-date-out" cross-validation. Using only remotely sensed predictors showed the highest overall accuracy when validating with conventional air temperature data, demonstrating that remote sensing data is essential to

accurately predicting air temperature in the gaps between citizen weather stations. Accuracy was high when validating with crowdsourced data, but decreased when validating with independent data, likely due to the effect of citizen weather stations' installation near building walls. However, accuracy was highest in highly sealed urban fabric, where heat risk is highest, demonstrating the utility of this approach for identifying dangerous hotspots within the city.

ET provides an essential cooling service and approximates vegetation water use, making it a key indicator for urban sustainability. ET is defined as the water transported from the land surface to the atmosphere and consists mainly of soil evaporation, plant transpiration, and evaporation of intercepted precipitation in terrestrial ecosystems. To improve ET modeling in urban areas, a novel approach was developed which incorporated diverse land cover contributing to urban ET measurements into AI algorithms. Flux footprint modeling, which estimates the source area of eddy covariance (EC) measurements, was used to extract weighted averages from remote sensing and GIS data. Incorporating remotely sensed data extracted with flux footprints enhanced the predictive accuracy of models. In the next phase, this approach was extended to map ET for the entire city at a high spatial (10-m) and temporal (hourly) resolution for one year. Validation showed that this approach is reliable for mapping urban ET. Lastly, to support urban planning, the spatial variation of both modeled thermal comfort indicators was analyzed in relation to land cover. Ultimately, using remote sensing and AI to characterize indicators of thermal comfort can support urban planning aiming to reduce heat risks for an increasingly urban world population.

ZUSAMMENFASSUNG

Wachsende Städte stehen weltweit zunehmend den Auswirkungen des globalen Wandels und Klimaveränderungen gegenüber. Extreme Hitzeperioden stellen eine der größten Bedrohungen für Stadtbewohner dar. Sie führen zu Todesfällen, gesundheitlichen Beeinträchtigungen, wirtschaftlichen Verlusten, erhöhtem Energieverbrauch und verstärkter Luftverschmutzung. Verschiedene städtebauliche Nachhaltigkeitskonzepte, wie z. B. die Sponge City, stützen sich vor allem auf die Ökosystemleistungen des städtischen Grüns und sollen die Resilienz von Städten gegenüber zahlreichen Umweltbedrohungen erhöhen. Nachhaltigkeitsinitiativen müssen auch auf der Ebene von Stadtteilen umgesetzt werden, um soziodemografische Anfälligkeiten für Hitzestress detailliert aufzugreifen. Entsprechend hochauflösende, räumlich verteilte Daten für die Identifizierung von Hotspots mit Hitzerisiko fehlen jedoch weitgehend. Insbesondere zwei wichtige Indikatoren für das thermische Wohlbefinden in Städten – Lufttemperatur und Evapotranspiration (ET) – müssen dringend charakterisiert werden. Ihre Abschätzung bleibt eine Herausforderung, da die städtische Umwelt horizontal und vertikal sehr heterogen ist. Die meisten der bestehenden prozessbasierten Modelle können die Komplexität von Städten nicht aufnehmen oder müssen auf die Stadtebene hochskaliert werden. Zur Bewältigung dieser Herausforderungen haben sich modernste Technologien, insbesondere künstliche Intelligenz (KI), Fernerkundung und Crowdsourcing, als Lösung für die Erstellung räumlich expliziter und genauer Informationen über die für den Klimaschutz relevanten Parameter erwiesen. Das übergeordnete Ziel dieser Arbeit war es daher, Indikatoren für das thermische Wohlbefinden in einer städtischen Umgebung mit hoher räumlicher Auflösung und unter Verwendung von Fernerkundung, offenen Geodaten und KI zu charakterisieren. Berlin wurde aufgrund des hohen Risikos von hitzebedingten Todesfällen und der etablierten Infrastruktur für atmosphärische Messungen als Studiengebiet ausgewählt. Die vorliegende Doktorarbeit verfolgt zwei Hauptziele: (1) Methoden zu untersuchen, wie Indikatoren zum thermischen Wohlbefinden in städtischen Gebieten mit Fernerkundung, Geodaten und meteorologischen Daten unter Verwendung von Regressionsalgorithmen des maschinellen Lernens (ML) in hoher räumlicher Auflösung modelliert werden können; und (2) die räumliche Variation der modellierten ET und Lufttemperatur in Berlin zu bewerten, um Erkenntnisse in die Stadtplanung und -politik zur Minderung des Hitzerisikos einfließen zu lassen.

Die Lufttemperatur, insbesondere nachts, ist einer der besten Indikatoren für Hitzestress. Hochauflösende Karten der Lufttemperatur waren jedoch bislang aufgrund der begrenzten Anzahl von Wetterstationen in Städten nicht verfügbar. Daher konzentrierte sich der erste Teil dieser Arbeit auf die Frage, wie Open-Source Fernerkundungs-, GIS- und Crowdsource-Wetterdaten genutzt werden können, um die räumliche Verteilung der nächtlichen Lufttemperatur einen Tag im Voraus mit einer Auflösung von 30 m für die gesamte Stadt mittels ML vorherzusagen. Es wurden drei Vorhersageszenarien getestet: (1) ausschließlich fernerkundlich basiert, (2)

ausschließlich basiert auf räumlich interpolierte Crowdsource-Lufttemperatur Daten vom Vortag, und (3) basiert auf sowohl fernerkundlichen als auch Crowdsource-Daten vom Vortag. Da Crowdsource-Wetterdaten mit mehr Unsicherheiten behaftet sind als herkömmliche Wetterdaten, wurde die Validierung sowohl mit Crowdsource- als auch mit herkömmlichen Lufttemperaturdaten unter Verwendung einer "leave-one-date-out"-Kreuzvalidierung durchgeführt. Die Verwendung von Fernerkundungsdaten zeigte die höchste Gesamtgenauigkeit bei der Validierung mit herkömmlichen Lufttemperaturdaten. Das zeigt, dass Fernerkundungsdaten eine genaue räumliche Vorhersage der Lufttemperatur, ergänzend zu den privaten Wetterstationen von Bürgern, ermöglichen. Die Genauigkeit war bei der Validierung mit Crowdsource-Daten hoch, nahm jedoch bei der Validierung mit unabhängigen Daten ab, was wahrscheinlich auf die Installation der privaten Wetterstationen in der Nähe von Gebäudewänden zurückzuführen ist. Die Genauigkeit war jedoch in stark versiegelten städtischen Strukturen am höchsten, in denen das Hitzerisiko zugleich am stärksten ausgeprägt ist. Das zeigt den Nutzen dieses Ansatzes für die Identifizierung gefährlicher Hotspots in der Stadt.

ET entspricht dem Wasserverbrauch von Vegetation und leistet somit einen wesentlichen Beitrag zur Abkühlung, welcher sie zu einem Schlüsselindikator für urbane Nachhaltigkeit macht. ET ist definiert als das Wasser, welches von der Landoberfläche in die Atmosphäre transportiert wird. Sie besteht hauptsächlich aus Bodenverdunstung, Transpiration über Pflanzen und Evaporation von interzeptierten Niederschlägen innerhalb terrestrischer Ökosysteme. Um ET-Modellierungen in städtischen Gebieten zu verbessern, wurde ein neuartiger Ansatz entwickelt, der verschiedene Landbedeckungen in KI-Algorithmen integriert. Die Flux-Footprint-Modellierung, welche den räumlichen Fußabdruck der Eddy-Kovarianz-Messungen (EC) schätzt, wurde verwendet, um gewichtete Mittelwerte aus Fernerkundungs- und Geoinformationsdaten zu gewinnen. Die Einbeziehung von Fernerkundungsdaten aus den Flux-Footprints, verbesserte die Vorhersagegenauigkeit der Modelle. In der nächsten Phase wurde dieser Ansatz erweitert, um die ET für die gesamte Stadt mit einer hohen räumlichen (10 Meter) und zeitlichen (stündlich) Auflösung für ein Jahr zu modellieren. Die Validierung zeigte, dass dieser Ansatz für die Kartierung der ET in Städten zuverlässig ist. Zur stadtplanerischen Verbesserung wurden schließlich die räumlichen Variationen der beiden modellierten Indikatoren für das thermische Wohlbefinden bezüglich verschiedener Landbedeckungsklassen analysiert. Somit kann der Einsatz von Fernerkundung und KI zur Charakterisierung von Indikatoren für das thermische Wohlbefinden Stadtkonzepte unterstützen, welche auf die Verringerung von Hitzerisiken für eine zunehmend urbane Weltbevölkerung abzielen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI: Artificial intelligence

avNNet: Model averaged neural network

BH: Building height

CNN: Convolutional neural network

CWS: Citizen weather stations

DL: Deep learning

DWD: German Meteorological Service

EC: Eddy covariance

ET: Evapotranspiration

ETo: Reference evapotranspiration

GBM: Stochastic gradient boosting

FAI: Frontal area index

GIS: Geographic information system

ISF: Impervious surface fraction

L-8: Landsat 8

LAI: Leaf area index

LCZ: Local Climate Zone

LE: Latent heat flux

LULC: Land use land cover

LST: Land surface temperature

MAE: Mean absolute error

ML: Machine learning

NBS: Nature-based solutions

NDBI: Normalized Difference Built-up Index

NDVI: Normalized Difference Vegetation Index

NRMSE: Normalized root mean square error

P: Precipitation

PAI: Plan area index

Pbias: Percent bias

QC: Quality control

RF: Random Forest

RH: Relative humidity

RMSE: Root mean square error

SEB: Surface energy balance

SVF: Sky view factor

SVP: Saturated vapor pressure

Tair: Air temperature

Tcrowd: Crowdsourced air temperature

$T_{crowddev}$: Deviation of crowdsourced air temperature from weather station Berlin-Tempelhof

$Testing_{crowd}$: Validation using crowdsourced weather data

$Testing_{ref}$: Validation using data from conventional weather stations

T_{mod}: Modeled air temperature

TIR: Thermal Infrared

UAV: Unmanned Aerial Vehicle

UCO: Urban Climate Observatory

UMEP: Urban Multi-scale Environmental Predictor (QGIS plug-in)

UHI: Urban heat island

UGI: Urban green infrastructure

VF: Vegetation fraction

VH: Vegetation height

VPD: Vapor pressure deficit

z0: Roughness length

zd: Zero-plane displacement length

zHstd: Standard deviation of building heights

Chapter 1: Introduction

This thesis is devoted to better modeling and understanding two key thermal comfort indicators (evapotranspiration (ET) and air temperature (Tair)) using remote sensing, open geodata and artificial intelligence. The following sections provide an overview of how urbanization affects human well-being, especially in the context of heat risk (Section 1.1.1), present challenges in this area of research (Section 1.1.2), and introduce how state-of-the-art technologies can be used to characterize two key thermal comfort indicators (ET and Tair) (Section 1.2). The study area is introduced in Section 1.3. Finally, the research questions and objectives and the structure of the thesis are presented in Sections 1.4 and 1.5, respectively.

1.1 BACKGROUND AND MOTIVATION

1.1.1 THE EFFECTS OF URBANIZATION ON HUMAN WELL-BEING AND HEALTH

A rapidly growing proportion of the world's population resides in cities; currently, 55% of the 7.6 billion global population is urban, which is expected to increase to 68% by 2050 (United Nations, 2019). Climate and global change are degrading the quality-of-life of urban dwellers, where heat extremes, water insecurity, and flooding are even greater risks than in rural areas (Kundzewicz et al., 2014; Li and Bou-Zeid, 2013; Manoli et al., 2019). One of the greatest threats to residents of urban settlements are heat extremes, whose risk is exacerbated from the interaction between increasingly frequent heat waves and the urban climate (Li and Bou-Zeid, 2013; Meehl and Tebaldi, 2004).

Heat stress is amplified in cities due to the urban heat island (UHI) effect, which describes the higher air and surface temperatures in dense urban areas compared to their rural surroundings (Oke, 1982; Voogt and Oke, 2003). The land use change associated with urbanization drives the UHI effect by increasing solar absorption, heat trapping, and anthropogenic heat from vehicles,

industry, and buildings, while decreasing ET from vegetation (Christen and Vogt, 2004; Oke et al., 1991). During heatwave events, the UHI effect magnifies exposure to extreme heat, leading to detrimental effects on human health and increased mortality (Dousset et al., 2011; Kovats and Hajat, 2008; Mora et al., 2017). Heatwaves are one of the most lethal climate-related events (Mora et al., 2017), with heat-related mortalities reported for cities around the world (Dousset et al., 2011; Schifano et al., 2009; Shaposhnikov et al., 2014). In Europe, heatwaves accounted for more than 123,000 fatalities between 1980 and 2020, or more than 85% of fatalities from all weather-related events (EEA, 2022a). Extreme heat affects human health detrimentally, aggravating respiratory and cardiovascular disease directly (Kovats and Hajat, 2008) and indirectly through the intensification of air pollution (Krug et al., 2019). Heat exposure is especially dangerous for vulnerable populations, including the elderly, infants, and persons with chronic disease (Kovats and Hajat, 2008). Urban heat is also associated with higher energy demand from air conditioning, leading to more carbon emissions and financial burden on low-income residents (Santamouris, 2020).

ET enhances thermal comfort, reduces stormwater runoff, and approximates the irrigation water demand of green spaces, making it a key indicator for urban sustainability and resilience (Saher et al., 2021). Terrestrial ET is defined as the water transported from the land surface to the atmosphere (Wang and Dickinson, 2012). The main sources of ET are soil evaporation, plant transpiration, and interception loss (Miralles et al., 2020) (Figure 1.1), with transpiration contributing 80 - 90% of ET in vegetated ecosystems (Jasechko et al., 2013). Transpiration, the release of water vapor into the atmosphere occurring when plants open their stomatal pores, is a necessary cost of photosynthesis (Wang and Dickinson, 2012). The energy required for the phase change from liquid water to water vapor (transpiration) removes heat from the leaf, leading to a cooling effect (Wang and Dickinson, 2012). Evaporation also cools the land surface through the energy absorbed during the phase change (Wang and Dickinson, 2012). Interception loss, or the evaporation of intercepted precipitation, is a significant source of ET in natural environments, especially in forests (Miralles et al., 2020). Interception loss plays an even larger role in cities due to their impervious surfaces (Ramamurthy and Bou-Zeid, 2014). In cities, anthropogenic emissions also contribute to ET. Human activities, including building energy use, vehicle traffic, industry, and human metabolism, release water vapor (Karsisto et al., 2016; Kotthaus and Grimmond, 2012; Ward et al., 2013). However, these anthropogenic sources of ET are not beneficial for urban heat mitigation (Rocha et al., 2022a).
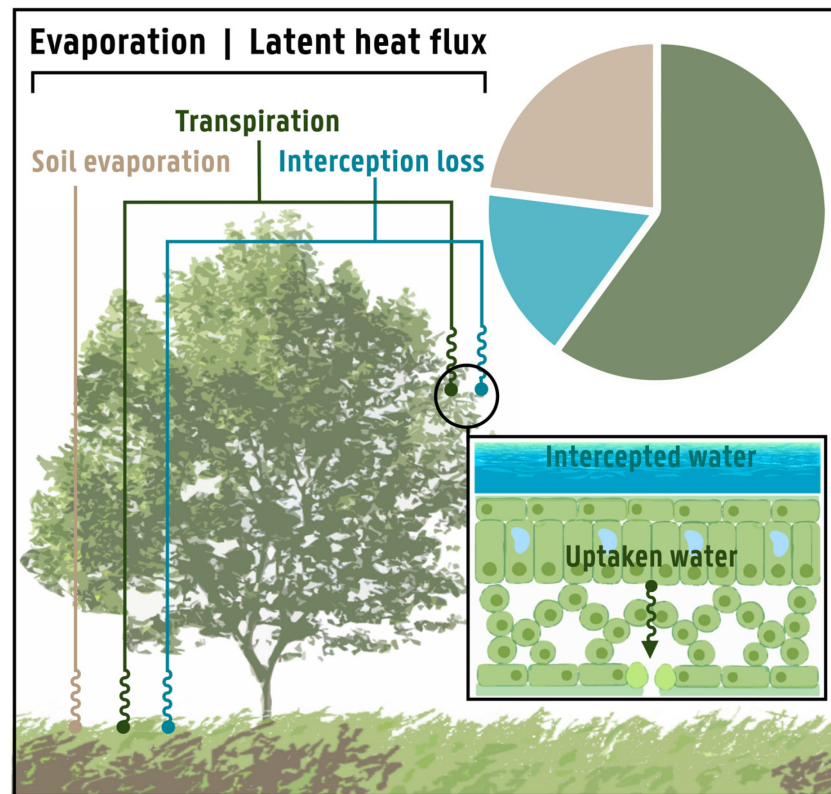
Figure 1.1: The main components of terrestrial ET (also referred to as 'evaporation' and 'latent heat flux') in natural ecosystems (Miralles et al., 2020). The pie chart shows the approximate share of each component globally according to Wei et al. (2017). The bottom-right box portrays a cross-section of a wet leaf, with liquid water transitioning into vapor on its surface (interception loss) and inside the leaf (transpiration).

Replacing vegetation with artificial surfaces generally leads to lower ET from transpiration over urban areas, causing higher Tair (the UHI effect) (Wong et al., 2021). While ET itself modifies the microclimate, it is at the same time deeply driven by atmospheric conditions, including Tair, solar radiation, and relative humidity (Foltýnová et al., 2020). Aside from microclimate, ET is dependent on land cover characteristics, including vegetation type, volume, species and soil moisture, which are highly spatially variable within an urban environment (Nouri et al., 2015).

The utility of urban ET goes beyond mitigating heat risk through evaporative cooling, as ET is also a central component in the hydrological cycle (Gessner et al., 2014; Nouri et al., 2015). Urban water interfaces are defined as boundary zones between components or subsystems of the entire urban water system between which fluxes of water, matter, and energy occur (Gessner et al., 2014) (Figure 1.2). ET constitutes a major part of a continuous flow of water within the soil-plant-atmosphere continuum, leaving soil surfaces through evaporation and plant stomata through transpiration (Damm et al., 2018; Wang and Dickinson, 2012). Thus, ET bridges one of the main urban water interfaces: the interface between the soil, vegetation, and atmosphere (Damm et al.,

2018; Gessner et al., 2014). Sustainably managing water resources in urban areas, which are increasingly under pressure from climate and global change, requires an improved understanding of urban water interfaces (Gessner et al., 2014). Quantifying ET is crucial to reducing the water footprint of cities by optimizing the irrigation needed to maintain the ecosystem services provided by urban vegetation, such as cooling (Nouri et al., 2019). Characterizing ET in cities can also resolve how much rainwater remains available for anthropogenic use, such as toilet flushing (Gessner et al., 2014). Finally, lower ET rates in cities, combined with limited stormwater infiltration over impervious surfaces, result in higher flooding risk (Berland et al., 2017; Wan et al., 2021).
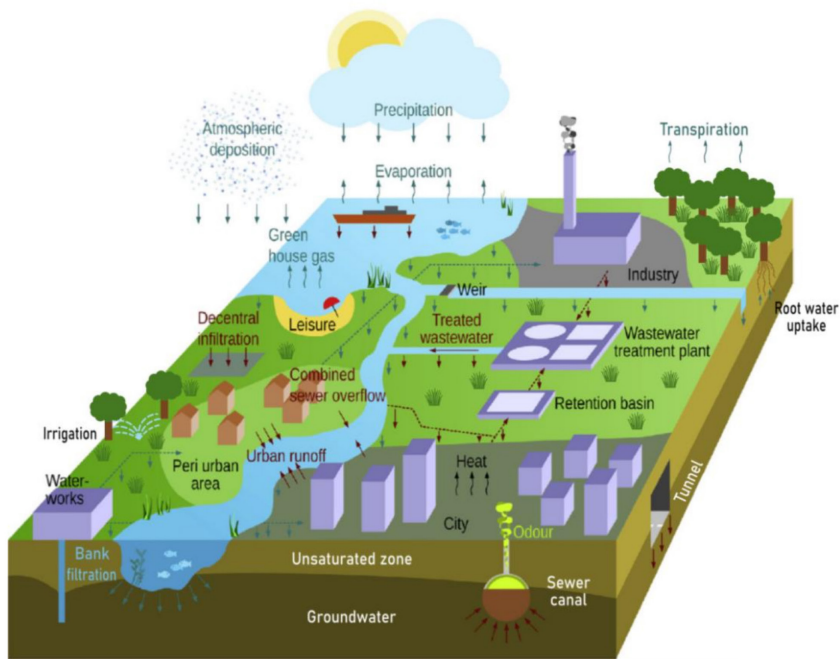
Figure 1.2: The urban water cycle (Gessner et al. (2014); modified by Kuhlemann (2022)).

Urban greening has emerged as one of the main solutions for the various threats to the livability of cities, including extreme heat and flooding (Gaffin et al., 2012; Wong et al., 2021). Vegetation and the ecosystem services it provides, with one of the major ones being cooling via ET, are central components of various urban sustainability concepts, including Nature-Based Solutions (Somarakis et al., 2019), Sponge Cities (Gaines, 2016), Water Sensitive Urban Design (Fletcher et al., 2008), and Low-Impact Development (Qin et al., 2013). Urban green infrastructure (UGI) such as urban parks and street trees can lower Tair during heat waves by enhancing ET and providing shade, while also providing numerous additional co-benefits, including recreational space, improved air quality, and carbon sequestration (Wong et al., 2021). To properly implement and manage UGI in order to promote thermal comfort in cities, however, spatially distributed, high-resolution data on where hotspots of heat risk (high Tair and low ET)

occur within an urban area is needed, but has until now remained largely unavailable (Saher et al., 2021; Zhou et al., 2019).

1.1.2 RESEARCH CHALLENGES IN PROVIDING DATA TO SUPPORT UHI MITIGATION

Characterizing indicators of urban heat, such as Tair and ET, is challenging due to the high heterogeneity of urban environments, where artificial and natural components intersect in a complex mosaic (Gessner et al., 2014; Middel et al., 2022). Most existing process-based models have not been developed to handle the horizontal and vertical complexity of cities, with the parameters of urban climate and water systems constantly varying dynamically in both space and time (Gessner et al., 2014; Ho et al., 2014; Saher et al., 2021). To plan sustainable and climate-resilient cities, accurately characterizing thermal comfort at a high spatio-temporal resolution is needed (Middel et al., 2022; Milojevic-Dupont and Creutzig, 2021). State-of-the-art technological and methodological approaches, such as remote sensing, artificial intelligence (AI), and novel sensor data offer a way forward to better modeling urban heat. This thesis aims to better characterize the soil-vegetation-atmosphere interface in urban areas, which can support both heat risk mitigation and sustainable water management.

1.1.2.1 TAIR IN URBAN AREAS

The majority of work on mapping urban heat has utilized land surface temperature (LST) derived from satellite imagery (Zhou et al., 2019). Tair is a more appropriate indicator of human health risk than LST (Ho et al., 2016). However, the overwhelming focus on LST in UHI studies has mainly been motivated by the availability of spatially explicit data, while Tair has until recently only been measured by a limited set of meteorological stations widely spaced within a city (Muller et al., 2013). Tair is highly spatially variable at a fine scale (1-100 m) in urban areas due to contrasting land cover, sun exposure, wind dynamics, and vegetation types (Oke and Maxwell, 1975; Voogt and Oke, 2003).

The existing approaches to model and map Tair using remote sensing data include (1) temperature vegetation index (TVX) methods, (2) energy balance models, and (3) statistical and machine learning regression models (Ho et al., 2014; Venter et al., 2020). The TVX method is based upon the negative semi-empirical relationship between NDVI and LST, but it can only estimate Tair for large areas with gradual temperature changes at a coarse resolution, making it inappropriate for urban areas (Ho et al., 2014; Stisen et al., 2007). Energy balance models estimate Tair based on the principles of thermodynamics, but rely on parameters for which spatially distributed, high-resolution data is rarely available in cities (Mostovoy et al., 2006; Sun et al., 2005). The UrbClim model estimates urban Tair based on a simple parameterization of the urban surface energy balance coupled to a 3-D atmospheric boundary layer scheme (De Ridder et al., 2015). However, due to model simplifications necessary to increase computational speed, the

UrbClim model cannot account for several key aspects of the urban climate, including local circulation patterns (the city breeze) and the influence of the urban canopy layer (De Ridder et al., 2015). Statistical approaches are based on empirical regression modeling, which includes linear regression (Nichol et al., 2009) or ML regression (see Section 1.2.1) and are most promising for Tair mapping in urban areas (Ho et al., 2014; Venter et al., 2020).

Recently, the proliferation of private weather stations providing crowdsourced weather data in cities worldwide has made it possible to measure microclimatic variation in a cost-effective way (Fenner et al., 2017; Meier et al., 2017; Muller et al., 2015; Napoly et al., 2018). Although much more densely spaced than traditional meteorological stations (with hundreds to thousands of stations within a city), citizen weather stations such as "Netatmo" are too unevenly distributed and spatially clustered (Fenner et al., 2017; Meier et al., 2017), making spatial interpolation insufficient to represent microscale Tair variation adequately (Ho et al., 2016; Venter et al., 2020). Linking crowdsourced Tair data (included within the term "geodata" throughout this thesis) to remotely sensed satellite imagery provides the opportunity to upscale these measurements based on attributes such as vegetation density, morphology, and albedo.

1.1.2.2 ET IN URBAN AREAS

Despite the importance of ET for urban heat mitigation and the urban water cycle, ET estimation in urban areas remains a critical area of research, with most ET studies focusing on agricultural or natural areas (Nouri et al., 2015, 2013; Saher et al., 2021). Characterizing ET is highly challenging even in a predominantly homogeneous, natural environment (Damm et al., 2018). However, the urban environment is exceptionally heterogeneous, with numerous interfaces between artificial and natural systems (Gessner et al., 2014) and a high variation in soil attributes, land cover, vegetation type and height, and microclimate (Nouri et al., 2013). Therefore, similarly to Tair, the spatial variability of ET occurs at a fine scale in cities. Point-based field measurements of ET, such as sap flow and lysimeters, cannot capture the spatial variation of ET over urban fabric (Litvak et al., 2017; Nouri et al., 2013); furthermore, they are difficult to deploy in cities in a distributed manner due to lack of accessibility and risk of vandalism (Gillefalk et al., 2022).

Eddy covariance (EC) is one of the best-suited methods to directly measure urban ET (Nouri et al., 2013). The EC micrometeorological technique, which is most commonly deployed on a fixed tower, measures high speed fluxes of water, heat, gas, and momentum in the atmospheric boundary layer in a large area (up to 1 km^2) (Nouri et al., 2013). Nevertheless, the EC instrumentation is still a point-based measurement which is limited by large data gaps (20-60 %) due to equipment failures and low turbulence conditions (Moffat et al., 2007). Furthermore, a constantly-changing upwind area (a 'footprint' or 'source area') contributes to the ET signal measured from an EC tower (Foken, 2016; Kljun et al., 2002; Schmid and Oke, 1990). Due to continuous footprint variation, heterogeneous land cover has a dynamic effect on EC

measurements (Kotthaus and Grimmond, 2014). Linking EC and remote sensing data holds potential for upscaling urban ET, an approach which has so far been hampered by the lack of ET measurements over a variety of land covers (Wang et al., 2016).

Due to the constraints of field measurements, physical models have mainly been used to estimate urban ET (Nouri et al., 2015; Saher et al., 2021). Two of the most common types of models, surface energy balance (SEB) models and physically based hydrological models, were not developed for urban areas and therefore do not perform well for urban ET estimation (Nouri et al., 2015; Saher et al., 2021). Models specifically developed for urban applications, such as urban land-surface models (ULSMs), are generally not applied to entire cities at a high resolution, due to the difficulty of estimating the input parameters in a heterogeneous urban terrain (Rocha et al., 2022b; Saher et al., 2021). For a more detailed overview of the available urban ET modeling approaches, refer to Sections 3.1 and 4.1 and to the review papers Saher et al. (2021) and Nouri et al. (2015). The premise of this thesis is to investigate if machine learning combined with remote sensing imagery and in situ data can overcome the shortcomings of physical models to better characterize urban ET at a high temporal and spatial resolution (Chapters 3-4).

1.2 CHARACTERIZING THERMAL COMFORT INDICATORS USING AI AND REMOTE SENSING

In this work, state-of-the-art techniques (AI and remote sensing) are used to model thermal comfort indicators at a high spatio-temporal resolution. The following sections introduce the principles of AI and remote sensing and their utility for characterizing Tair and ET in an urban environment.

1.2.1 ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is an umbrella term comprising both machine learning (ML) and deep learning (DL) (Chollet and Allaire, 2018). AI falls within the domain of predictive modeling or statistical learning, a process of uncovering relationships within data to predict an outcome (Hastie et al., 2009; Kuhn and Johnson, 2013; Vapnik, 1999). ML and DL methods allow for the generalization of patterns on new data outside of the training sample (Vapnik, 1999). ML algorithms can address different learning tasks, with the main ones being classification and regression (Hastie et al., 2009; Kuhn and Johnson, 2013). The major paradigms of ML are supervised learning (where a target value is specified during training), unsupervised learning (where the target value is unknown), and reinforcement learning (learning through actions within an environment) (Hastie et al., 2009). For a more comprehensive introduction to ML, refer to Hastie et al. (2009) and Kuhn and Johnson (2013). Within the scope of this thesis, the focus is on supervised regression algorithms, which predict a numeric outcome (such as ET) by learning from ground-truth data (Kuhn and Johnson, 2013).

AI methods have shown several advantages over process-based approaches due to their capacity to uncover patterns from data which are not yet well-defined in physical models and their superior accuracy for numerous applications (Reichstein et al., 2019). AI algorithms can also integrate diverse and large data into prediction and can facilitate interpretability by, for example, providing variable importance (Kuhn and Johnson, 2013; Roscher et al., 2020a). Spatial prediction, or mapping, of biogeophysical parameters such as photosynthesis or ET at local to global scales is one realm where supervised ML regression has been especially successful (Reichstein et al., 2019). AI also holds great potential to advance resilience to climate change in urban areas (Creutzig et al., 2019; Middel et al., 2022; Milojevic-Dupont and Creutzig, 2021; Rolnick et al., 2023). Using AI methods to link satellite data with other datasets can generate spatially explicit products relevant for climate change mitigation, which can facilitate neighborhood-scale policy and urban planning solutions (Creutzig et al., 2019; Milojevic-Dupont and Creutzig, 2021).

One of the main sub-categories of supervised ML regression are tree-based models, which contain nested "if-then" statements that split the data into possible outcomes (Hastie et al., 2009; Kuhn and Johnson, 2013). While tree-based approaches are advantageous due to their capacity to easily integrate different types of predictors and implicit feature selection, simpler tree-based models suffer from model instability and poor accuracy (Kuhn and Johnson, 2013). To resolve this problem, ensemble techniques such as bagging were introduced in the 1990s to provide more stable predictions by combining many models' predictions (Kuhn and Johnson, 2013). One of the resulting algorithms, Random Forests, has proved to be one of the most popular ML algorithms due to its high and consistent predictive accuracy for various applications and particularly in the remote sensing field (see Chapters 2-4). Boosting, another ensemble approach, which combines (or 'boosts') the predictions of several "weak" classifiers to produce a superior accuracy, was also introduced in the early 1990s (Hastie et al., 2009; Kuhn and Johnson, 2013). Combining the concept of boosting with regression trees has resulted in powerful algorithms such as stochastic gradient boosting (Chapter 2), where numerous trees with a restricted depth (number of splits) are used to generate a prediction (Kuhn and Johnson, 2013).

Neural networks are a type of nonlinear empirical model which estimates the outcome using hidden units, or an intermediary set of unobserved variables (Hastie et al., 2009; Kuhn and Johnson, 2013). The simplest type of neural network, still categorized as just ML, has a single layer of hidden units (a single-layer feed-forward network) and is referred to as an artificial neural network (ANN) (Hastie et al., 2009; Kuhn and Johnson, 2013). Averaging the predictions of several ANNs, similarly to ensemble methods for regression trees, is beneficial for producing a more stable prediction (Kuhn and Johnson, 2013); such an algorithm (model averaged neural network) is applied in Chapter 2. A deep neural network (DNN) contains more layers of hidden

units and is therefore considered 'deep' learning, which is in fact a subfield of machine learning (Chollet and Allaire, 2018). Adding convolutional layers, which can learn local spatial or temporal patterns, to DNNs produces a convolutional neural network (CNN) (Chollet and Allaire, 2018). For instance, a type of CNN suitable for time-series analysis (1D CNN) is used in Chapter 3 to predict ET.

1.2.2 REMOTE SENSING

Remotely-sensed data provides spectral, thermal, and morphological information about the landscape in a spatially explicit manner. Satellite remote sensing is a cost-effective way of providing regularly-updated, spatially distributed data, especially in cities, where dense networks of in situ measurements are not feasible due to both cost and permits. Most of the processes related to the soil-plant-atmosphere continuum, such as ET, cannot directly be observed by remote sensing (Damm et al., 2018). However, various proxies, which have a relationship with the process of interest, can be detected by changes in absorption or scattering observable by remote sensing (Damm et al., 2018).

Both ET and Tair vary based on vegetation density, vitality, height, 3D structure (or exposure to sun and wind), imperviousness, soil moisture, and land cover (or the thermal properties of the surface cover), which are highly variable in urban areas (Nouri et al., 2013; Oke and Maxwell, 1975; Voogt and Oke, 2003). Satellite imagery can directly estimate some of these attributes (such as vegetation height) or serve as a proxy of others (such as vegetation indices as a proxy of vegetation vitality). Medium-resolution satellite imagery is especially appropriate for large-scale modeling of urban heat and for urban planning applications due to its relatively high spatial resolution (10-30 m for spectral imagery) and open access (Venter et al., 2020), with the best-known examples being the Landsat (Landsat 7 and 8) and Sentinel-2 satellites. Vegetation indices, which are usually derived from spectral data, can be used as proxies of leaf area index (LAI), canopy density, and vegetation health (Saher et al., 2021). The Normalized Difference Vegetation Index (NDVI) is the most commonly used vegetation index in both ET and Tair studies in urban areas (Ho et al., 2014; Nouri et al., 2015; Saher et al., 2021).

Thermal infrared (TIR) remote sensing data has most commonly been used to estimate land surface temperature (LST) and the surface urban heat island (SUHI), which is defined as the radiative temperature difference between urban and non-urban surfaces (Zhou et al., 2019). TIR data can be used to estimate leaf temperature, which is related to transpiration rates (Damm et al., 2018). Though most UHI studies have focused on utilizing LST directly as an indicator of urban heat, this research field has recently shifted its focus to linking satellite imagery to in-situ measurements in order to model more appropriate indicators of human health and thermal comfort, such as Tair (Ho et al., 2016; Zhou et al., 2019). In fact, LST and Tair are considerably spatially and temporally mismatched over urban areas due to the influence of wind, turbulence,

the 3-D urban structure, and other factors (Azevedo et al., 2016; Mirzaei and Haghighat, 2010). Nevertheless, LST can approximate the thermal properties of the land surface, including heat capacity and thermal conductivity, and is therefore frequently used as an input parameter in empirical and physical models estimating Tair and ET (Bastiaanssen et al., 1998; Zhou et al., 2019).

Active remote sensing systems such as light detection and ranging (LiDAR) can be used to estimate 3-D structure (building height and vegetation height), which has a major influence on urban microclimate as a driver of wind flow, shade, and the volume of vegetation available for ET (Damm et al., 2018; Masson et al., 2020). Vegetation structure, including its height and morphology, is an indicator of several factors influencing transpiration, including the number of leaf stomata, canopy conductance, and canopy roughness (Damm et al., 2018; Yebra et al., 2013). Finally, several sources of satellite imagery, such as Sentinel-1 Synthetic Aperture Radar (SAR) data, another product of active remote sensing, still remain unexplored for urban ET and Tair estimation and are discussed in section 5.3. For a general overview, Damm et al. (2018) provides a comprehensive review of the remote sensing imagery suitable for studying plant-water relations, with a focus on transpiration (Figure 1.3). Figure 1.3 shows how remote sensing technologies can approximate physiological, biochemical, and structural parameters related to plant water availability, several of which have a relationship to transpiration (Damm et al., 2018). As a note, the terms "geodata" and "GIS data" are used in this thesis to refer broadly to georeferenced information, which includes products derived from remote sensing imagery (such as building height data) in addition to crowdsourced weather data.
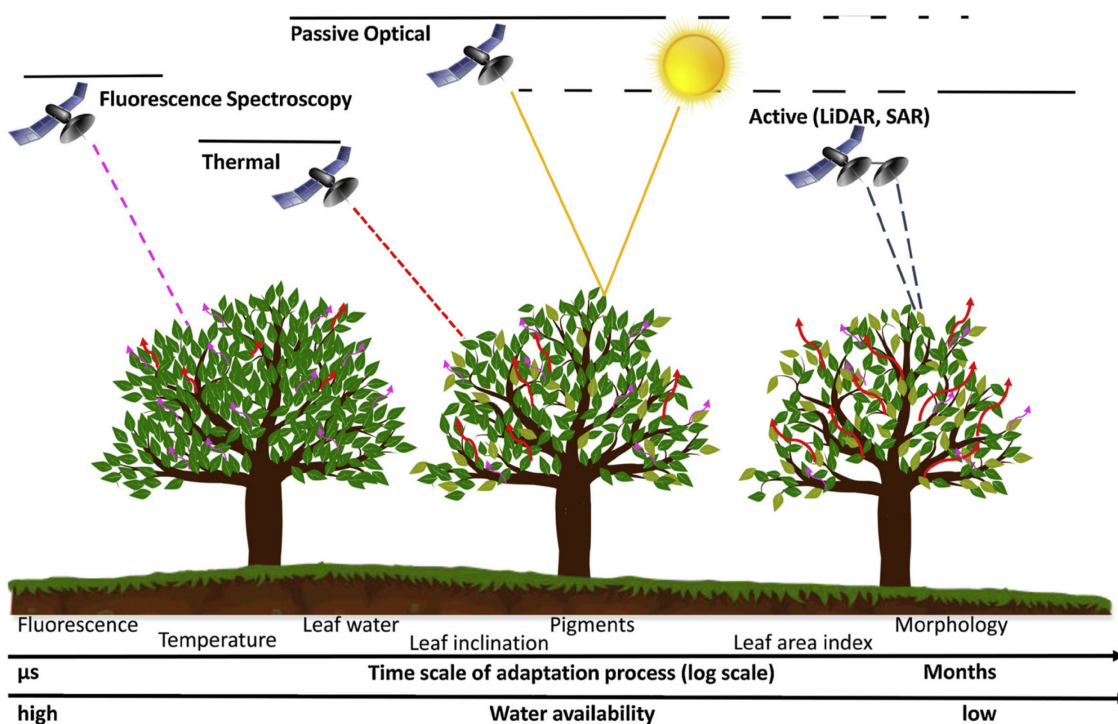
Figure 1.3: An overview of the remote sensing technology relevant for assessing plant-water relations, including transpiration (Damm et al., 2018). Bold black lines refer to the application spectrum of the remote sensing technology, while dashed lines imply limited applicability. Moving in the right direction, water availability decreases, with corresponding adaptation mechanisms by plants. Purple arrows show fluorescence radiation emission from plants, while red arrows indicate thermal radiation emission.

Remote sensing predictors of Tair used in regression modeling include LST, NDVI, sky view factor (SVF), and land cover (Benali et al., 2012; Ho et al., 2014; Mostovoy et al., 2006). For an overview, the main modeling approaches used to estimate urban Tair and ET are outlined in section 1.1.2. Until recently, regression models estimating Tair relied on Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data at a 1 km resolution, which is insufficient for urban planning (Venter et al., 2020). This thesis itself is situated at the forefront of the use of ML regression to map urban Tair at a higher resolution by integrating crowdsourced weather data and medium-resolution satellite imagery (Landsat 8) into modeling (Chapter 2).

Recent studies have shown that integrating a variety of remote sensing and GIS data is indispensable to modeling urban ET (Nouri et al., 2015; Saher et al., 2021). Satellite imagery has generally been used to provide information on soil moisture and plant traits in ET models (Nouri et al., 2015). Remote sensing data most commonly used in urban ET models are vegetation indices and LST (Nouri et al., 2015). Satellite imagery has recently shown potential to serve as a proxy of anthropogenic heat, defined as the heat emitted from human metabolism, air conditioning and heating, vehicle exhaust, and industry (Allen et al., 2011; Faridatul et al., 2020). For instance, Faridatul et al. (2020) adapted the Surface Energy Balance Algorithm for Land (SEBAL) to the urban environment by adding anthropogenic heat to the energy budget based upon a land cover map. Although anthropogenic heat plays a substantial role in the urban surface energy budget, it is omitted from most remote sensing-based ET models, leading to an underestimation of ET (Allen et al., 2011; Cong et al., 2017; Faridatul et al., 2020). Finally, it should be noted that though land cover drives the spatial variation in ET in an urban environment, meteorological conditions force most of the temporal variation in ET, making meteorological data essential to ET estimation (Nouri et al., 2015; Rocha et al., 2022b). Accurate estimates of Tair in particular are critical to quantifying ET (Damm et al., 2018).

1.3 STUDY AREA

Berlin is the largest city in Germany with respect to both population (3.7 million) and area (891 km^2) (Statistical Office of Berlin-Brandenburg, 2019). Berlin is a mid-latitude city characterized by a climate transitioning between a temperate oceanic and a humid continental climate with warm summers (Beck et al., 2018). The long-term annual mean (1991-2020) Tair

and precipitation are 10.4 °C and 575 mm, respectively (DWD, 2022). Berlin has a fairly flat topography with an average elevation of 34 m above sea level.

Berlin is a hotspot for heat stress-related mortalities (Gabriel and Endlicher, 2011; Krug et al., 2019), with 5% of all deaths between 2001 and 2010 attributed to elevated Tair (Scherer et al., 2013). Around 34% of Berlin's land surface is impervious (Senate Department for Urban Development and Housing, 2017), which is mainly concentrated around the city center (Figure 1.4). Heat-stress risk is most pronounced in the mid-rise dense residential areas in the city center and high-rise multi-story houses in the eastern part of the city, where higher Tair coincides with a high proportion of elderly residents (Dugord et al., 2014). At the same time, Berlin has been identified as a pioneer city worldwide in the use of green infrastructure to reduce its heat risk and water footprint (Liu and Jensen, 2018). According to the European Environment Agency, Berlin has one of the highest percentages of 'green infrastructure' (51%), a term comprising both green (vegetated) and blue (water) spaces, among the 38 European capital cities (EEA, 2022b). Surface water accounts for 6.6% of the city's area (SenUVK, 2021).

Berlin is a suitable pilot city for modeling thermal comfort indicators due to its exceptional infrastructure of atmospheric measurements. The Urban Climate Observatory (UCO) Berlin has been operated by the Chair of Climatology at the Technische Universität Berlin since 1990 and thus represents one of the longest time-series of data on urban climates in the world (Scherer et al., 2019). The UCO Berlin provided additional independent Tair data for validation of modeled Tair in Chapter 2. Furthermore, two flux towers, TUB Campus Charlottenburg (TUCC) and Rothenburgstrasse (ROTH), were installed in 2014 and 2018, respectively, by the UCO Berlin (Figure 1.4). These flux towers provide continuous measurements of heat fluxes using the EC technique, including ET, and are central to the work in Chapters 3-4. For a detailed description of the flux towers, refer to Section 3.2.1.
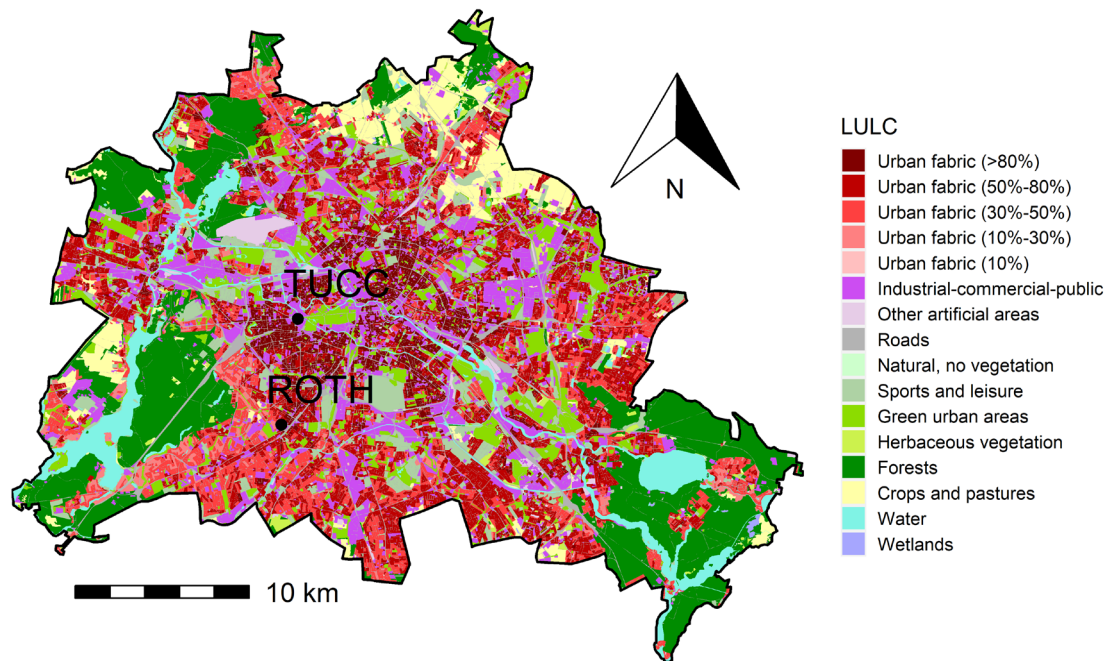
Figure 1.4: Land use land cover (LULC) map of Berlin, Germany (Urban Atlas (2018); Rocha et al. (2022a); modified). The two flux towers are marked with black dots.

1.4 RESEARCH QUESTIONS AND OBJECTIVES

The overall goal of this thesis is to better characterize the spatial variation of two essential thermal comfort indicators (Tair and ET) in an urban environment using open-source remote sensing imagery, geodata, and AI. To support this goal, this thesis has two main objectives:

Objective 1: Investigate how accurately thermal comfort indicators in urban areas (ET and Tair) can be modeled at a high spatial resolution with remote sensing imagery, geodata, and meteorological data using ML regression algorithms (Chapters 2-4).

Objective 2: Assess the spatial variation of the modeled ET and Tair across Berlin to provide insights for urban planning and policy focused on heat risk mitigation (Chapters 2 and 4).

Tair, especially at nighttime, is one of the best indicators of heat-related mortality and adverse health effects (Dousset et al., 2011; Gabriel and Endlicher, 2011; Ho et al., 2016; Kovats and Hajat, 2008). Nevertheless, the remote sensing field has so far focused on mapping LST due to the relative ease of attaining spatially explicit data, despite its poor link to human heat exposure (Ho et al., 2016; Zhou et al., 2019). Tair products with a high spatial resolution in urban areas are currently lacking, but could greatly benefit urban planners aiming to reduce heat-related human mortality and health risks. To investigate how the emergence of crowdsourced information can

facilitate the mapping of Tair at a high spatial resolution, the first research question of this thesis is:

Research Question I: How can open source remote sensing, GIS and crowdsourced weather data be used to predict the spatial distribution of nocturnal Tair in an urban area? (Chapter 2)

The developed methodology demonstrated the potential of ML regression to accurately model and map nocturnal Tair using medium-resolution satellite data and other open-source geodata. However, another key indicator of thermal comfort, ET, still remained poorly characterized in urban areas. The importance of ET in cities is far-reaching, as this parameter can approximate the cooling capacity of green spaces as well as their water demand and consumption (Saher et al., 2021; Shashua-Bar et al., 2011). Thus, a better understanding of the spatio-temporal distribution of ET is essential to inform the management of green infrastructure in order to reduce heat risk. Similarly to Tair, the potential of using AI to model ET in urban areas was hampered by the lack of sufficient training data until now.

In order to map ET at a high spatial resolution using an AI approach, it was necessary to first establish a workflow of integrating remote sensing and GIS data into ML regression algorithms. Flux footprint modeling, a method to estimate the constantly-changing source area of ET measurements from flux towers, was tested as a solution to linking heterogeneous urban land cover to the measured ET signal. The second research question of this thesis therefore addressed the gains in predictive accuracy from integrating remote sensing and GIS data over simply using meteorological data:

Research Question II: Is there a benefit to integrating remote sensing and GIS data extracted by flux footprints to model urban ET at a high temporal (half-hourly) resolution? (Chapter 3)

Having demonstrated the potential of this methodology for modeling urban ET at a high temporal resolution, it was then possible to further develop this approach to map ET at a high spatio-temporal resolution in Berlin. Thus, the third research question of this thesis focused on upscaling ET to the city-scale:

Research Question III: Can ET be accurately mapped at a high spatial (10-m) and temporal (hourly) resolution in an urban environment with remote sensing imagery, geodata, and meteorological data using ML and flux footprint modeling? (Chapter 4)

While this thesis was grounded in methodological advances which facilitated the modeling and mapping of thermal comfort indicators (Objective 1), its ultimate goal was to provide insights relevant for urban planners and policy-makers focused on alleviating the risk of heat stress (Objective 2). Therefore, the fourth research question focused on understanding the spatial variation of the modeled thermal comfort indicators:

Research Question IV: What is the spatial variation of modeled ET and air temperature across Berlin and what implications does this variation have for urban planners? (Chapters 2-4)

The spatially explicit products of this thesis can support urban sustainability measures focused on heat risk mitigation and greening. Furthermore, the use of AI and remote sensing to identify hotspots of heat risk can serve as a blueprint for "Smart City" applications tackling the challenges of climate change.

1.5 THESIS OUTLINE

Following this introduction, three main research chapters each constitute a stand-alone manuscript published or submitted to an international peer-reviewed journal. A schematic overview of the main topics of each research chapter is given in Figure 1.5. Chapter 2 addresses Research Question I and objectives 1 and 2 (published in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*). Chapter 3 covers Research Question II and Objective 1 (published in *Science of the Total Environment*). Chapter 4 covers Research Question III and objectives 1 and 2 (in review with *Remote Sensing of Environment*). Please refer to Appendix D for a complete list of publications. Research Question IV, a cross-cutting question concerning the spatial variation of the modeled thermal comfort indicators, is covered in all research chapters. In Chapter 5, the thesis is synthesized, providing a summary of the main results and an outlook for future research.
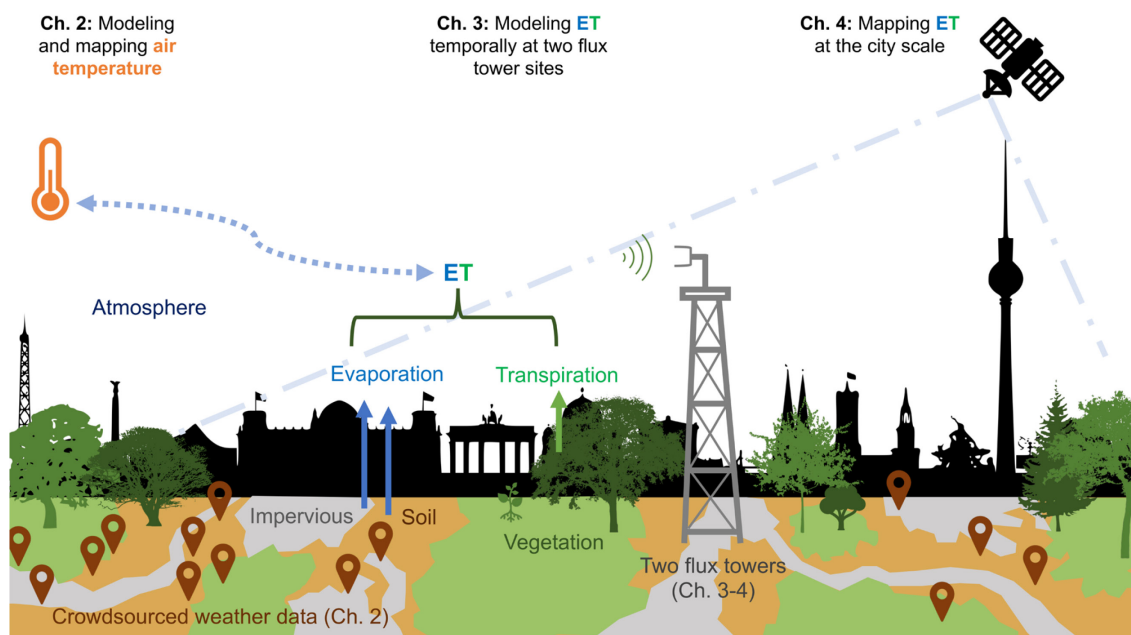
Figure 1.5: Conceptual graphic visualizing the thesis structure.

Chapter 2: Summer nights in Berlin, Germany: Modeling air temperature spatially with remote sensing, crowdsourced weather data, and machine learning

This study was published as:

This is the postprint version of the article.

ABSTRACT

Urban areas tend to be warmer than their rural surroundings, well-known as the "urban heat island" effect. Higher nocturnal air temperature (Tair) is associated with adverse effects on human health, higher mortality rates, and higher energy consumption. Prediction of the spatial distribution of Tair is a step towards the "Smart City" concept, providing an early warning system for vulnerable populations. The study of the spatial distribution of urban Tair was thus far limited by the low spatial resolution of traditional data sources. Volunteered geographic information (VGI) provides alternative data with higher spatial density, with citizen weather stations monitoring Tair continuously in hundreds or thousands of locations within a single city. In this study, the aim was to predict the spatial distribution of nocturnal Tair in Berlin, Germany, one day in advance at a 30-m resolution using open-source remote sensing and geodata from Landsat and Urban Atlas, crowdsourced Tair data, and machine learning (ML) methods. Results were tested with a "leave-one-date-out" training scheme (testing$_{crowd}$) and reference Tair data (testing$_{ref}$). Three ML algorithms were compared - Random Forest (RF), Stochastic Gradient Boosting (GBM), and Model Averaged Neural Network (avNNet). The optimal model based on accuracy and computational speed is RF, with an average RMSE for testing$_{crowd}$ of 1.16 °C ($R^2 = 0.512$) and RMSE for testing$_{ref}$ of 1.97 °C ($R^2 = 0.581$). Overall, the most important GIS predictors

were morphometric parameters and albedo. The proposed method relies on open-source datasets and can, therefore, be adapted to many cities worldwide.

2.1 INTRODUCTION

Urbanization has been accelerating for the past few decades. According to the United Nations (United Nations, 2018), 60% of the global population will be concentrated in urban areas by 2030. Urban areas experience higher air and surface temperatures than surrounding rural areas, well-known as the urban heat island (UHI) effect (Voogt and Oke, 2003). The UHI is mainly driven by the modification of three-dimensional structure, albedo, photosynthetic capacity, land cover, and imperviousness of the landscape, which alters the surface energy balance, and anthropogenic heat from air conditioning, automobiles, and industry (Christen and Vogt, 2004). The UHI of mid-latitude cities is especially distinct during summer nights, when absorbed energy from daytime solar radiation is released, preventing ambient air temperature (Tair) from declining (Christen and Vogt, 2004).

By 2050, urban land expansion is projected to result in a warming of urban Tair of up to 3 °C, increasing extreme heat risks for about half of the future urban population (Huang et al., 2019). Extreme urban heat increases human mortality, with well-documented heat waves such as Paris in 2003 (>4000 deaths) (Dousset et al., 2011). The heat load to city dwellers is more pronounced at night, since daytime vulnerability may be reduced by modifying behavior, whereas the options to enhance sleeping conditions are limited (Gabriel and Endlicher, 2011; Kovats and Hajat, 2008). Elevated nighttime Tair prevent the human body from properly resting and enhance the risk of respiratory and cardiovascular health issues (Dousset et al., 2011). Populations such as elderly people, infants, and hospital patients are especially vulnerable to heat stress (Gabriel and Endlicher, 2011; Kovats and Hajat, 2008; Scherer et al., 2013). Summertime heat increases energy consumption due to air conditioning, straining electricity infrastructure and placing an additional financial burden on low-income households (Santamouris and Kolokotsa, 2015).

The population of Berlin, Germany, has already been adversely impacted by heat waves. Mortality rates increased up to 67.2% during extreme heat waves in Berlin from 1990 to 2006 (Gabriel and Endlicher, 2011). From 2001 to 2010, 5% of all deaths in Berlin were linked to high summer Tair, primarily affecting people over the age of 65 (Scherer et al., 2013). Dugord *et al.* (2014) used simulated Tair for 6 AM in August 2000, population density, and the share of vulnerable people to assess heat-stress risk in Berlin, finding the highest risk in central residential areas and in high-rise multi-story houses in the eastern part of Berlin. Furthermore, the impact on human mortality during heat waves in Berlin is amplified by high ozone concentrations, highlighting the interactive effects of high Tair and air pollution (Krug et al., 2019).

In order to mitigate heat risks for the urban population, high-resolution and spatially distributed Tair data is urgently needed (Scherer et al., 2013; Venter et al., 2020). Traditionally, ground-based measurements of Tair were limited to conventional weather stations, which are not sufficiently dense to portray intra-urban Tair variation. Higher-density Tair monitoring has been achieved by high-resolution measurement networks (Muller et al., 2013) and vehicle-mounted temperature sensors ('traverses') (Schmidt, 1927; Seidel et al., 2016; Shandas et al., 2019), but these approaches are costly to implement and maintain. Recently, crowdsourced atmospheric data has emerged as a source for low-cost, spatially dense, and long-term measurements of Tair in urban environments (Fenner et al., 2017; Meier et al., 2017; Napoly et al., 2018; Seebacher et al., 2019; Venter et al., 2020). Crowdsourced Tair data from 'Netatmo' citizen weather stations (CWS), which belong to the realm of the 'Internet of things' and volunteered geographic information (VGI), has proved to be reliable to study urban Tair with its worldwide distribution, easy and free access, and high spatial density (Chapman et al., 2017; de Vos et al., 2020; Feichtinger et al., 2020; Fenner et al., 2017; Meier et al., 2017; Napoly et al., 2018; Nipen et al., 2020).

Simply spatially interpolating Tair measurements based upon techniques such as inverse distance weighting (IDW) or kriging to map Tair is insufficient, as microscale variations in Tair occur due to differences in surface roughness, land cover, and vegetation density (Voogt and Oke, 2003). Remotely sensed satellite data are indispensable for regression modeling of Tair measurements based on spectral, thermal and morphological information about the urban landscape (Shandas et al., 2019; Venter et al., 2020; Yoo et al., 2018). Spectral imagery can characterize urban land cover through vegetation indices (such as the commonly used Normalized Difference Vegetation Index (NDVI)) and built-up indices (Shandas et al., 2019; Venter et al., 2020; Wicki et al., 2018; Yoo et al., 2018; Zha et al., 2003). Spectral data can also be used to derive albedo (Liang, 2001; Zhu et al., 2019), which controls the energy balance partitioning of urban surfaces (Christen and Vogt, 2004). Land Surface Temperature (LST) derived from thermal imagery is an indicator of the thermal conductivity and heat capacity of urban surfaces, which allows some surfaces to absorb incoming radiant energy more efficiently (Wicki et al., 2018). Energy is stored in the urban fabric during the day and released during the night, causing elevated Tair, particularly in summertime conditions with high solar irradiance (Christen and Vogt, 2004; Wicki et al., 2018). LST and NDVI provided by the Moderate Resolution Imaging Spectroradiometer (MODIS) have been utilized as predictors for Tair mapping at 1 km resolution (Yoo et al., 2018; Zhu et al., 2019). Recently, higher spatial resolution satellites such as Landsat 7 and 8 (30-100 m) and Sentinel-2 (10-60 m), which are better-suited for urban applications, have been employed for Tair mapping (Nichol, 2005; Shandas et al., 2019; Wicki et al., 2018). LiDAR remote sensing and building height products characterize the 3D structure of the urban canopy (Venter et al., 2020), which drives the variation of urban Tair by influencing wind flow and solar

radiation trapping (Quanz et al., 2018; Wicki et al., 2018). Building height is increasingly freely available through open-source datasets such as the ALOS World 3D model (Ren et al., 2020; Takaku et al., 2014) and Urban Atlas (European Environment Agency, 2018).

Urban Tair is increasingly mapped by combining Tair measurements and remote sensing data (e.g., (Shandas et al., 2019; Venter et al., 2020; Wicki et al., 2018; Yoo et al., 2018)). Venter *et al.* (2020) categorized available approaches of Tair mapping into three groups of: (1) temperature vegetation index methods (TVX); (2) thermodynamic balance models; and (3) statistical and machine learning (ML) regression models. ML approaches are superior in prediction accuracy to other methods (Kuhn and Johnson, 2013). Moreover, in contrast to parametric approaches, ML algorithms can handle a large set of available predictor variables and provide mechanisms for estimating variable importance (Kuhn and Johnson, 2013). However, very few studies so far have fused crowdsourced Tair data and remote sensing data to map Tair (Seebacher et al., 2019; Venter et al., 2020). Shandas *et al.* (2019) used Random Forest (RF) regression using Sentinel-2 and vehicle-mounted temperature sensors to model Tair spatially. Seebacher *et al.* (2019) predicted UHI occurrences using private weather station network data, land-use and land-cover data from OpenStreetMap, meteorological data, and RF classification. They recommended the addition of other influencing variables in prediction models of UHIs, including the morphology of urban surfaces (Seebacher et al., 2019). Notably, Venter *et al.* (2020) combined Sentinel, Landsat, and LiDAR data with crowdsourced Netatmo data to model annual mean, daily maximum and minimum Tair using RF regression in Oslo, Norway. They found that accuracy in modeling Tair was lowest in summer when Tair variability was higher and recommended that future research focus on mapping in these conditions due to their high societal relevance (Venter et al., 2020). The fusion of crowdsourced Tair data and geodata to model urban Tair is therefore still in its infancy and needs to be applied in new locations and conditions.

In this study, we aim to predict nocturnal urban Tair spatially one day in advance by integrating geodata (Landsat 8 and morphometric indicators), crowdsourced Tair data, and ML methods. Previous studies used satellite data and Tair measurements as an interpolation tool (Shandas et al., 2019; Venter et al., 2020; Wicki et al., 2018), but did not aim to predict Tair distribution in advance. Our study aims at predicting Tair as deviation from a conventional weather station, allowing for the integration of high-resolution mapping with sophisticated forecasting models. In addition, we added crowdsourced Tair as a predictor in addition to 13 GIS predictors, exploring temporal autocorrelation for its predictive potential. Furthermore, this study explores how accurately Tair can be modeled using crowdsourced Tair, which integrates other uncertainties compared to traditional meteorological data, by independently testing with conventional Tair data in addition to "leave-one-date-out" testing. Although RF regression has been applied in previous studies (Shandas et al., 2019; Venter et al., 2020), comparisons of ML

algorithms to map urban Tair are thus far lacking. This study demonstrates this methodology in Berlin, Germany, but can easily be adapted to cities worldwide due to the free availability of the majority of datasets.

We addressed the following research questions: (1) How can open source geodata and crowdsourced Tair data be applied to predict spatial distribution of nocturnal Tair in a mid-latitude metropolitan region? and (2) How is nocturnal Tair distributed in the metropolitan area and how is it affected by urban land cover?

2.2 METHODS

2.2.1 STUDY AREA

Berlin is the largest city in Germany with 3.7 million inhabitants distributed across 891 km^2 in 2018 (Statistical Office of Berlin-Brandenburg, 2019). It is located in eastern Germany (52.52 °N, 13.40 °E) and has a humid warm temperate climate (Cfb) (Kottek et al., 2006). Berlin is covered by 45% green and blue infrastructure, nearly 20% transportation and infrastructural areas and around 35% built-up areas (Dugord et al., 2014).

2.2.2 DATA AND PREPROCESSING

The workflow for this study is shown in Figure 2.1. An overview of the 14 predictors used to model Tair one day in advance is provided in Table 2.2.

Figure 2.1: Flowchart of the study, including preprocessing steps, training using machine learning (ML) models, generation of spatial predictions of Tair, and testing. Abbreviations can be found in the text. Data products are represented as rectangles, whereas processes

(such as data processing or statistical analyses) are represented as ovals. The 14 predictors used in machine learning models are represented as parallelograms. Letters refer to the parts of the "Methods" section which describe steps shown in the flowchart.

2.2.2.1 CROWDSOURCED TAIR DATA

Crowdsourced Tair, referred to as "Tcrowd" hereafter, was obtained from CWS of the private company 'Netatmo' (www.netatmo.com). Each CWS provides instantaneous values at approximately five-minute intervals, which are automatically uploaded to the Netatmo server via Wi-Fi. An automatic workflow to acquire these data at one-hourly intervals is described in Meier *et al.* (2017). In total, data from 2959 Netatmo stations were included in this study (see Figure 2.2, Table 2.2) for four summer dates in the years 2017-2019 (see section "Landsat data" below).

Quality assessment and control is crucial for the application of crowdsourced Tair data (Chapman et al., 2017; Meier et al., 2017; Muller et al., 2015; Napoly et al., 2018). The quality control (QC) procedure developed by Napoly *et al.* (2018) was used in this study to statistically identify implausible measurements of erroneous crowdsourced data due to misplacement of sensors, solar exposition (radiative errors), inconsistent metadata or device malfunctions. This QC procedure needs no reference meteorological data and is easy to apply via an available software R-package "CrowdQC" (Grassmann et al., 2018). The first step (M1) of the QC flags CWS with equivalent longitude and latitude coordinates. The second step (M2) applies a modified z-score approach for the detection of statistical outliers from the hourly Tair distribution. Therefore, as an intermediate step, altitude-corrected Tair was calculated to account for natural variation of Tair due to different elevations of Netatmo stations. Elevation data were derived from the globally available Digital Elevation Model (DEM) derived from the 2000 NASA Shuttle Radar Topography Mission (SRTM) (Jarvis et al., 2008). The third step (M3) removes data for a complete month when more than 20 % of data points during this month are flagged during QC step M2 (per station). The fourth step (M4) excludes any station that, when correlated against median hourly Tcrowd over a month, produces a Pearson's correlation coefficient of less than 0.9. The last QC step applied here (O1) is the linear interpolation of missing individual hourly values at each CWS.

Tcrowd data was temporally averaged between (and including) the hours of 22 UTC and 2 UTC in order to represent nighttime Tair. Averaging of Tcrowd data improves its robustness and further improves data quality by reducing the influence of erroneous data points that could still remain in the data set, even after QC (Napoly et al., 2018). As the nighttime UHI typically requires a few hours to be fully developed (Fenner et al., 2017), three hours after sunset (22 UTC) was selected as the starting time of the averaging interval. As Netatmo data is rounded to the closest hour (e.g., 3:20 UTC is rounded to 3 UTC), 2 UTC was selected as the endpoint to ensure that all

data was before sunrise. Only stations where five out of five observations were available between 22 UTC and 2 UTC were included.
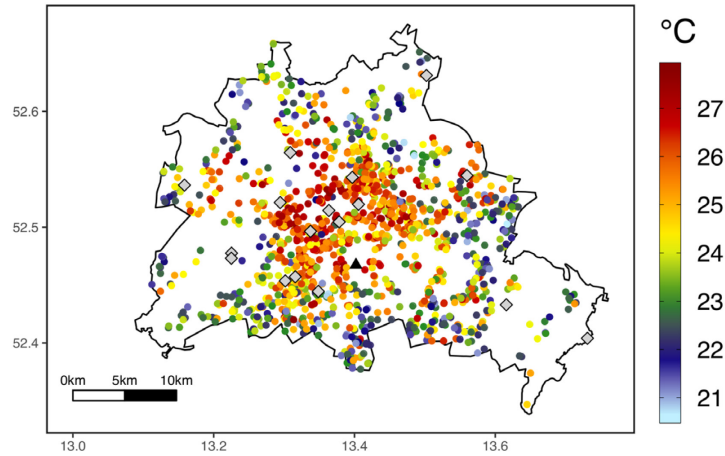
Figure 2.2: Spatial distribution of crowdsourced air temperature (Tcrowd) measured at Netatmo weather stations (colored circles) in Berlin on the night starting on 25 June 2019 (averaged from 22 UTC to 2 UTC), which coincides with the June 2019 heat wave in Germany. The 1338 Netatmo stations shown were available after quality control (QC) and removing Netatmo stations outside of the Berlin city boundary and without data on the previous night (24 June 2019). Symbols partially overlap due to the high spatial density of Tcrowd. The location of reference meteorological stations operated by the German Meteorological Service (DWD) and the Urban Climate Observation Network (UCON) is indicated with gray diamonds. The DWD station Berlin-Tempelhof, which was used to compute Tcrowd deviation, is indicated with a black triangle. The black line denotes the administrative border of Berlin.

After initial preprocessing, the deviation of Tcrowd data was computed. Tair data from the German Meteorological Service (DWD) weather station Berlin-Tempelhof (Station ID: 00433), $T_{Tempelhof}$, was used as a proxy for the "mean" temperature of Berlin. DWD station data were averaged between 22 and 2 UTC, corresponding to the same time as Tcrowd data. Deviation of Tcrowd ($Tcrowd_{dev}$) was then computed as Tcrowd minus $T_{Tempelhof}$ for each night.

2.2.2.2 OTHER METEOROLOGICAL DATA

Meteorological data used for Land Surface Temperature (LST) derivation (Tair (°C), relative humidity (%), and air pressure (hPa)) were acquired from the weather station Berlin-Tempelhof via DWD climate data center (DWD, 2021). Weather data used for LST calculations were collected at 10:00 UTC in order to correspond with the Landsat 8 day scene acquisition time (see Table 2.1). Modeled Tair was independently validated using six DWD (DWD, 2021) stations and eleven stations of the Urban Climate Observation Network (UCON) of the Chair of Climatology at Technische Universität Berlin (Fenner et al., 2014) (see Figure 2.2 and Appendix A Table A.1).

Table 2.1: Meteorological conditions. "LS day" refers to the acquisition time of Landsat imagery. "Night" refers to conditions averaged between 22 UTC and 2 UTC starting on the same day as Landsat acquisition. "Night (+1 day)" refers to conditions averaged between 22 UTC and 2 UTC one day after Landsat acquisition. Data is acquired from DWD weather station Berlin-Tempelhof (Station ID: 00433), except for cloudiness data (acquired from DWD weather station Berlin-Dahlem (FU) (Station ID: 00403)).

Image acquisition date and time (UTC)	Air temperature (°C)			Relative humidity (%)			Wind speed (m/s)			Wind direction (°)			Cloudiness (oktas)		
	LS day	Night	Night (+1 day)	LS day	Night	Night (+1 day)	LS day	Night	Night (+1 day)	LS day	Night	Night (+1 day)	LS day	Night	Night (+1 day)
2 June 2017, 10:02	22.7	13.7	18.3	30.0	56.6	83.2	3.8	3.1	1.6	300	90	154	2.0	2.0	7.6
23 July 2018, 10:01	24.8	19.3	21.4	46.0	66.0	70.2	5.4	1.3	1.5	300	86	70	1.0	0	0.2
24 June 2019, 10:02	26.7	20.6	23.3	31.0	52.2	54.6	4.0	2.8	1.2	100	94	124	4.0	1.8	3.6
26 July 2019, 10:02	27.9	21.8	20.0	41.0	59.2	64.6	4.2	4.7	5.1	60	72	78	1.0	0.6	2.0

2.2.2.3 LANDSAT DATA

Four Landsat 8 (L-8) scenes were selected based on the following criteria: (a) cloud-free; (b) summertime (June-August); and (c) matching Netatmo data availability (2015-present): 2 June 2017, 23 July 2018, 24 June 2019, and 26 July 2019 (Table 2.1).

L-8 Thermal Infrared Sensor (TIRS) data were used to characterize LST. L-8 data were acquired from the United States Geological Survey (USGS) EarthExplorer (U.S. Geological Survey (USGS), 2019a). Land surface emissivity (LSE), a key input variable for estimating LST, was estimated using the Normalized Difference Vegetation Index (NDVI) thresholds method (Sobrino et al., 2004; Wang et al., 2015). For water bodies, the emissivity value of water (0.991) was assigned for the "Water" LULC class provided by Urban Atlas 2012 (Wang et al., 2015). The mono-window algorithm developed by Wang *et al.* (2015) was applied to derive LST from the thermal band (band 10):

$$LST = \frac{a \cdot (1 - C - D) + (b \cdot (1 - C - D) + C + D) \cdot T_{AS} - D \cdot T_a}{C} \tag{1}$$

where T_a is the effective mean atmospheric temperature (K); a and b are the coefficients approximating the derivative of the Planck radiance function for the thermal band; C and D are scene-dependent based on LSE and atmospheric transmittance; and T_{AS} is the effective brightness temperature received at the sensor (Wang et al., 2015).

NDVI was computed as an indicator of photosynthetic activity and vegetation density (Tucker, 1979). NDVI represents the reflection properties of vegetated surfaces in the visible red (RED) and near-infrared (NIR) range:

$$NDVI = \frac{NIR - RED}{NIR + RED} \qquad (2)$$

The Normalized Difference Built-up Index (NDBI) was computed as an indicator of built-up areas, which have a higher reflectance in the shortwave-infrared (SWIR) region (Band 6 for L-8) compared to the near-infrared (NIR) region (Band 5 for L-8) (Zha et al., 2003):

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \qquad (3)$$

Surface broadband albedo (α) was computed based upon the normalized version (Smith, 2010) of the Liang (2001) algorithm, adapted to L-8 data. Five Top of Atmosphere (TOA) reflectance L-8 bands ($\rho2$-7) as provided by the USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) were used for the calculations (U.S. Geological Survey (USGS), 2019b):

$$\alpha = \frac{0.356 \cdot \rho2 + 0.13 \cdot \rho4 + 0.373 \cdot \rho5 + 0.085 \cdot \rho6 + 0.072 \cdot \rho7 - 0.0018}{1.016} \qquad (4)$$

Since Landsat-based predictors vary between study dates and are weather-dependent, their spatial deviation was used to train and test models. Landsat deviation (Landsat$_{dev}$) was calculated for all different Landsat data (LST, NDVI, NDBI, albedo) as follows:

$$Landsat_{dev} = Landsat - Landsat_{mean} \qquad (5)$$

where Landsat$_{mean}$ represents the spatial mean value of the respective variable within Berlin.

2.2.2.4 OTHER GEODATA

Building height (BH) and land use and land cover (LULC) classes data from 2012 were provided by the Urban Atlas (European Environment Agency, 2018). Urban Atlas 2012 datasets are freely available for 785 Functional Urban Areas in 39 countries, allowing for direct transferability of this methodology. Vegetation height (VH) was provided by the Berlin Digital Environmental Atlas (Berlin Senate Department for Urban Development and Housing, 2014). Impervious surface fraction (ISF) was assigned based on LULC classes from Urban Atlas data. Impervious coefficients were assigned to each LULC class following Jennings *et al.* (2004) and the degree of soil sealing indicated by LULC classes.

Urban morphology parameters, which describe the spatial structure of the urban environment, were calculated using the Urban Multi-scale Environmental Predictor (UMEP) QGIS plug-in (Lindberg et al., 2018) and BH. The Sky View Factor (SVF) represents the proportion of the observed sky divided by the total upper hemisphere. A 100 x 100m cell grid (as

in URBANFLUXES, Marconcini *et al.* (2017)) and BH were used to compute several parameters including the plan area index (PAI) (fraction of roof area), the frontal area index (FAI) (area of building walls normal to wind direction relative to the total ground area), the standard deviation of BH (zHstd), roughness length (z0) (the height at which wind speed theoretically becomes zero in the log wind profile), and zero-plane displacement length (zd) (height above ground where the wind speed is zero as a results of obstacles to the flow) (Grimmond and Oke, 1999; Lindberg et al., 2018). The roughness parameters (z0 and zd) were calculated following the method of Kanda *et al.* (2013). The parameters used here are isotropic (not depending on a specific wind direction).

2.2.3 FOCAL BUFFER ANALYSIS

Land surface parameters influence Tair with a distance-decay effect, which can be accounted for using spatial aggregation (Shandas et al., 2019; Venter et al., 2020; Wicki et al., 2018). Therefore, a moving window average of varying spatial distances (referred to as a 'focal buffer') was computed for all 13 GIS predictors (see Fig. 1(c)). Focal buffers with the following circular radii (0 m, 50 m, 100 m, 200 m, 300 m, 500 m, 750 m, 1000 m, 1500 m), coregistered to Landsat resolution (30 x 30 m), were created. The optimal radii were determined based on Spearman's rho and are shown in Table 2.1. Further details on focal buffer analyses are given in Appendix A (Table A.5; Figure A.1). Focal buffer values were extracted at the location of the CWS.

2.2.4 MACHINE LEARNING MODELING

2.2.4.1 DATA COMPILATION AND MODELING APPROACH

A time lag of one day was selected for the prediction as longer time lags are associated with greater uncertainty based on weather forecasting (Zhang et al., 2019). The dates on which Landsat data were available are referred to as the predictor dates, as the Tcrowd data from these dates were used as input for prediction (one of the independent variables). The outcome dates are one day after the corresponding predictor dates, referring to the Tcrowd values being predicted in the models (the dependent variable). For instance, 2 June 2017 is a predictor date and 3 June 2017 is an outcome date. Tcrowd data from a predictor and outcome date pair (such 2 and 3 June 2017) are matched based upon the station ID. For model tuning and training, three out of the four study dates were used ("leave-one-date-out"), meaning that training was conducted with three predictor and outcome date pairs. The remaining study date (one predictor and outcome date pair) was used as an independent test subset to assess model performance. For example, training was conducted with Tcrowd data from 2-3 June 2017, 23-24 July 2018, and 24-25 June 2019, and this trained model was applied to data from 26 July 2019 (Tcrowd and 13 GIS predictors) to predict Tair on 27 July 2019. CWS without data on both the paired predictor and outcome dates and outside of the Berlin city boundary were excluded from training and testing datasets (see Figure 2.2). After

preprocessing, 1130 observations for 2-3 June 2017, 1258 observations for 23-24 July 2018, 1338 observations for 24-25 June 2019 (Figure 2.2), and 1387 observations for 26-27 July 2019 remained. "GIS predictors" refers to all 13 remote sensing and GIS predictors (without Tcrowd). The following combinations of predictors were tested in models: (a) Tcrowd and all 13 GIS predictors (all); (b) only the 13 GIS predictors (GIS); and (c) only Tcrowd as a predictor (Tcrowd) (Table 2.2). Models were additionally run without VH, a dataset which is not available in all cities worldwide, in order to demonstrate the applicability of our method using only open-source data with worldwide (or at least European) coverage. Comparison of model performance with and without VH is given in Appendix A (Figure A.2 and Figure A.3).

Table 2.2: Overview of the predictor variables used to model Tair one day in advance. Optimal focal buffer radii were identified using Spearman's rho. Landsat 8 predictors are depicted for the exemplary scene 23 July 2018, 10:01 UTC. Tcrowd is shown for the night of 23 July 2018 (averaged between 22 UTC and 2 UTC). GIS predictors are depicted with a spatial resolution of 30m x 30m, not as focal buffers. Landsat predictors and Tcrowd are shown with true values (not deviation). Citation numbering refers to Vulova et al. (2020).

Predictor	Data source	Preprocessing method	Original spatial resolution [m]	Optimal radius [m]	
Albedo	Landsat 8	Liang [37]; Smith [36]	30	200	
Building height (BH)	Urban Atlas 2012	-	10	300	
Crowdsourced Tair (Tcrowd)	Netatmo	Napoly *et al.* [19]; Grassmann *et al.* [28]; Meier *et al.* [18]	-	-	
Frontal area index (FAI)	Urban Atlas 2012 (BH)	Lindberg *et al.* [42]	100	200	
Impervious surface fraction (ISF)	Urban Atlas 2012 (LULC)	Jennings *et al.* [41]	10	1500	
Land Surface Temperature (LST)	Landsat 8	Wang *et al.* [33]	100	500	
Normalized Difference Built-up Index (NDBI)	Landsat 8	Zha *et al.* [35]	30	500	
Normalized Difference Vegetation Index (NDVI)	Landsat 8	Tucker [34]	30	500	
Plan area index (PAI)	Urban Atlas 2012 (BH)	Lindberg *et al.* [42]	100	1500	
Roughness length (z0)	Urban Atlas 2012 (BH)	Lindberg *et al.* [42]; Kanda *et al.* [45]	100	750	
Sky View Factor (SVF)	Urban Atlas 2012 (BH)	Lindberg *et al.* [42]	10	300	
Standard deviation of building heights (zHstd)	Urban Atlas 2012 (BH)	Lindberg *et al.* [42]	100	500	
Vegetation height (VH)	Berlin Digital Environmental Atlas	-	5	50	
Zero-plane displacement length (zd)	Urban Atlas 2012 (BH)	Lindberg *et al.* [42]; Kanda *et al.* [45]	100	300	

2.2.4.2 MACHINE LEARNING ALGORITHMS

We tested three supervised regression ML algorithms: Random Forest (RF), Stochastic Gradient Boosting (GBM), and Model Averaged Neural Network (avNNet). Detailed descriptions of the ML algorithms and the 'caret' package implementations are provided in Kuhn and Johnson (Kuhn and Johnson, 2013). The root mean square error (RMSE) was used to select the optimal hyperparameters during 10-fold cross-validation. With 10-fold cross-validation, the training samples are randomly partitioned into ten equally sized folds, with nine folds used for training and one fold used for validation (Kuhn and Johnson, 2013). This procedure is repeated with each of the ten folds used for validation once (Kuhn and Johnson, 2013). Tested hyperparameters during tuning were based upon recommendations from Kuhn and Johnson (Kuhn and Johnson, 2013). Optimal hyperparameters determined by this procedure are given in the Appendix A (Tables A.2 – A.4).

RF is an ensemble machine learning technique commonly applied in remote sensing (Shandas et al., 2019; Venter et al., 2020; Yoo et al., 2018; Zhu et al., 2019). RF generates predictions by creating a large number of decision trees based on randomly selected subsets through bootstrapping from both training samples and predictor variables at each tree node (Kuhn and Johnson, 2013). The number of predictor variables randomly selected at each split ('mtry') was tuned for each value between two and the number of predictor variables (Kuhn and Johnson, 2013). RF models with one predictor (Tcrowd) were not tuned, with 'mtry' kept constant at a default value of 2. The number of trees ('ntree') was set to the default 'caret' value of 500, as tuning 'ntree' is shown to be unnecessary and the largest performance gain is generally achieved with the first 100 trees (Probst and Boulesteix, 2018).

GBM incorporates the strengths of regression trees and boosting in a single algorithm (Kuhn and Johnson, 2013). With boosting, new trees are generated based upon previous trees by fitting the model to the residuals from the prior step (Kuhn and Johnson, 2013). The depth of trees ('interaction.depth') was tuned from 1 to 11 with increments of 2 (Kuhn and Johnson, 2013). Three values of the learning rate ('shrinkage') were evaluated during tuning (0.01, 0.05, and 0.1) (Kuhn and Johnson, 2013). The number of trees ('n.trees') was tuned from 100 to 1000 with increments of 50 (Kuhn and Johnson, 2013). The minimum observations in terminal nodes ('n.minobsinnode') was tuned using three values (6, 8, and 10).

Neural networks are powerful nonlinear regression methods which model the outcome by an intermediary set of unobserved variables (hidden units) (Kuhn and Johnson, 2013). To generate avNNet predictions, the 'caret' package fits the same neural network model five times using different random seeds and averages the output from each network (Kuhn and Johnson, 2013). Prior to modeling with avNNet, the predictors were centered and scaled (Kuhn and Johnson, 2013). Weight decay was tuned from 0 to 0.1 with increments of 0.02 (Kuhn and Johnson, 2013). The number of hidden units ('size') was tuned from 2 to 14 (the maximum number of predictors) with increments of 2 (Kuhn and Johnson, 2013). Relative variable importance scores (scaled from 0 to 100%) were extracted from models with only GIS predictors (Kuhn and Johnson, 2013).

2.2.4.3 SPATIAL PREDICTIONS

Trained models were used to predict Tair on the testing dates. In order to generate spatial predictions, a spatially continuous raster of all predictors was necessary. Thus, ordinary kriging was performed on $Tcrowd_{dev}$ using a 30m x 30m grid using the R package 'automap' (Hiemstra et al., 2009). CWS outside of the Berlin city boundary were not excluded in order to enhance accuracy on the city outskirts. Spatial predictions were converted from $Tcrowd_{dev}$ to true Tair by adding $T_{Tempelhof}$ for the corresponding night (Table 2.1, Figure 2.1).

2.2.5 Testing

Testing$_{crowd}$ refers to applying trained models to the date left out of training and comparing the predictions to measured Tcrowd. Additionally, we compared observed Tair (T_{obs}) of 17 reference weather stations (see Figure 2.2) and modeled Tair (T_{mod}) at the pixel matching the geographic coordinates of the stations (see Figure 2.2 and Appendix A Table A.1), which is referred to as testing$_{ref}$. Reference Tair data is essential in order to assess the bias and inaccuracy which may be incorporated by the use of Tcrowd (instead of conventional Tair data) as a predictor (Fenner et al., 2017; Meier et al., 2017). Both testing$_{crowd}$ and testing$_{ref}$ were evaluated with RMSE, mean absolute error (MAE), percent bias (pbias), and coefficient of determination (R^2). R^2 is included as a reference, as similar studies (e.g., (Venter et al., 2020; Yoo et al., 2018; Zhu et al., 2019)) include this statistic. However, R^2 is an inadequate measure to assess the performance of nonlinear models (Spiess and Neumeyer, 2010) and is highly influenced by the variation in the dependent variable (Kuhn and Johnson, 2013). Thus, the other performance metrics were emphasized in evaluating model performance.

2.2.6 LULC Analysis

A single spatial prediction representing a typical heat wave was selected for analysis of the variation of T_{mod} between LULC classes and different distances from the geometric city center: 24 July 2018 using RF and all predictors. LULC classes comprising less than 5% of the Berlin area were omitted from analysis, leaving 9 LULC classes. The following buffer classes from the city center were analyzed: 0-5 km, >5-10 km, >10-20 km, and >20-30 km. Kruskall-Wallis tests and Dunn's post-hoc tests (with p-values adjusted with the Benjamini-Hochberg method (Benjamini and Hochberg, 1995)) were applied to assess the significance of differences in Tair between LULC classes.

An analysis of the relationship between T_{mod} and the GIS predictors was conducted in order to quantify how nocturnal Tair is affected by urban land cover. The relationship between T_{mod} using RF and all predictors was analyzed using Spearman rank correlation coefficient (Spearman's rho) (see Appendix A; Figure A.4). The GIS predictors without focal buffers and coregistered to Landsat (30-m) resolution were used for this analysis.

If not mentioned otherwise, all modeling and analysis were completed using R version 3.6.2 (2019-12-12) and RStudio (version 1.2.5033) (R Core Team, 2020). The 'caret' package for R was used as a wrapper package for the ML algorithm implementations in R (Kuhn, 2008). Parallel processing was performed on seven cores using the R package 'doSNOW' (Microsoft Corporation and Weston, 2019). Processing time of model training was measured using the R package 'tictoc' (Izrailev, 2014). Benchmarking results are given in Appendix A (Table A.10).

2.3 RESULTS

2.3.1 AIR TEMPERATURE MODELING

Using all predictors (Tcrowd and 13 GIS predictors) or only GIS predictors in all models, RMSE is <1.3 K for testing$_{crowd}$ and <2.4 K for testing$_{ref}$ (Table 2.3). Based on testing$_{crowd}$, RF is the best-performing algorithm with an average RMSE of 1.23 K across the three predictor sets, closely followed by GBM (RMSE = 1.24 K) and avNNet (RMSE = 1.26 K). Testing$_{crowd}$ reveals that models with GIS predictors show on average the lowest RMSE (1.22 K). However, all predictors most frequently show the lowest RMSE for testing$_{crowd}$ for each date and ML algorithm (in 6 out of 12 cases) (see Appendix A; Tables A.6-A.9). According to RMSE for testing$_{crowd}$, all predictors performed best in 2018 and July 2019, whereas GIS predictors and Tcrowd performed best for 2017 and June 2019, respectively. The lowest RMSE for testing$_{crowd}$ averaged for all four study dates is achieved using RF and GIS predictors (RMSE = 1.16 K).

Testing$_{ref}$ confirms RF as the best algorithm, with an average RMSE of 2.21 K across the three predictor sets, followed by avNNet (RMSE = 2.24 K) (Table 2.3). Although RF shows the lowest average RMSE for testing$_{ref}$, avNNet most frequently shows the lowest RMSE for testing$_{ref}$ for each date and predictor set (see Appendix A; Tables A.6-A.9). However, ML algorithms generally show a similar performance for both testing$_{crowd}$ and testing$_{ref}$. On average, GIS predictors outperform other predictor sets based on RMSE for testing$_{ref}$ (RMSE = 1.99 K). GIS predictors also most frequently have the lowest RMSE for testing$_{ref}$ for each date and ML algorithm (in 9 out of 12 cases), with other predictor combinations showing lower RMSE only on 27 July 2019. The lowest RMSE for testing$_{ref}$ averaged for all study dates is found using avNNet and GIS (RMSE = 1.95 K).

Table 2.3: "Leave-one-date-out" testing performance metrics averaged across all test dates. The performance metrics are root mean square error (RMSE), mean absolute error (MAE), percent bias (pbias), and coefficient of determination (R^2). The best performance metrics for testing$_{crowd}$ and testing$_{ref}$ are shown in bold.

Predictors	Model	RMSE (K)		MAE (K)		pbias (%)		R^2 (-)	
		Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$
All	avNNet	1.26	2.27	1.08	1.98	35.85	9.08	0.667	0.515
	GBM	1.23	2.33	1.06	2.07	33.77	9.65	0.694	0.496
	RF	1.23	2.31	1.06	2.04	34.85	9.85	0.710	**0.631**
GIS	avNNet	1.24	**1.95**	1.04	**1.64**	**15.58**	**7.20**	0.372	0.578
	GBM	1.25	2.04	1.04	1.71	16.05	7.27	0.435	0.465
	RF	**1.16**	1.97	**0.96**	1.67	15.88	7.35	0.512	0.581
Tcrowd	avNNet	1.27	2.51	1.12	2.23	39.15	10.70	**0.769**	0.489
	GBM	1.25	2.50	1.10	2.22	36.48	10.75	0.761	0.503
	RF	1.30	2.35	1.13	2.06	40.13	9.90	0.714	0.501

Testing$_{ref}$ indicates an overestimation of Tair based upon pbias (Table 2.3). The models' inability to predict lower Tair values can be seen in Figure 2.3 and Figure 2.4, showing results for 24 July 2018. Generally, Figure 2.3 shows underestimation of Tair for testing$_{crowd}$, especially for

high Tair, and overestimation for testing$_{ref}$. Average positive pbias for testing$_{crowd}$ is driven by the 2017 study date, which was cooler than the other nights (see Appendix A; Table A.6). T$_{mod}$ underestimates the true Tair range on 24 July 2018 due to the influence of the July 2019 date in training. On 27 July 2019, Tcrowd was spatially homogeneous, with a range and standard deviation of 4.15 K and 0.94 K, respectively. In contrast, on 24 July 2018, Tcrowd had a range of 6.25 K and a standard deviation of 1.54 K.
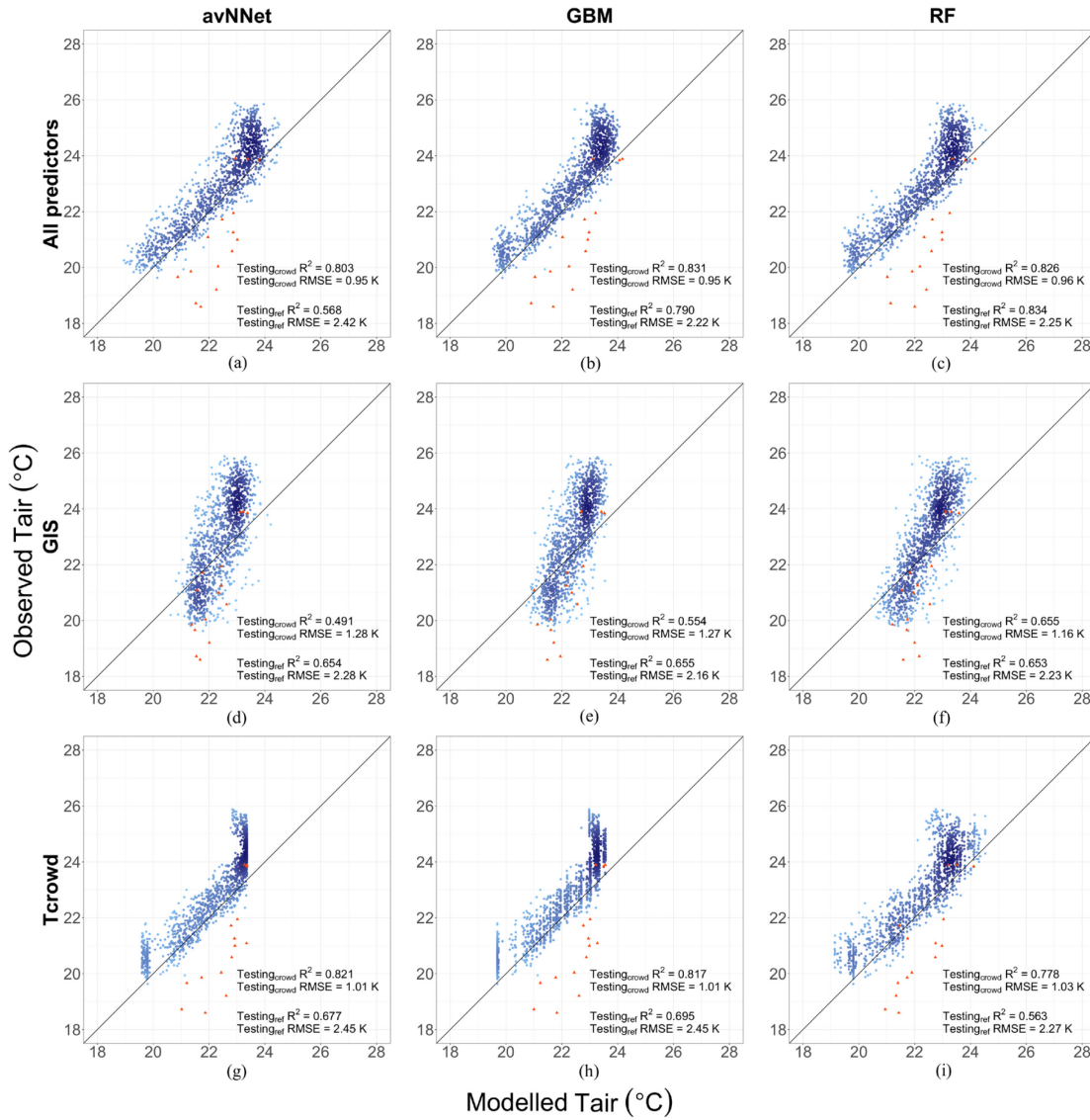
Figure 2.3: Scatterplots for a single study date (24 July 2018) showing the correlations between modelled Tair (T$_{mod}$) and observed Tair for nine models: (a) all predictors with avNNet, (b) all predictors with GBM, (c) all predictors with RF, (d) GIS predictors with avNNet, (e) GIS predictors with GBM, (f) GIS predictors with RF, (g) Tcrowd predictor with avNNet, (h) Tcrowd predictor with GBM, and (i) Tcrowd predictor with RF. The intensity of the blue color represents the testing$_{crowd}$ data point density. Red triangles represent testing$_{ref}$ data points.

All spatial predictions show a typical Tair spatial gradient (warmer to cooler) from the city center into the outskirts (Figure 2.4). Though T_{mod} is well-validated in the inner city, the lower Tair in the city outskirts are consistently overestimated for all three predictor sets (Figure 2.4). Spatial predictions based on a single Tcrowd kriging product did not fully capture the fine-scale variability of Tair. The spatial prediction based on only GIS predictors better reflect urban morphology and land cover but do not predict lower Tair in the north and northeast of Berlin (Figure 2.4). For 2018, the optimal RF model based on testing$_{crowd}$ (RMSE = 0.96 K) incorporates all predictors, reflecting the mixed influence of larger-scale meteorological influences and urban structure. Testing$_{ref}$ RMSE for RF is the lowest based on GIS predictors (2.23 K), demonstrating the utility of GIS predictors to predict Tair in new locations.
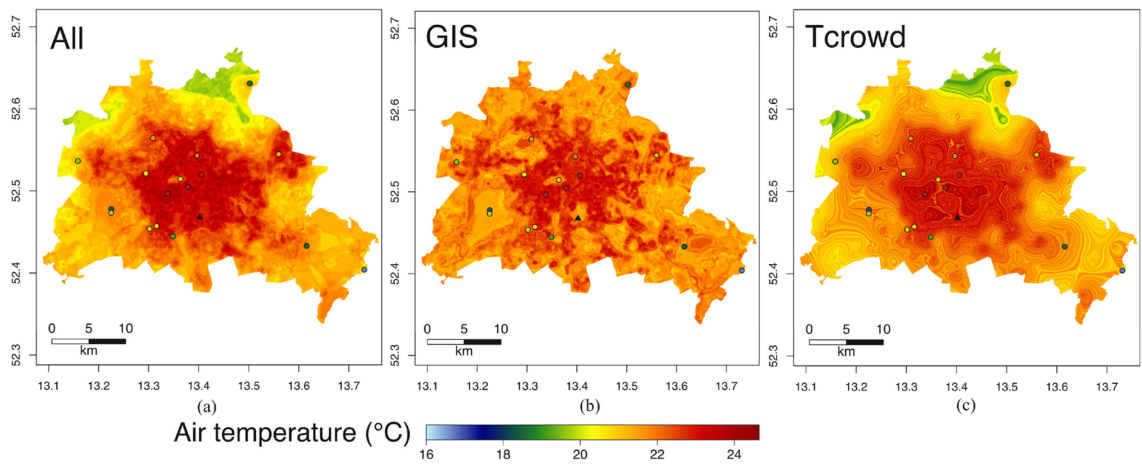
Figure 2.4: Visualization of Tair predictions of the three combinations of predictors for the exemplary date of 24 July 2018 using RF regression with (a) all, (b) GIS, and (c) Tcrowd as predictors. Points with a black outline represent UCON/DWD reference stations. The Tempelhof station is marked with a black triangle. The projection and datum are WGS-84.

2.3.2 VARIABLE IMPORTANCE

Figure 2.5 shows the percentage of relative influence of GIS predictors in models. A large variability of importance between algorithms is apparent. Morphometric parameters are particularly influential. On average, zd is the most important predictor (67.7% relative influence), followed by zHstd (62.4%), and albedo (54.8%). The least important predictors are, on average, NDBI (29.9%), BH (34.0%), and ISF (34.0%).

The importance of GIS predictors is highly affected by the ML algorithm employed. Figure 2.5 shows that the variable importance for the same parameter can differ greatly between the three ML algorithms. Zd is the most important parameter on average for avNNet, followed by BH and SVF. VH is the least influential predictor for avNNet, followed by LST and PAI. In contrast, for RF, on average, VH is the most important GIS predictor, followed by albedo and PAI. The least

important predictors for RF are BH, SVF, and zd. For GBM, the most important predictors are zd, zHstd, and albedo, and least important predictors are BH, PAI, and NDBI.

Variable importance also differs based on the test date excluded. The exclusion of two dates is analyzed as an example: 23 July 2018 (a spatially variable night) and 26 July 2019 (a spatially homogeneous night). When July 2019 is excluded from training, the most important predictors are zHstd, zd, and z0 and the least important are LST, ISF, and NDBI. With July 2018 excluded from training, the most important predictors are zd, zHstd, and albedo and the least important are BH, VH, and z0. However, the algorithm employed has a greater effect on variable importance than the dates used in training. While this analysis of variable importance focuses only on models with GIS predictors, Tcrowd is the most important predictor in all cases for models with all predictors, emphasizing the high autocorrelation of Tcrowd 24 hours apart.
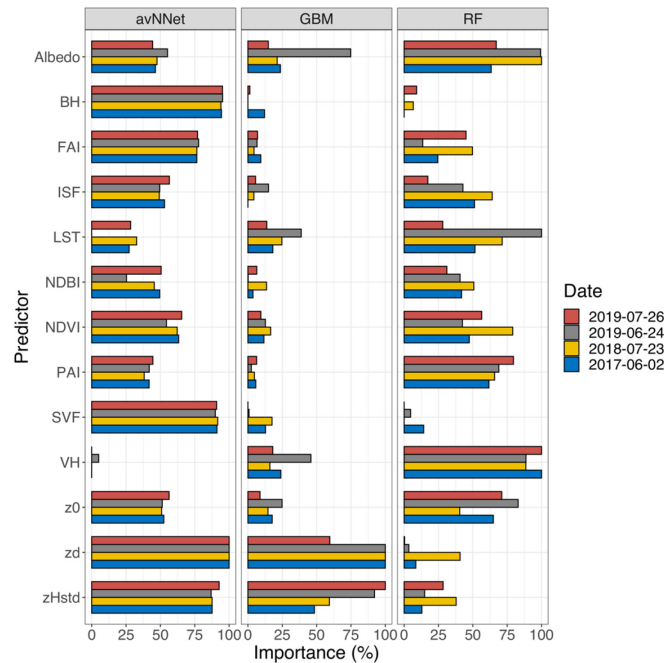
Figure 2.5: Relative variable importance revealed by models with only GIS predictors. Date refers to the test date excluded in model training.

2.3.3 Tair Distribution and LULC Effects

Mean Tair for each LULC class using the spatial prediction for 24 July 2018 with RF and all predictors is shown in Table 2.4. Greater distance to the city center is associated with lower Tair for nearly every LULC class. The only exceptions are noted for the "Forests" and "Water" classes, where Tair at >20-30 km is slightly higher than at >10-20 km. Overall, more pronounced differences are found between buffer sizes than between LULC classes. Testing within each LULC class, there are significant differences between all buffer sizes for all LULC classes (Dunn's test, adjusted p-values < 0.05).

LULC classes with predominantly sealed surfaces are generally warmer than other LULC classes. The warmest LULC class is "Continuous urban fabric" at 0-5 km, >5-10 km, and >10-20 km. At >20-30 km, "Forests" is the warmest LULC class. The second-warmest LULC classes are "Discontinuous dense urban fabric" at 0-5 km and >10-20 km, "Other roads and associated land" at >5-10 km, and "Water" at >20-30 km.

Vegetated LULC classes are generally cooler. The coolest LULC classes are "Green urban areas" at 0-5 km, "Forests" at >5-10 km, "Water" and "Forests" at >10-20 km, and "Sports and leisure facilities" at >20-30 km. The second-coolest LULC classes are "Sports and leisure facilities" at 0-5 km, >5-10 km, and >10-20 km and "Green urban areas" at >20-30 km. "Green urban areas" are defined as public green areas for primarily recreational use, including gardens, parks, and zoos (European Union, 2016). "Sports and leisure facilities" is also a largely vegetated LULC class, as it includes public arenas for sports and their associated green areas (European Union, 2016). Insignificant differences between LULC classes are given in Appendix A.

Table 2.4: Mean Tair ± standard deviation (SD) for nine LULC classes in Berlin at buffers representing different distances (in km) from the geometric city center. The LULC analysis was conducted using the spatial prediction for 24 July 2018 (averaged 22-2 UTC) with RF and all predictors. Spatial SD was calculated across all pixels per LULC class. Sealing level (S.L.) is indicated for urban fabric LULC classes. Blue: mean of 21-21.5 °C, green: mean of >21.5-22 °C, yellow: mean of >22-22.5 °C, orange: mean of >22.5-23 °C, red: mean of >23-23.5 °C.

LULC class [code]	Tair ± SD (°C)			
	0-5 km	>5-10 km	>10-20 km	>20-30 km
Continuous urban fabric (S.L.: >80%) [11100]	23.46 ± 0.30	23.28 ± 0.50	21.93 ± 0.81	21.22 ± 0.40
Discontinuous dense urban fabric (S.L.: 50% - 80%) [11210]	23.44 ± 0.34	22.54 ± 0.75	21.62 ± 0.71	21.14 ± 0.30
Discontinuous medium density urban fabric (S.L.: 30% - 50%) [11220]	23.22 ± 0.30	22.33 ± 0.63	21.37 ± 0.64	21.31 ± 0.35
Forests [31000]	-	21.61 ± 0.33	21.19 ± 0.58	21.40 ± 0.28
Green urban areas [14100]	22.96 ± 0.47	22.34 ± 0.56	21.55 ± 0.71	21.10 ± 0.30
Industrial, commercial, public, military and private units [12100]	23.31 ± 0.31	22.52 ± 0.72	21.59 ± 0.77	21.25 ± 0.42
Other roads and associated land [12220]	23.39 ± 0.34	22.55 ± 0.76	21.54 ± 0.76	21.20 ± 0.30
Sports and leisure facilities [14200]	23.10 ± 0.30	22.27 ± 0.62	21.25 ± 0.75	21.02 ± 0.48
Water [50000]	23.32 ± 0.28	22.54 ± 0.55	21.19 ± 0.49	21.33 ± 0.36

The modeled Tair distribution using RF models with all predictors (see Figure 2.4 and Table 2.4) emphasizes the cooling influence of green spaces in Berlin for all study dates. The cooling effect of both larger (>1 km^2) and smaller (<1 km^2) green spaces is visible in T_{mod}, with forests, public parks, and cemeteries all associated with lower Tair. Aside from the hot spot of higher Tair in the city center, the central eastern part of Berlin, which is composed of high-rise multi-story houses, is a pronounced hot spot. This area of Berlin was identified as having a high heat stress risk based upon the proportion of elderly residents and simulated Tair at 6 AM (Dugord et al., 2014). However, the cooling influence of the Erholungspark Marzahn (~1 km^2) on this borough is evident during the extreme heat wave event in 2019, demonstrating the beneficial cooling effect of urban green spaces.

2.4 DISCUSSION

With increasing extreme heat risks for urban residents (Huang et al., 2019), the prediction of hotspots at a high resolution is vital. Using the example of the city of Berlin, we have demonstrated the capacity of fusing crowdsourced Tair data, Landsat data, and other open-source geodata to spatially predict summertime nighttime Tair one day in advance at a 30-m resolution. Using the optimal algorithm (RF) and all predictors, testing$_{crowd}$ and testing$_{ref}$ show an average RMSE of 1.23 K and 2.31 K, respectively. Based on RF and only GIS predictors, accuracy is even higher, with an average RMSE for testing$_{crowd}$ of 1.16 K and testing$_{ref}$ 1.97 K. RF performance is lowest with only Tcrowd as a predictor, with an average RMSE for testing$_{crowd}$ of 1.30 K and RMSE for testing$_{ref}$ of 2.35 K. Our results are comparable to Venter *et al.* (2020)'s accuracy in modeling daily maximum (RMSE = 1.85 K) and minimum Tair (RMSE = 1.46 K) using Netatmo data, particularly as they showed that accuracy decreased (approaching 2 K) when modeling for summertime.

This study investigated the suitability of three different ML algorithms for their applicability in modeling Tair. Based on both testing$_{crowd}$ and testing$_{ref}$ performance, RF regression proves to be the most suitable algorithm. The second-best algorithm is avNNet, which most frequently had the lowest RMSE for testing$_{ref}$ compared to other algorithms. RF's comparatively faster computation time with tuning (up to three times faster than avNNet) is a clear advantage when processing large datasets (see Appendix A; Table A.10). However, for model training with optimal hyperparameters, RF is the slowest algorithm (twice as slow as avNNet).

We demonstrated that combining geodata and crowdsourced Tair in a ML model is advantageous. Due to high temporal autocorrelation, Tcrowd is useful in predicting its own spatial distribution one day in the future during relatively stable weather conditions. However, Netatmo stations are unevenly distributed throughout the city (see Table 2.2, Figure 2.2), concentrated in built-up central areas, while natural landscapes are not covered (Fenner et al., 2017; Meier et al.,

2017). GIS predictors facilitate more accurate predictions in the gaps between CWS, reflecting the spatial structure and other properties of the urban landscape. The appropriate predictors may depend on atmospheric conditions that favor a certain level of spatial variability of Tair. For prediction on a relatively spatially invariable date (e.g., 27 July 2019; Tcrowd range = 4 K), GIS predictors alone overestimate variation, with lowest testing$_{ref}$ performance for GIS predictors (RF RMSE = 1.05 K) and highest performance for Tcrowd only (RF RMSE = 0.43 K). In contrast, for a date with higher spatial Tair variability (e.g., 25 June 2019; Tcrowd range = 7 K), Tair is best modeled with only GIS predictors (RF RMSE = 2.75 K) according to testing$_{ref}$. Of the four study dates, 27 July 2019 represents the night with highest wind speed (5.1 m/s) and a more homogenous Tair distribution. Venter *et al.* (2020) showed that mapping on hotter, calm days is more challenging than on windy days, leading to overestimation of Tair. This observation is also reflected in our results, where RF with all predictors for 27 July 2019 shows a RMSE for testing$_{ref}$ of 0.47 K, in contrast with an average RMSE of 2.31 K for all study dates. Overall, we recommend the approach of combining GIS predictors with crowdsourced Tair data in order to harness the advantages of both: fine-scale urban structure (GIS) and Tair distribution from a prior time step (Tcrowd).

An analysis of the relative variable importance of GIS-only models highlighted the importance of morphometric predictors (particularly zd and zHstd) and albedo (see Figure 2.5). Variable importance differs based on the algorithm applied (Kuhn and Johnson, 2013); for RF models, for instance, the most important predictors were VH, albedo, and PAI. Wicki *et al.* (2018) and Venter *et al.* (2020) also found morphometric parameters, especially z0, zd (Wicki et al., 2018), and zHstd (Venter et al., 2020), to be useful predictors for modeling urban Tair. A study in Berlin found that nighttime Tair is influenced by surface structure, such as building surface fraction and SVF, and distance to open green spaces (Quanz et al., 2018). Considering that BH is open-source for hundreds of cities and morphometric parameters can easily be computed with open-source tools (e.g., Lindberg et al. (2018)), their integration into models is highly recommended. Acquiring cloud-free Landsat images greatly constrained the study dates in this study. One solution to expand the archive is to integrate Sentinel-2 images, which have been found to be well-suited to modeling Tair (Shandas et al., 2019; Venter et al., 2020). Additionally, model runs without VH did not significantly differ in performance from those with VH, confirming that our methodology is feasible using only open-source parameters with worldwide coverage (see Appendix A; Figure A.2 and Figure A.3).

Focal buffers optimize the performance of GIS parameters for climatological modeling, accounting for the distance of influence of landscape features on Tair (Shandas et al., 2019; Venter et al., 2020; Wicki et al., 2018). Therefore, we provided a recommendation for optimal focal buffer radii for each of the 13 GIS predictors (Table 2.2) for other urban climatology studies.

Focal buffers warrant further investigation in other cities and on other dates, as weather conditions were found to influence the optimal radius length (see Appendix A; Figure A.1).

Testing with an independent dataset can best approximate the true capacity of modeling using crowdsourced data. The lower accuracy indicated by testing$_{ref}$ indicates that assessing the performance of modeling by simply leaving out crowdsourced data from training (as in Venter *et al.* (2020)) is insufficient. Despite extensive QC procedures (e.g., (Meier et al., 2017; Napoly et al., 2018; Nipen et al., 2020)), crowdsourced data include biases and inaccuracies compared to high-quality meteorological observations which affect model predictions. Netatmo stations are mainly installed near building walls and in built-up areas (Fenner et al., 2017; Meier et al., 2017), which causes microscale effects on Tair. Meier *et al.* (2017), Fenner *et al.* (2017), and Napoly *et al.* (2018) found a consistent positive deviation between Tcrowd and reference Tair in Berlin, with the highest deviation occurring during summer and nighttime.

For urban planning applications, it is crucial to understand where Tair predictions are most accurate. Accuracy decreases with distance to the city center based on testing$_{ref}$ (see Appendix A; Table A.11). Whereas at 0-5 km distance from the city center RMSE for testing$_{ref}$ is 1.58 K on average, buffers of 5-10 km, 10-20 km, and 20-30 km show an average RMSE for testing$_{ref}$ of 2.70 K, 2.83 K, and 3.96 K, respectively. Overestimation is also lowest 0-5 km from the city center (pbias = 3.5%). The highest accuracy is found for urban fabric with a 50% to 100% sealing level (average RMSE for testing$_{ref}$ = 1.76 K), whereas the lowest accuracy is found for meadows (average RMSE for testing$_{ref}$ = 2.93 K) (see Appendix A; Table A.12). Urban fabric with a sealing level of 10% to 50% and forests show an average testing$_{ref}$ RMSE of 2.64 K and 2.79 K, respectively. Urban fabric with a higher sealing level also shows the lowest percent bias (5.9%). A study modeling Tair in Oslo using Netatmo Tair and Landsat data also found that the highest errors occurred in Local Climate Zones (LCZs) (Stewart and Oke, 2012) with vegetation cover (Venter et al., 2020).

A spatial analysis of a summer night during the 2018 heat wave shows significant differences in Tair between different distances from the city center and between LULC classes. Public green areas and other open green spaces are among the coolest LULC classes during nighttime. Considering the greater overestimation of modeled Tair for vegetated areas further away from the city center, Tair varies even more drastically than shown by modeled results. Reference weather stations show Tair as low as 16.7 °C (Kaniswall, class "Pastures", 20-30 km from city center) and as high as 23.9 °C (Dessauer, class "Discontinuous dense urban fabric", 0-5 km from city center). A study in London also emphasized the cooling effect of urban green spaces during calm warm nights (Vaz Monteiro et al., 2016). Furthermore, open green spaces are associated with greater cooling for nocturnal Tair than areas with a high proportion of trees, as the reduction in SVF by the canopy impedes radiation loss (Spronken-Smith and Oke, 1999; Vaz Monteiro et al., 2016).

However, trees also contribute to cooling of nighttime Tair (Vaz Monteiro et al., 2016) and are critical for cooling daytime Tair through shade and evapotranspiration (Spronken-Smith and Oke, 1999). Urban planners should consider the benefit of preserving and expanding open green spaces in mid-latitude cities, such as Tempelhofer Feld in Berlin, in light of their cooling effect on warm summer nights (Quanz et al., 2018), while also maintaining urban tree cover for its stronger daytime cooling effect.

Using daytime LST as a predictor may contribute to Tair overestimation in urban green spaces in addition to geographic clustering of Netatmo stations. We found a positive association between daytime LST and nighttime Tair. However, daytime LST indicates that forests are cooler than meadows (Vulova and Kleinschmit, 2019), whereas the opposite is true for nighttime Tair (Nichol, 2005; Spronken-Smith and Oke, 1999; Vaz Monteiro et al., 2016). Nighttime LST may be a more appropriate indicator for nighttime Tair (Nichol, 2005; Yoo et al., 2018), but inconsistent data availability did not allow us to integrate it into models. It is also important to note that the original spatial resolution of the Landsat 8 Thermal Infrared Sensor (TIRS) (100-m) is a limiting factor in using LST to model Tair, as only the lower temperature of larger urban green spaces can be detected at this resolution (Parlow et al., 2014). Despite these drawbacks, daytime LST was the 5th-ranked (out of 13) variable in variable importance for RF with an average importance of 62.7%, showing its value in Tair modeling. In order to reduce the overestimation of Tair, calculating Tcrowd deviation by subtracting the Tcrowd mean from all Tcrowd values could reduce positive bias in model runs.

The lack of Netatmo stations in green spaces is more difficult to resolve: to depict a more complete picture of the urban climate, crowdsourcing efforts could expand their network away from the city center. CWS could also be installed further away from buildings to better characterize the urban street canyon with well-mixed air, yet this might be challenging for many owners of such devices. The presented LULC analysis does not fully capture the cooling capacity of urban green spaces, which is crucial to highlight for sustainable urban planning. Nevertheless, the current methodology is well-suited for its key application: a short-term heat wave warning system for high-risk populations. A study on heat waves in Berlin found that the highest mortality rates occurred in the most densely built-up districts (Gabriel and Endlicher, 2011)—which are well-represented by Netatmo data (Fenner et al., 2017; Meier et al., 2017) and showed the highest modeling accuracy in our study.

2.5 CONCLUSION

This study presented a method for predicting summertime nocturnal Tair one day in advance in a mid-latitude temperate city using Landsat data, open-source geodata, and crowdsourced Tair data. We compared the performance of three ML algorithms and three different combinations of

predictors. The most accurate ML algorithm was found to be RF, which is encouraging due to its intuitive use and fast computational speed. RF models with only GIS predictors showed a high predictive capacity, with an average RMSE for testing$_{crowd}$ of 1.16 K and RMSE for testing$_{ref}$ of 1.97 K. Models with GIS predictors were generally more accurate, whereas models incorporating Tcrowd as a predictor enhanced accuracy on a night with a relatively spatially homogeneous Tair distribution and higher wind speed. The integration of both geodata and crowdsourced Tair as predictors is recommended in modeling in order to benefit from both high-resolution satellite and morphological data, and Tair distribution from a prior time step. Based on an analysis of variable importance, we recommend the inclusion of morphometric predictors (particularly zd and zHstd) and albedo in Tair prediction. Furthermore, we provided a recommendation for optimal focal buffer radii for the 13 GIS predictors, ranging from 50m to 1500m.

Testing with reference data revealed an overestimation of Tair, particularly in vegetated areas further from the city center. While modeling Tair using crowdsourced atmospheric data holds great promise, crowdsourced datasets incorporate biases due to sensor location and spatial distribution. Nevertheless, the present method is well-suited for prediction of heat risk in the built-up city center, where urban residents face the highest risk of heat stress.

An understanding of the spatial distribution of Tair hot spots can guide sustainable urban planning, which is crucial as urban residents face increasing extreme heat risks. The strength of the developed method is that it can easily be fused with weather forecasts from other models to provide accurate estimates of heat wave hotspots in advance. Therefore, the prediction of Tair in advance can be used as a warning system to detect when vulnerable communities could be affected by hot spots, such as retirement homes, hospitals, and schools. Future studies could test this methodology in other urban areas.

ACKNOWLEDGEMENT

Chapter 3: Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence

This study was published as:

This is the postprint version of the article.

GRAPHICAL ABSTRACT

ABSTRACT

As climate change progresses, urban areas are increasingly affected by water scarcity and the urban heat island effect. Evapotranspiration (ET) is a crucial component of urban greening initiatives of cities worldwide aimed at mitigating these issues. However, ET estimation methods in urban areas have so far been limited. An expanding number of flux towers in urban environments provide the opportunity to directly measure ET by the eddy covariance method. In this study, we present a novel approach to model urban ET by combining flux footprint modeling, remote sensing and geographic information system (GIS) data, and deep learning and machine learning techniques. This approach facilitates spatio-temporal extrapolation of ET at a half-hourly

resolution; we tested this approach with a two-year dataset from two flux towers in Berlin, Germany. The benefit of integrating remote sensing and GIS data into models was investigated by testing four predictor scenarios. Two algorithms (1D convolutional neural networks (CNNs) and random forest (RF)) were compared. The best-performing models were then used to model ET values for the year 2019. The inclusion of GIS data extracted using flux footprints enhanced the predictive accuracy of models, particularly when meteorological data was more limited. The best-performing scenario (meteorological and GIS data) showed an RMSE of 0.0239 mm/ hour and R^2 of 0.840 with RF and an RMSE of 0.0250 mm/ hour and a R^2 of 0.824 with 1D CNN for the more vegetated site. The 2019 ET sum was substantially higher at the site surrounded by more urban greenery (366 mm) than at the inner-city site (223 mm), demonstrating the substantial influence of vegetation on the urban water cycle. The proposed method is highly promising for modeling ET in a heterogeneous urban environment and can support climate change mitigation initiatives of urban areas worldwide.

3.1 INTRODUCTION

Meeting the United Nations' Sustainable Development Goal 11 of "sustainable cities and communities" is critical as urban residents worldwide are increasingly affected by climate change and water scarcity (Liu and Jensen, 2018). By 2050, nearly 70% of the world population is projected to be urban (United Nations, 2019). The paradigm of integrating green infrastructure to support sustainable urban water management is growing in importance and is known by various terms such as water sensitive urban design (WSUD), sponge city, and low impact development (LID) (Liu and Jensen, 2018; Nguyen et al., 2019). These city policies are part of a "co-benefits approach" of climate change mitigation, which simultaneously addresses urban flooding and the urban heat island (He et al., 2019). The ecosystem services provided by green infrastructure, which include the enhancement of flood water infiltration, air temperature modification, energy saving, air quality, biodiversity, and numerous socio-economic benefits, are central to this approach (He et al., 2019; Nguyen et al., 2019; Nouri et al., 2013).

Berlin, Germany, a city internationally recognized for its sustainable water supply system, is one example of a city preparing for climate change impacts by adopting the sponge city concept (Liu and Jensen, 2018). Taking advantage of the cooling effect of evapotranspiration (ET) is a key principle of this concept (Gunawardena et al., 2017). Adding more green infrastructure such as green roofs to the city can provide an evaporative cooling effect during the increasingly common heat waves expected in northern Europe (Liu and Jensen, 2018; Meehl and Tebaldi, 2004; Vulova and Kleinschmit, 2019). Therefore, an enhanced understanding of ET in urban areas is essential to implementing climate change mitigation and urban greening schemes (Gunawardena et al., 2017).

Despite its significance, ET estimation in urban environments remains a nascent area of research, as most ET studies have focused on natural or agricultural areas (Nouri et al., 2013; Saher et al., 2021). The urban environment is spatio-temporally heterogeneous, with a mixture of land cover, microclimate, and soil and water characteristics (Nouri et al., 2013). Point-based ET measurements such as lysimeters and sap flow may not be representative or practical in an urban setting (Litvak et al., 2017).

Surface energy balance (SEB) models use remotely sensed imagery and auxiliary meteorological data as input and estimate ET as the residual of the energy balance equation (Norman et al., 1995; Saher et al., 2021; Su, 2002). Some commonly applied SEB models include the two-source energy balance (TSEB) model (Norman et al., 1995), Surface Energy Balance Algorithm for Land (SEBAL) (Bastiaanssen et al., 1998), and Surface Energy Balance System (SEBS) (Su, 2002). Jiang and Weng (2017) applied TSEB to model daily ET in an urban area in the United States; however, ET was not validated with field data in this study. Cong et al. (2017) compared SEBS to a water balance method to estimate ET in Beijing, China, finding that SEBS severely underestimated annual ET. SEB models are not suitable for characterizing urban ET without major adjustments, as they omit anthropogenic heat and diurnal advection effects, assume homogeneous vegetation, operate on a regional scale, and ignore atmospheric variability in urban areas (Saher et al., 2021).

Physically based hydrological models simulate hydrological processes, including soil evaporation, transpiration, interception, subsurface flow, and channel flow, using representations of partial differential equations (Dwarakish et al., 2015). Recent hydrological models incorporate remote sensing data such as the Normalized Difference Vegetation Index (NDVI) as a proxy for leaf area index (LAI), biomass, and urban impervious fraction (Boegh et al., 2009, 2004). Physically based hydrological models are challenging to implement as they necessitate the estimation of numerous parameters, which often require expert knowledge or field measurements to be determined (Boegh et al., 2009; Dwarakish et al., 2015).

Eddy covariance (EC), a micrometeorological technique which measures turbulent fluxes of sensible and latent heat, momentum and other gases e.g. CO_2, is one of the few methods which can directly measure ET in a heterogeneous urban environment (Nouri et al., 2013). However, EC measurements have mostly been used in agricultural, forestry and riparian studies so far (Holl et al., 2020; Järvi et al., 2012; Menzer et al., 2015; Moffat et al., 2007; Papale and Valentini, 2003). Limited EC applications over urban areas were predominantly used to assess exchanges of energy, greenhouse gases and air pollutants rather than LE or ET (Holl et al., 2020; Järvi et al., 2012; Kordowski and Kuttler, 2010; Menzer et al., 2015; Moffat et al., 2007; Papale and Valentini, 2003). Moreover, EC measurements are frequently affected by instrument failures, low turbulent atmospheric conditions, and system "spikes", leading to data gaps accounting for 20-

60% of data on an annual basis (Moffat et al., 2007). Gap-filling or modeling is therefore indispensable in order to obtain daily, monthly, and annual sums of ET, which can be used to better characterize the urban water cycle.

Current modeling and gap-filling approaches are insufficient in a heterogeneous urban environment due to the changing source area of EC measurements and urban heterogeneity (Kotthaus and Grimmond, 2014; Menzer et al., 2015). Previous studies modeling CO_2 have addressed the issue of urban heterogeneity by training models independently on different wind direction sectors (Menzer et al., 2015) or adding wind direction as a binary predictor (Järvi et al., 2012). However, these approaches do not integrate the influence of wind speed and atmospheric stability in controlling footprint area (Kljun et al., 2002; Kotthaus and Grimmond, 2014). Footprint models, on the other hand, can estimate the likely surface area affecting turbulent flux measurements at a given point in time (Kotthaus and Grimmond, 2014). Combining footprint modeling with a geospatial database is highly beneficial for the interpretation of urban fluxes (Christen et al., 2011; Kotthaus and Grimmond, 2014). However, footprint modeling has not thus far been combined with machine learning (ML) regression in order to model urban ET.

Although studies modeling EC fluxes at a half-hourly scale have applied ML approaches to gap-fill the datasets, the application of ML models in new locations has not thus far been assessed. Extracting remotely sensed data by footprint modeling opens up the opportunity to spatially upscale ET to the city scale at a high resolution in the future (Crawford and Christen, 2015). We therefore test the potential to model ET in areas where no flux tower is available by training in one location and testing in another. We model urban ET using 1D convolutional neural networks (1D CNNs) and random forest (RF), which have never and rarely been applied to model EC fluxes so far, respectively (Kim et al., 2020).

In this study, we evaluated an approach combining flux footprint modeling, deep learning (DL) and ML, and GIS and remotely sensed data in order to model ET at a half-hourly resolution in a heterogeneous urban environment. We assessed the hypothesis that integrating remote sensing and GIS data by footprint modeling, rather than relying solely on meteorological data, can better characterize urban ET. We tested this methodology in Berlin, Germany with a two-year dataset from two flux towers. Our key objectives were to: 1) assess the benefit of remote sensing and GIS data extracted by flux footprints to modeling urban ET; 2) identify the main drivers of ET in an urban environment; 3) compare the performance of two ML and DL algorithms (1D CNN and RF) and four predictor scenarios in modeling urban ET; 4) compare model performance at two urban flux tower sites with differing land cover and evaluate to what extent models can be applied in new locations; and 5) derive monthly and annual ET estimates.

3.2 DATA

A set of temporally dynamic and static data extracted from satellite imagery, flux towers, and meteorological stations was used to predict ET in an urban landscape. An overview of the 23 predictors used to model ET at a half-hourly resolution is given in Table 3.1.

3.2.1 STUDY AREA

The study is based in Berlin, the largest city in Germany with 3.7 million inhabitants distributed across 891 km^2 in 2018 (Statistical Office of Berlin-Brandenburg, 2019) (Figure 1.1). Berlin is located in eastern Germany (52.52 °N, 13.40 °E) and has a humid warm temperate climate (Kottek et al., 2006).

The two flux towers are part of the Urban Climate Observatory (UCO) Berlin maintained by the Chair of Climatology at the Technische Universität Berlin (TUB) (Scherer et al., 2019). The Rothenburgstraße (ROTH) flux tower is located in an urban research garden belonging to the TUB in the southwest of the city (Figure 3.1). The TUB Campus Charlottenburg (TUCC) flux tower is located on top of the main building of the TUB in the center of the city (Figure 3.1). The measurement heights are 39.75 m and 56 m for the ROTH and TUCC towers, respectively. The tower at ROTH is situated in a more vegetated area, with 50.6% vegetation, 26.3% impervious surface, 22.8% buildings, and 0.3% water bodies within a 1000-m radius. The tower at TUCC, in contrast, is characterized by 32.6% vegetation, 35.5% impervious surface, 27.2% buildings, and 4.7% water bodies within a 1000-m radius.
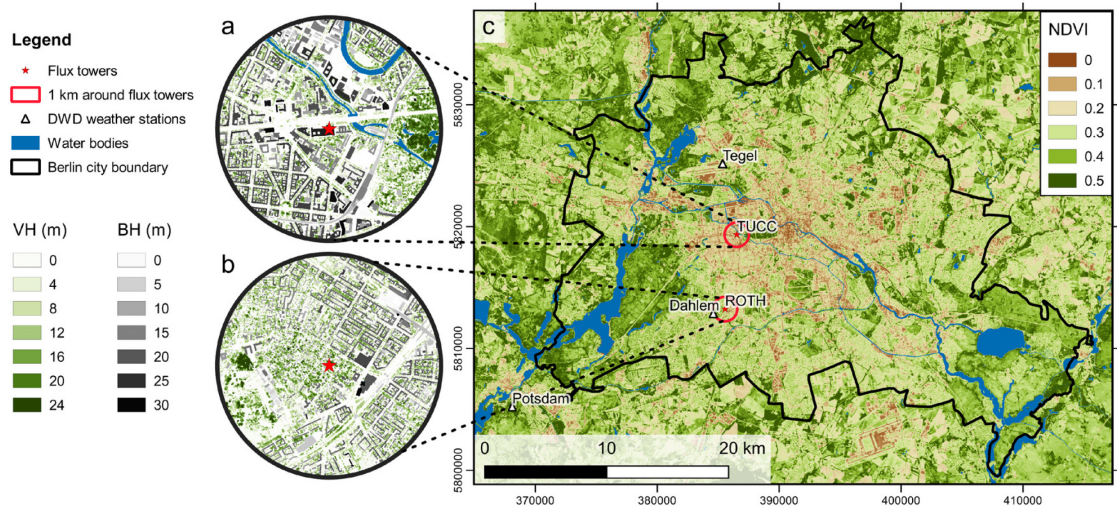
Figure 3.1: Locations of the two flux towers in Berlin, Germany. A 1-km radius showing vegetation height (VH) and building height (BH) around the towers is depicted for a) TUCC and b) ROTH. c) The Normalized Difference Vegetation Index (NDVI) from Landsat 8 is shown for 24 June 2019 in Berlin and its surrounding area. The datum is WGS-84 and the projection is UTM.

3.2.2 METEOROLOGICAL DATA

Hourly meteorological data were acquired from the German Meteorological Service (DWD) Climate Data Center (DWD, 2021) and used as ET predictors (Table 1). Solar radiation data were acquired from the station Potsdam (19.2 km and 23.2 km distance from ROTH and TUCC, respectively) for both towers. All other DWD data were acquired at stations Berlin-Dahlem (1.04 km distance from ROTH) and Berlin-Tegel (5.94 km distance from TUCC) for ROTH and TUCC, respectively. Linear interpolation was applied for short (<4 h) gaps. Lastly, all DWD data were linearly interpolated to a half-hourly resolution.

Reference ET (ETo) was calculated with the hourly ASCE "Standardized Reference Evapotranspiration Equation" for short crops (Allen et al., 2005) using the "hourlyET" function of the R package "water" (Olmedo et al., 2016). For further details on the parametrization of ETo and saturated vapor pressure (SVP), refer to Appendix B.

Figure 3.2a-d shows daily air temperature, shortwave downward radiation, precipitation, and NDVI throughout the study period. Figure 3.2e shows the temporal distribution of half-hourly ET after quality control (in units of mm/ hour) measured at the two flux towers during the study period. Both towers follow a typical seasonal pattern for ET, with the highest maximum values in summertime and the lowest maximum values in wintertime. As expected, the maximum ET values at the more vegetated site (ROTH, green dots) are nearly double the values at the less vegetated site (TUCC, orange dots) in the growing season.
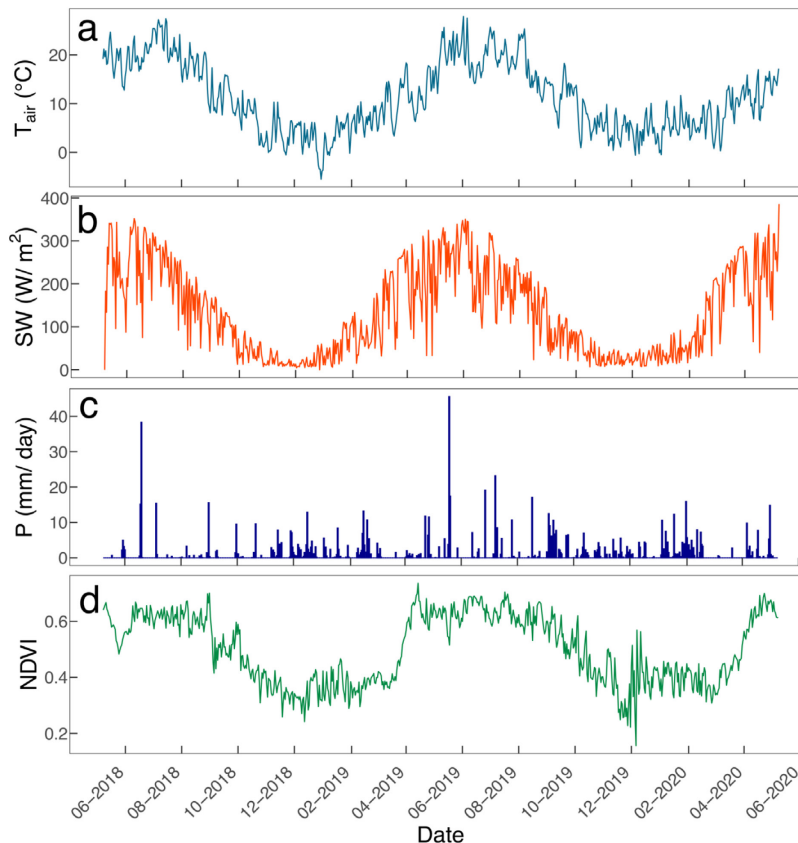
Figure 3.2: Meteorological and vegetation greenness conditions during the study period (1 June 2018 to 1 June 2020): (a) daily averaged air temperature (Tair), (b) daily averaged shortwave downward radiation (SW), (c) daily precipitation (P), (d) Normalized Difference Vegetation Index (NDVI), and (e) half-hourly ET after quality control measured at the two flux towers during the study period, with ROTH depicted in green and TUCC in orange. NDVI data were extracted from flux footprint modeling (at a half-hourly temporal scale) at the ROTH flux tower and averaged to a daily scale. Tair and P were derived from the German Meteorological Service (DWD) station Berlin-Dahlem (FU) and SW from the DWD station Potsdam. The study period was drier and warmer than the long-term (1981-2010) mean, with an average annual P of 448 mm (long-term mean: 591 mm) and an average Tair of 11.3 °C (long-term mean: 9.5 °C) (DWD, 2021, 2020).

3.2.3 FLUX MEASUREMENTS, FILTERING AND DATA PROCESSING

The EC system at both sites combines an open-path gas analyzer and a three-dimensional sonic anemometer-thermometer (IRGASON, Campbell Scientific) for simultaneous measurements of water vapor density as well as orthogonal wind components. The EC software package EddyPro (Version 6.2.1) was used to quality-control EC raw data and to calculate ET from 20-Hz time series over 30-min intervals.

Data quality control and processing with EddyPro included elimination of spikes, filtering values based on physical thresholds, and statistical screening based on the method developed by Vickers and Mahrt (1997). We applied double coordinate rotation, correction of sonic temperature for humidity, high- and low-frequency spectral corrections (Moncrieff et al., 1997), and corrections for air density (Webb et al., 1980). Furthermore, instrument diagnostic flags not equal zero and data with signal strength <0.8 were withheld. EddyPro output data at 30-minute resolution with quality flag 2 were excluded (Foken, 2016) as well as data during and 4 hours after precipitation events based upon DWD precipitation data. Additionally, data from wind directions 17° - 35° at TUCC and 54° - 72° at ROTH were excluded because this sector is influenced by flow distortion due to the instrument and mounting setup (Foken, 2016). De-spiking based on standard deviation (SD) was applied, with LE data five times greater than the SD removed as spikes using the R package "FREddyPro" (Xenakis, 2016). A simple threshold was applied, excluding data where LE was below -100 W/ m^2 and above 500 W/ m^2. Furthermore, negative values of ET were removed as they represent condensation.

Based on overlapping observations from the two towers, the study period was restricted to two years (1 June 2018 to 1 June 2020). A total of 16,707 and 16,013 (48% and 46% of the study period) high quality half-hourly ET observations (in mm/ hour) then remained available for modeling at ROTH and TUCC, respectively.

3.2.4 TURBULENT FLUX FOOTPRINTS

Modeling urban ET measured by flux towers is challenging due to the dynamic influence of heterogeneous land cover. Turbulent fluxes measured by an EC measurement system do not represent the energy exchange of a fixed radius around the site, but rather that of a constantly changing upwind area which contains the sources and sinks contributing to measurements (Foken, 2016; Kljun et al., 2002; Schmid and Oke, 1990). This area is referred to as a footprint or turbulent flux source area and can be estimated by footprint models (Crawford and Christen, 2015; Kljun et al., 2002). In this study, footprint areas were estimated with the Kormann and Meixner (2001) analytical footprint model using the R package "FREddyPro" (Xenakis, 2016). Further details on the parametrization of footprint models are given in Appendix B. The resulting footprint grids were reduced to the 90% probability area. Weighted averages of the surface cover were extracted by multiplying these grids with the raster layers and summing the resulting pixel values on a half-hourly basis (Figure 3.3).

3.2.5 REMOTE SENSING AND GIS DATA

Six remote sensing and GIS datasets were used as predictors of urban ET (Table 3.1; Figure 3.3), which are referred to as "GIS predictors."

The Harmonized Landsat and Sentinel-2 (HLS) surface reflectance product provided by NASA was used to compute NDVI, an indicator of vegetation density and vitality (Claverie et al., 2018; Tucker, 1979). This open-source product is well-suited for vegetation phenology monitoring in urban areas with a revisit period of 3-5 days and a 30-meter resolution (Claverie et al., 2018). The preprocessing and interpolation to daily resolution of HLS data are further described in Appendix B.

Vegetation height (VH) and building height (BH) were provided by the Berlin Digital Environmental Atlas at 1-meter resolution (Berlin Senate Department for Urban Development and Housing, 2014). Vegetation fraction was derived by converting VH pixels higher than 0.01 m to a value of 1 and all other pixels to 0. Land use and land cover (LULC) classes data were provided by Urban Atlas 2012 (European Environment Agency, 2018). Water fraction was derived by assigning the LULC classes "Water" and "Wetlands" to a value of 1 and all other classes to 0. Impervious surface fraction (ISF) was also derived from the Urban Atlas LULC dataset, as described in Vulova et al. (2020).

Table 3.1: An overview of the predictor variables used to model evapotranspiration (ET) at half-hourly resolution; details of the source of the data, preprocessing method, and their spatial and temporal resolutions are listed. In the "Temporal resolution" column, "static" refers to predictors which remain constant over time (such as building height). In the "Spatial resolution" column, P refers to point data (derived from a single point in space),

while F refers to remote sensing and GIS data extracted with flux footprints. All meteorological data (including ETo) were linearly interpolated to a half-hourly resolution.

Predictor (acronym)	Unit	Data source	Preprocessing method	Spatial resolution (m)	Temporal resolution
GIS predictors					
Building height (BH)	m	Berlin Digital Environmental Atlas (2014)	-	1 (F)	Static
Impervious surface fraction (ISF)	%	Urban Atlas 2012 (LULC)	Vulova et al. (2020); rasterization with 1-m resolution	1 (F)	Static
Normalized Difference Vegetation Index (NDVI)	-	Harmonized Landsat and Sentinel-2 (HLS) surface reflectance product, NASA	Claverie et al. (2018); cloud-masking; pixel-wise linear interpolation to daily scale	30 (F)	Daily
Vegetation fraction	%	Berlin Digital Environmental Atlas (2014) VH	Conversion of VH higher than 0.01 m to 1 and all other pixels to 0	1 (F)	Static
Vegetation height (VH)	m	Berlin Digital Environmental Atlas (2014)	-	1 (F)	Static
Water fraction	%	Urban Atlas 2012 (LULC)	Rasterization of water bodies with 1-m resolution (1 = present, 0 = absent)	1 (F)	Static
Meteorological predictors					
Air pressure	hPa	DWD	-	P	Hourly
Air temperature	°C	DWD	-	P	Hourly
Diffuse solar radiation	W/m^2	DWD	Unit conversion from J/cm^2 to W/m^2	P	Hourly
Dry bulb temperature	°C	DWD	-	P	Hourly
Longwave downward radiation	W/m^2	DWD	Unit conversion from J/cm^2 to W/m^2	P	Hourly
Reference evapotranspiration (ETo)	mm/hour	DWD	Allen et al. (2005) equation for short crops; "water" R package (Olmedo et al., 2016)	P	Hourly
Relative humidity (RH)	%	DWD	-	P	Hourly
Saturated vapor pressure (SVP)	hPa	DWD	Allen et al. (1998); "MeTo" R package (Dettmann and Grimma, 2019)	P	Hourly
Shortwave downward radiation	W/m^2	DWD	Unit conversion from J/cm^2 to W/m^2	P	Hourly
Soil temperature at 5 cm, 10 cm, 20 cm, 50 cm, and 100 cm depth	°C	DWD	-	P	Hourly
Solar zenith angle	°	DWD	-	P	Hourly
Vapor pressure deficit (VPD)	hPa	DWD	-	P	Hourly
Wind speed	m/s	DWD	-	P	Hourly

3.3 METHODS

3.3.1 ET MODELING

3.3.1.1 PREDICTOR SCENARIOS

To explore the contribution of different data sources (remote sensing and GIS, meteorological, and ETo) to estimating ET, four predictor scenarios were tested in models: (1) only ETo as a predictor (one predictor; "ETo"), (2) ETo and GIS predictors (seven predictors; "ETo and GIS"), (3) only meteorological predictors (16 predictors; "Met"), and (4) meteorological and GIS predictors (22 predictors; "Met and GIS"). GIS predictors refer to the six remote sensing and GIS datasets (BH, VH, ISF, NDVI, vegetation fraction, and water fraction) (Table 3.1; Figure 3.3). Meteorological predictors refer to all predictors except GIS predictors and ETo (Table 1). In total, 32 models were run, with two towers, two temporal training/ testing splits, four predictor scenarios, and two artificial intelligence (AI) algorithms.

Figure 3.3 presents an overview of the entire methodological approach.
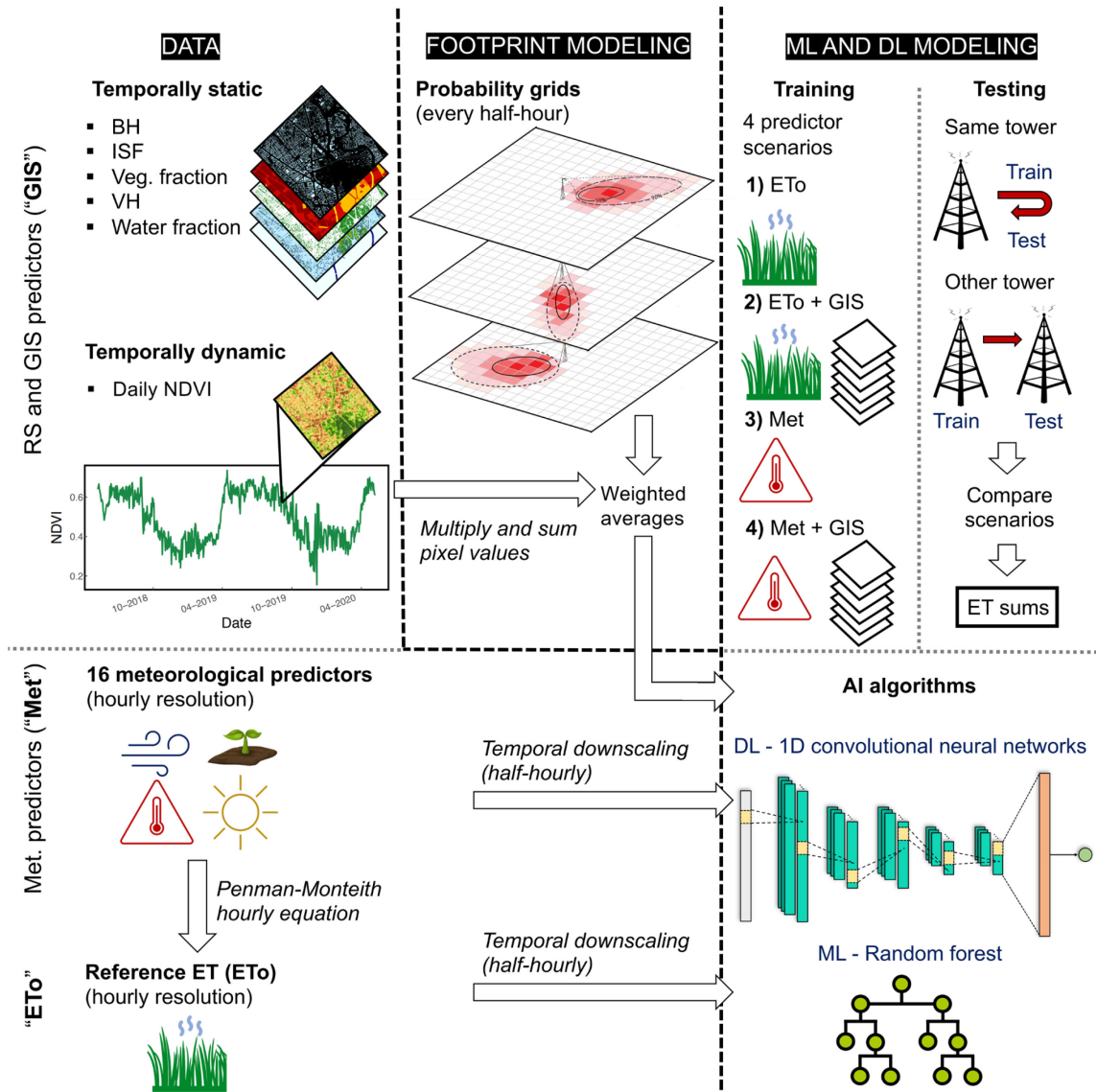
Figure 3.3: Flowchart showing an overview of the study, including the data used for ET modeling, footprint modeling, and ML and DL modeling. Abbreviations can be found in the text. The probability grids diagram is adapted from Christen (2016).

3.3.1.2 ML AND DL ALGORITHMS

ML and DL algorithms, which broadly fall under the umbrella of AI, have emerged as particularly accurate methods for gap-filling urban fluxes (Järvi et al., 2012; Menzer et al., 2015; Schmidt et al., 2008). ML and DL algorithms were selected to model urban ET as they can generalize from sample data without relying on a process-based formulation (Kuhn and Johnson, 2013; Menzer et al., 2015; Schmidt et al., 2008). Furthermore, ML and DL algorithms allow for the integration of a large set of predictor variables and the estimation of variable importance (Kuhn and Johnson, 2013). Here, we tested two supervised regression DL and ML algorithms: a 1D CNN and RF.

3.3.1.2.1 1D CONVOLUTIONAL NEURAL NETWORKS (CNN)

CNNs are a type of DL model defined by convolutional layers, which learn local patterns automatically (Chollet and Allaire, 2018). Whereas in computer vision 3D convolutions are used to extract local spatial features, 1D convolutions essentially treat time as a spatial dimension, extracting local patterns in a sequence and recognizing them at a different temporal position (Chollet and Allaire, 2018). This capacity makes them highly relevant to sequence processing, with recent applications in hydrological modeling (Chollet and Allaire, 2018; Ferreira and da Cunha, 2020; Haidar and Verma, 2018).

The CNN architecture used in this study consisted of three convolutional layers, with each of the first two convolutional layers immediately followed by max pooling layers. The third convolutional layer was followed by a flatten layer in order to convert the 3D outputs to 2D outputs (Chollet and Allaire, 2018). The final (seventh) layer is a 1-dimensional dense layer with linear activation, generating the ET prediction. CNN architecture was set up according to the recommendations of Chollet and Allaire (2018) and trial and error experimentation. In convolutional layers, padding was set to "same", stride was set to 1, and the activation function was set to rectified linear unit (ReLU). To train the models, the Adam algorithm (Kingma and Ba, 2015) and the mean squared error (MSE) loss function were used. The learning rate was set to the default of 0.001. To limit overfitting, early stopping was used to define the number of training epochs, with a maximum of 100 training epochs and a patience of 10 epochs. In-depth information on CNNs and their implementation using the R interface to Keras can be found in Chollet and Allaire (2018). A detailed description of hyperparameter tuning for CNNs is given in Appendix B.

3.3.1.2.2 RANDOM FOREST (RF)

RF was used to model urban ET due to its capacity to handle highly nonlinear relationships, avoid overfitting, and produce a stable prediction (Kim et al., 2020; Kuhn and Johnson, 2013). RF is an ensemble decision tree-based algorithm proposed by Breiman (2001). Further details on RF and its implementation are given in Kuhn and Johnson (2013). Hyperparameter tuning for RF is described in Appendix B.

3.3.1.2.3 VARIABLE IMPORTANCE

Relative variable importance scores (scaled from 0 to 100%) were extracted from models trained with the "Met and GIS" predictor scenario. Variable importance was extracted by the permutation approach described by Breiman (2001) using the "vip" R package for CNNs (Greenwell et al., 2020) and the "caret" R package for RF (Kuhn, 2008).

3.3.2 ET PERFORMANCE METRICS

To evaluate the temporal and spatial extrapolation capability of models (Roberts et al., 2017), we split the data into training and testing sets. For the temporal extrapolation capability, we used 2018 and 2020 data for training and 2019 data for testing and vice versa. For ROTH, 8617 and 8090 observations were available for 2018/ 2020 and 2019, respectively. For TUCC, 8055 and 7958 observations were available for 2018/ 2020 and 2019, respectively. For the spatial extrapolation capability, we tested two scenarios: (1) training and testing on the same tower and (2) training on one tower and testing on the other.

The model performance was evaluated using the testing set, and the prediction accuracy was assessed using five metrics: root mean square error (RMSE), percent bias (pbias), coefficient of determination (R^2), mean absolute error (MAE), and normalized root mean square error (NRMSE). The value used to normalize NRMSE was the difference between the maximum and minimum observed values.

3.3.3 TOTAL ET

Monthly and annual ET sums for the year 2019 were calculated by two methods: (1) by gap-filling using the ML and DL algorithms and (2) by exclusively modeling ET using the ML and DL algorithms without using the high-quality ET data for 2019. All models used for ET sums were trained in 2018 and 2020 and on the same tower for which data were predicted. In all cases, the "Met and GIS" predictor scenario was applied first, as it was found to be the best-performing scenario. In some cases, flux footprints could not be modeled due to unsuitable atmospheric conditions for footprint modeling and missing EC input data. Thus, any remaining gaps were filled using models trained with the "Met" predictor scenario. For ROTH, 9430 half-hours in 2019 were missing and thus gap-filled (53.8 % of the data). For TUCC, 9562 half-hours in 2019 were missing and thus gap-filled (54.6 % of the data). Since ET was modeled at a half-hourly scale but is given in units of mm/ hour, half-hourly values were averaged to an hourly scale before summing ET.

Unless stated otherwise, all modeling and analysis were completed using R version 3.6.3 (2020-02-29) and RStudio (version 1.3.1073) (R Core Team, 2020). CNNs were implemented through an R interface using the Keras DL framework and the TensorFlow backend (Allaire and Chollet, 2020; Chollet and Allaire, 2018). CNN models were trained on a local workstation using the CUDA environment (GPU-based processing) and a NVIDIA graphics card (Quadro P1000). The "caret" R package was used as a wrapper package for the RF algorithm implementation in R (Kuhn, 2008). For RF model tuning and training, parallel processing was performed on 14 cores using the R package "doSNOW" (Microsoft Corporation and Weston, 2019).

3.4 RESULTS

3.4.1 ET MODELING

When modeling both towers independently, the scenarios including GIS predictors show the highest prediction accuracy. Table 3.2 presents the testing performance metrics averaged across the two temporal training and testing splits, providing an overall performance estimate of models. The "Met and GIS" scenario shows the lowest RMSE, NRMSE, and MAE and highest R^2 in the majority of cases. With RF, the "Met and GIS" scenario shows the highest accuracy overall, with an RMSE of 0.0239 mm ($R^2 = 0.840$ and MAE = 0.0154 mm) for ROTH and 0.0170 mm ($R^2 = 0.544$ and MAE = 0.0115 mm) for TUCC (Table 3.2). The "Met and GIS" scenario has a similar performance with CNN, with an RMSE of 0.0250 mm ($R^2 = 0.824$ and MAE = 0.0160 mm) for ROTH and 0.0178 mm ($R^2 = 0.502$ and MAE = 0.0119 mm) for TUCC.

The increase in performance is greater when adding GIS predictors to a simple "ETo" scenario than a scenario with 16 meteorological predictors. Comparing the simple predictor scenarios of "ETo" and "ETo and GIS", "ETo and GIS" achieves a higher performance (lower RMSE, NRMSE, and MAE and higher R^2) for both towers (Table 3.2). The "Met and GIS" scenario shows increased prediction accuracy than the "Met" scenario (Table 3.2). For instance, at ROTH with RF, "ETo and GIS" has a nearly 0.01 mm lower RMSE (26% decrease) and a 0.15 higher R^2 (22% increase) than "ETo", whereas the increased performance of "Met and GIS" compared to "Met" is less pronounced: 0.002 mm lower RMSE (7% decrease) and 0.02 higher R^2 (3% increase).

RF outperforms CNN in most predictor scenarios, although the performance of both algorithms is comparable (Table 3.2). In a few scenarios, CNN performs better ("ETo" for ROTH; "ETo" and "ETo and GIS" for TUCC).

The RMSE and R^2 values cannot be directly compared between the two towers, as they are dependent on the variation of ET. The range and maximum of ET are lower at TUCC, a site with less surrounding vegetation, than at ROTH (Figure 3.2e), leading to lower RMSE and R^2 values at TUCC (Table 3.2). In the study period, the maximum ET at ROTH is nearly double (0.29 mm/ hour) the maximum ET at TUCC (0.16 mm/ hour) (Figure 3.2e). R^2, in particular, is highly affected by the variation in the dependent variable (Kuhn and Johnson, 2013). NRMSE is normalized RMSE which facilitates a comparison of the prediction accuracy which takes into account the different ranges of ET at the two towers. When training and testing occurs on the same tower, NRMSE is generally lower at the more vegetated site (ROTH) than at the less vegetated site (TUCC). In most cases, training on the more vegetated site (ROTH) and testing on the less vegetated site (TUCC) is associated with a higher NRMSE than the opposite scenario due to the lower ET range at TUCC.

Models for both towers have a low pbias when training and testing in different years (Appendix B; Table B.3 and Table B.4). However, training and testing on different towers increases the RMSE substantially while showing a relatively small effect on the R^2 (Table 3.2). This increase in RMSE occurs due to the considerable difference in the average of ET between the two towers (0.05 vs. 0.03 mm/ hour at ROTH and TUCC, respectively), while the similar R^2 values indicate that the relationship between most predictors and ET is similar for both towers.

The addition of GIS data is associated with lower RMSE, NRMSE, and MAE for RF models when training in ROTH and testing in TUCC. Otherwise, the inclusion of GIS data is associated with a higher RMSE, NRMSE, and MAE when training and testing on different towers. This effect can be explained by the contrasting correlations between ET and the GIS predictors observed at the two towers (Appendix B; Figure B.3). At ROTH, ET correlates positively with indicators of higher vegetation presence (VH, Pearson's r = 0.30; vegetation fraction, Pearson's r = 0.32) and negatively with indicators of impervious land cover (ISF, Pearson's r = -0.45; BH, Pearson's r = -0.27), as expected. In contrast, at TUCC, ET has a weak positive correlation with BH (Pearson's r = 0.06) and an insignificant correlation with ISF. Unexpectedly, ET at TUCC correlates negatively with both vegetation fraction (Pearson's r = -0.15) and VH (Pearson's r = -0.15).

Training on TUCC leads to an underestimation of ET at ROTH, with a pbias ranging from -44% to -55% (Table 3.2). As ET at TUCC is on average lower (Figs. 3 and 7), models trained on TUCC data cannot predict the higher ET observed at ROTH. Conversely, most models trained on ROTH data overestimate the lower ET observed at TUCC, as indicated by a positive pbias (52% - 87%) for all predictor scenarios except "ETo and GIS." The decrease in pbias when including GIS data in models trained in ROTH and tested in TUCC can be explained by the contradictory relationship between ET and GIS predictors at the two towers (Appendix B; Figure B.3).

Monthly and diurnal modeling accuracy are discussed in Appendix B (Figures B.6 - B.10).

Table 3.2: Testing performance metrics averaged across the two training and testing splits (training in 2018/ 2020 and testing in 2019 and vice versa). The performance metrics are root mean square error (RMSE), mean absolute error (MAE), percent bias (pbias), coefficient of determination (R^2), and normalized root mean square error (NRMSE). The best performance metrics for each tower training and testing combination (e.g., training in ROTH and testing in TUCC) are shown in bold.

Tower (train)	Tower (test)	Predictors	RMSE (mm/ hour)		MAE (mm/ hour)		pbias (%)		R^2 (-)		NRMSE (%)	
			CNN	RF	CNN	RF	CNN	RF	CNN	RF	CNN	RF
ROTH	ROTH	ETo	0.0304	0.0355	0.0195	0.0226	-6.20	0.95	0.745	0.660	10.55	12.30
		ETo and GIS	0.0266	0.0263	0.0177	0.0169	1.05	**-0.20**	0.801	0.805	9.25	9.10
		Met	0.0274	0.0256	0.0178	0.0164	4.65	2.35	0.793	0.817	9.50	8.85
		Met and GIS	0.0250	**0.0239**	0.0160	**0.0154**	-3.15	1.80	0.824	**0.840**	8.65	**8.30**
TUCC	TUCC	ETo	0.0186	0.0208	0.0121	0.0140	-4.00	0.80	0.486	0.360	11.70	13.05
		ETo and GIS	0.0174	0.0175	**0.0114**	0.0116	-4.65	1.40	0.525	0.514	10.95	11.00
		Met	0.0178	0.0173	0.0119	0.0119	-0.60	5.90	0.508	0.529	11.15	10.85
		Met and GIS	0.0178	**0.0170**	0.0119	0.0115	**0.35**	4.45	0.502	**0.544**	11.15	**10.65**
ROTH	TUCC	ETo	0.0399	0.0475	0.0247	0.0285	70.90	83.95	0.486	0.452	25.05	29.80
		ETo and GIS	0.0635	**0.0226**	0.0398	**0.0153**	-137.10	**-7.45**	0.067	0.369	39.90	**14.20**
		Met	0.0473	0.0446	0.0290	0.0273	90.50	87.15	0.507	**0.521**	29.70	27.95
		Met and GIS	0.0636	0.0318	0.0393	0.0202	-119.00	52.10	0.145	0.514	39.70	19.90
TUCC	ROTH	ETo	0.0509	0.0524	0.0297	0.0311	-48.60	-46.65	0.744	0.524	17.65	18.20
		ETo and GIS	0.0583	0.0538	0.0349	0.0321	-61.70	-54.75	0.609	0.703	20.25	18.65
		Met	0.0516	**0.0491**	0.0313	**0.0289**	-47.10	**-44.00**	0.665	0.775	17.95	**17.05**
		Met and GIS	0.0582	0.0505	0.0352	0.0297	-61.15	-49.15	0.596	**0.780**	20.20	17.50

3.4.2 VARIABLE IMPORTANCE

The relative variable importance of predictors for both towers was tested. Figure 3.4 shows the percentage of relative influence of predictors with the "Met and GIS" scenario, which includes most of the predictors (except ETo). Despite how strongly ET is driven by meteorological conditions, the influence of GIS predictors is apparent in the RF variable importance. For RF models at ROTH, NDVI, ISF, and BH are the third, fifth, and sixth most important predictors, respectively. The most important meteorological predictors at both ROTH and TUCC with RF are wind speed, air pressure, and solar zenith angle.

RF variable importance differs between the two towers, reflecting the higher influence of impervious cover at TUCC. The second and fourth most important predictors at TUCC with RF are ISF and BH, respectively. Water fraction also has a minor influence in TUCC, whereas in ROTH water fraction has essentially no influence due to the lack of water bodies around the tower (Figure 3.4).

CNN variable importance differs substantially from RF variable importance. With CNNs, the most important meteorological predictors are shortwave downward radiation, solar zenith

angle, and diffuse solar radiation at ROTH and solar zenith angle, wind speed, and shortwave downward radiation at TUCC. ISF is the 7th most important predictor at ROTH, whereas the other GIS predictors are low ranked in importance.
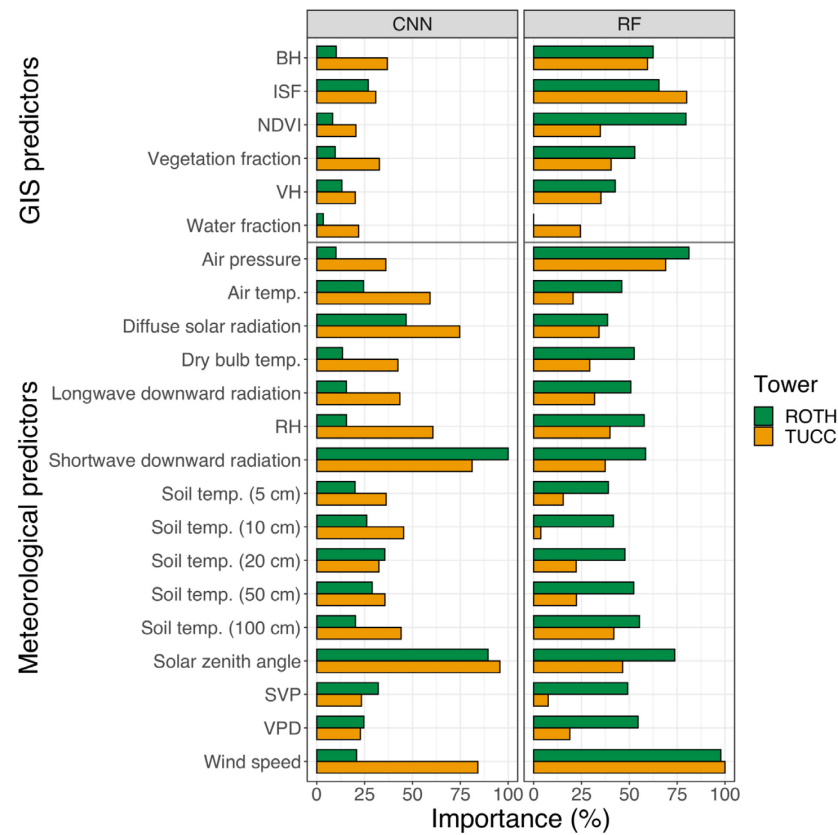
Figure 3.4: Relative variable importance revealed by models. The "Met and GIS" (meteorological and GIS predictors) predictor scenario is depicted. The variable importance was averaged across the two training and testing splits (training in 2018/ 2020 and testing in 2019 and vice versa).

3.4.3 TOTAL ET

Monthly and annual ET sums were computed for both towers (Figure 3.5). ET sums show a substantially higher ET at the more vegetated site (ROTH), as expected (Figure 3.5). The annual ET sum for 2019 is ~140 mm higher at ROTH than at TUCC, with an annual sum of 366 mm and 223 mm, respectively, when gap-filling with the best-performing algorithm (RF) (Figure 3.5). The highest monthly ET sums are in the summer months; for ROTH, gap-filled ET sums with RF are 63 mm in June, 54 mm in July, and 50 mm in August (Figure 3.5). For TUCC, gap-filled ET sums with RF are around half the sums at ROTH: 31 mm in June, 28 mm in July, and 27 mm in August. Annual ET sums differ by only 3-5 mm when modeling the entire year of 2019 rather than gap-filling, confirming the dependability of the approach for ET estimation. Modeling the entire year without gap-filling reproduces the diurnal and annual cycle of the original data well

for both algorithms (Figure 3.6). CNN better reproduces the high summer daytime ET at ROTH, whereas these values are underestimated by RF (Figure 3.6; Appendix B Table B.3). Nevertheless, differences in annual ET sums between the two algorithms are negligible, even when solely modeling (4 mm and 10 mm difference at ROTH and TUCC, respectively) (Figure 3.5). For both towers, 14 hours remained where ET could not be modeled due to missing DWD data, which were assigned to zero.
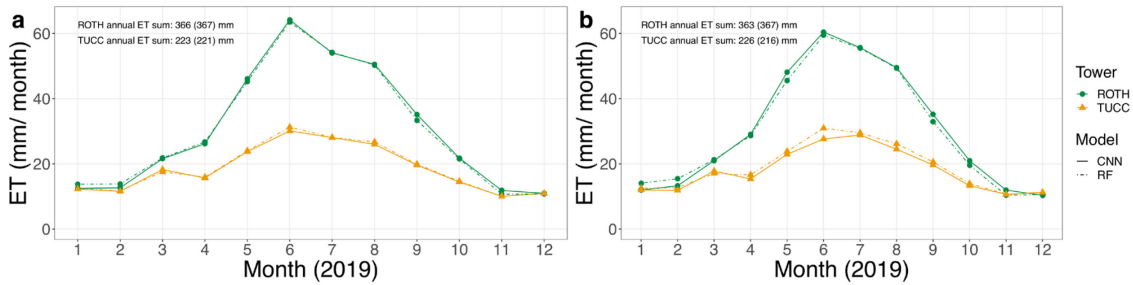
Figure 3.5: Monthly evapotranspiration (ET) sums for the year 2019 calculated by (a) gap-filling using random forest (RF) and 1D convolutional neural networks (CNN) and (b) by only modeling using RF and 1D CNN without using the original ET data for 2019. The first annual ET sum values listed represent the RF-based annual sums, while the CNN-based annual sums are in parentheses. All models used to obtain ET sums were trained in 2018 and 2020 and on the same tower for which data were predicted. In all cases, the "Met and GIS" (meteorological and GIS) predictor scenario was applied first. Any remaining gaps were filled using models trained with the "Met" (meteorological) predictor scenario (see main text for further details).
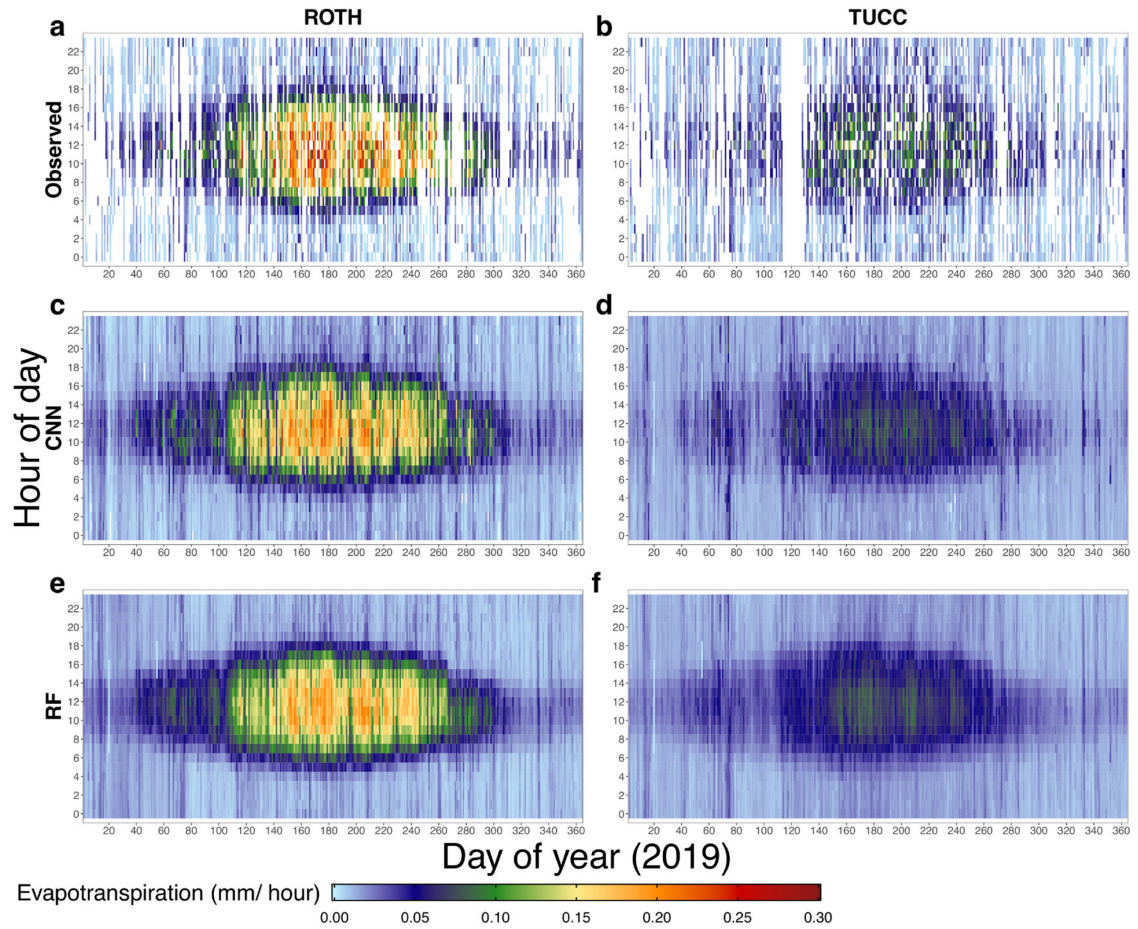
Figure 3.6: Annual and diurnal course of (a) observed evapotranspiration (ET) at ROTH, (b) observed ET at TUCC, (c) modeled ET at ROTH with 1D convolutional neural networks (CNN), (d) modeled ET at TUCC with 1D CNN, (e) modeled ET at ROTH with random forest (RF), and (f) modeled ET at TUCC with RF. ET depicted in (c)-(f) is exclusively modeled with ML algorithms (not gap-filled). All models used to model ET were trained in 2018 and 2020 and on the same tower for which data were predicted. White areas represent missing data (for (a) and (b)) and modeled negative values (for (c)-(f)).

3.5 DISCUSSION

3.5.1 ET IN AN URBAN ENVIRONMENT

Aligning with greening, carbon-neutrality, and sustainability plans of many cities worldwide and in light of the impact of climate change on the urban water cycle and urban heat, it is crucial to accurately characterize urban ET. In this study, we presented a novel approach combining footprint modeling, ML and DL, and remote sensing to accurately model urban ET at a half-hourly scale. Integrating GIS, remote sensing and meteorological datasets and using ML and DL techniques paves the way for new approaches in urban ET estimation and facilitates enhancing its accuracy and transferability. When available meteorological data is more limited (e.g., only

ETo data), the addition of remotely sensed satellite and GIS data is especially beneficial for ET modeling in an urban environment. Some urban areas, particularly in developing countries, have limited access to data such as solar radiation (Shafieiyoun et al., 2020; Shojaei et al., 2018), whereas medium-resolution satellite imagery is freely available worldwide (Claverie et al., 2018). Modeling with Penman-Monteith ETo still assumes extensive meteorological data is available for an urban area, which is not the case for data-scarce regions (Shafieiyoun et al., 2020). Therefore, future work should consider even more limited data scenarios.

We tested different predictors of urban ET on two EC towers located in different neighborhoods in Berlin. We found a combination of meteorological, remote sensing, and GIS data to be the optimal predictor scenario of urban ET. For the tower surrounded by more urban greenery, meteorological and GIS data shows an RMSE of 0.0239 mm/ hour and R^2 of 0.840 with RF and an RMSE of 0.0250 mm/ hour and a R^2 of 0.824 with 1D CNN. For the tower surrounded by a greater proportion of impervious cover, the same scenario shows an RMSE of 0.0170 mm/ hour and R^2 of 0.544 with RF and an RMSE of 0.0178 mm/ hour and R^2 of 0.502 with 1D CNN.

Our approach compares favorably with other modeling approaches for estimating hourly urban LE; we converted LE (W/ m^2) from other studies to ET (mm/ hour) for comparison. For instance, the Surface Urban Energy and Water Balance Scheme (SUEWS) was used to model LE on an hourly scale in Los Angeles, USA and Vancouver, Canada; an RMSE of 0.0295-0.0826 mm/hour and R^2 0.47-0.79 were reported (Järvi et al., 2011). SUEWS was applied to estimate urban LE at an hourly scale at two EC sites in the UK, a dense urban site in London and a suburban site in Swindon; an RMSE of 0.0364 mm/ hour and R^2 of 0.245 and an RMSE of 0.0333 and R^2 of 0.721 were reported, respectively (Ward et al., 2016). Three urban land-surface models (LSMs) were applied to model EC fluxes in a dense city center site and a suburban site in Helsinki, Finland (Karsisto et al., 2016). The LSMs showed an RMSE of 0.0168 mm/ hour (winter) to 0.0889 mm/ hour (summer) at the suburban site and an RMSE of 0.0226 mm/ hour (winter) to 0.0733 mm/ hour (summer) at the urban site (Karsisto et al., 2016). A micrometeorological approach (ARM) was applied to model LE in three temperate cities, finding a poor to moderate agreement with EC measurements; an RMSE of 0.0268-0.0450 mm/ hour and R^2 of 0.37-0.42, an RMSE of 0.0280 mm/ hour and R^2 of 0.09, and an RMSE of 0.0311 mm/ hour and R^2 of 0.07 were reported for Basel, Switzerland, London, UK, and Heraklion, Greece, respectively (Chrysoulakis et al., 2018).

3.5.2 MODELING ET

Comparing ML and DL algorithms used in our study, RF showed a slightly higher accuracy than CNN, although the performance of the algorithms was comparable. RF also outperformed ANNs and SVM in a study gap-filling methane fluxes (Kim et al., 2020). Kim et al. (2020) attributed the superior performance of RF for methane gap-filling to its capacity to incorporate a

variety of model inputs, including noise, while avoiding overfitting. 1D CNNs are nevertheless promising for hydrological modeling due to their high computational speed and capacity to extract temporal patterns (Ferreira and da Cunha, 2020; Haidar and Verma, 2018). As data-driven empirical models, the ML and DL algorithms needed to be trained at each location to perform optimally, shown by the increase in RMSE when training and testing on different towers. However, the similar R^2 values whether training and testing on the same or different towers show that the relationship between the predictors and ET are similar for both towers. The inclusion of remote sensing and GIS data opens up the possibility to fit general models incorporating the influence of surface cover.

3.5.3 ET DRIVERS IN AN URBAN ENVIRONMENT

Deeper investigation of predictors by variable importance analysis highlighted the contribution of remote sensing and GIS data to models, even though ET is primarily driven by meteorological variables. The most important GIS predictors based on RF were NDVI, ISF, and BH at ROTH and ISF, BH, and vegetation fraction at TUCC. Land cover may play a role in which predictors are most relevant. For instance, NDVI may be more relevant for modeling in a more vegetated urban area (e.g., ROTH), as NDVI has been shown to be very beneficial to modeling ET of urban vegetation (Boegh et al., 2009; Nouri et al., 2015). CNN models assigned lower importance to GIS predictors, with ISF as the most important GIS predictor at ROTH and BH as the most important GIS predictor at TUCC. Wind speed, air pressure, and solar zenith angle are highly relevant in ET estimation according to RF models. CNN models also emphasized a similar set of meteorological variables: shortwave downward radiation, solar zenith angle, diffuse solar radiation, and wind speed.

The ET sums showed a substantial difference between the less vegetated site and the more vegetated site, demonstrating how green infrastructure can substantially alter the urban water cycle. In summertime, monthly ET sums of a more highly vegetated site (ROTH) were double the ET sums of the less vegetated site (TUCC). ET at ROTH constituted a larger share of available precipitation (P) annually (72% of 506 mm P measured at Berlin-Dahlem) than at TUCC (56% of 400 mm P measured at Berlin-Tegel or 38% of 584 mm P measured at the TUCC flux tower), which can be attributed to the greater vegetation cover at ROTH. P was not calculated from the ROTH flux tower due to large data gaps in the dataset. A study in a suburban area (Swindon) in the UK estimated annual ET to be 370 mm using EC data (Ward et al., 2013), closely matching the annual ET sum at the more vegetated site (ROTH) in our study (366 mm). The study area in Swindon, UK was similar to our study site at ROTH, as their flux tower was also installed in a residential garden with significant surrounding vegetation cover (44% within a 500-m radius) (Ward et al., 2013). However, the ratio of ET to P annually at Swindon (57%) was lower than the ratio at ROTH (Ward et al., 2013). A study simulating annual ET in Copenhagen, Denmark with

an ecohydrological model estimated an annual ET of 210 mm, corresponding to 27% of P (Boegh et al., 2009); this ET sum is comparable to the value at the less vegetated site (223 mm), although the percentage of P is lower in Copenhagen than in both sites in this study. The site in Copenhagen was surrounded by 28% vegetation cover in the averaged daily flux footprint (Boegh et al., 2009), which is even lower than the fraction of vegetation cover at TUCC (33%). A source of uncertainty for ET sums are precipitation events, during which the EC system cannot accurately record fluxes. On average, modeled ET was lower during precipitation events, likely reflecting the reduced shortwave radiation during rain events (Appendix B; Figure B.4 and Table B.5). In reality, a rapid evaporation may occur directly following rainfall from rainwater that is intercepted by impervious surfaces or vegetation (Ward et al., 2013). Methods relying on open-path gas analyzers therefore underestimate ET sums (Ward et al., 2013).

3.5.4 FUTURE APPLICATIONS

The future of sustainable cities, elaborated in the sponge city or WSUD concept, hinges upon ecosystem services provided by green infrastructure (He et al., 2019; Liu and Jensen, 2018). ET is a key indicator of the ecosystem services provided by urban green spaces due to its role in the water balance and its cooling capacity; as such, urban initiatives are already explicitly aiming to enhance ET by integrating more vegetation, green facades, and green roofs into the urban landscape (Liu and Jensen, 2018). Expanding green infrastructure in cities, however, brings trade-offs in urban sustainability, as more water resources are used for landscape irrigation (Litvak et al., 2017; Nouri et al., 2019; Pataki et al., 2011). While urban water scarcity has mainly been the concern of (semi-)arid regions so far (Litvak et al., 2017; Pataki et al., 2011; Saher et al., 2021), extreme drought and heat waves in recent years in western and northern Europe have demonstrated that these challenges are also increasingly relevant for temperate cities suffering from summer heat waves such as Paris or Berlin (Dousset et al., 2011; Fenner et al., 2019; Gabriel and Endlicher, 2011).

The growing interest in managing and quantifying the impacts of climate change on cities therefore demands accurate ET modeling approaches (Boegh et al., 2009; Cong et al., 2017; Saher et al., 2021). The proposed methodology can be applied to simulate the effects of climate change and land use scenarios on urban ET, with implications for the urban energy and water balance. Furthermore, the presented approach can be used to upscale urban ET spatially at the city scale at a high resolution in the presence of EC and GIS data as the most important model inputs are freely available. Spatial upscaling can facilitate the exploration of how adding green infrastructure such as green roofs in cities can augment evaporative cooling and where to install them for the greatest benefit to urban residents (Besir and Cuce, 2018). Furthermore, large-scale urban land use change, such as the conversion of the former Berlin Tempelhof Airport to a built-up area, can be simulated to anticipate its effect on ET and its related ecosystem services. In future applications, urban ET

maps can be made publicly available to support initiatives aiming for sustainable irrigation management of urban green spaces and heat risk mitigation for urban residents. The presented approach can also be applied to accurately gap-fill urban EC flux data and produce long-term ET time series, which can be used to better understand ET seasonality and trends (Foltýnová et al., 2020).

Generating spatial predictions of ET simply requires predicting with trained models on a GIS database of all predictors, as has previously been applied for urban air temperature (Vulova et al., 2020). However, some challenges need to be resolved before upscaling. Extracted GIS data by footprint modeling is averaged over different spatial scales (Kotthaus and Grimmond, 2014). Upscaling spatially would thus require spatial aggregation of GIS data corresponding to an average footprint.

The ET dynamics of the less vegetated site (TUCC) need to be further investigated, particularly in relation to the influence of land cover. A study modeling LE from EC data at an hourly resolution with SUEWS also showed lower accuracy (higher RMSE and lower R^2) in modeling LE in a dense urban site compared to a suburban site surrounded by more vegetation (Ward et al., 2016). The lower range of ET values and therefore lower R^2 is expected in a site with less vegetation (Ward et al., 2016). Furthermore, at the more vegetated site (ROTH), footprint modeling revealed the expected positive correlation of vegetation cover and ET (Appendix B; Figure B.3). At the less vegetated site, however, the correlations were counter-intuitive, showing a negative relationship between ET and indicators of vegetation presence (vegetation fraction and vegetation height) (Appendix B; Figure B.3). As the TUCC tower is an inner-city tower situated on a building roof, it may be more affected by water vapor released through anthropogenic activities, which can confound the contribution of surrounding vegetation to ET (Karsisto et al., 2016; Kotthaus and Grimmond, 2012; Nordbo et al., 2012; Ward et al., 2013). In addition, observed ET from EC towers is not reliable during rain. In this study, four hours of data after precipitation were removed; however, the effect of increased ET by wet surfaces after rainfall may persist for more than 12 hours (Kotthaus and Grimmond, 2014). Advection effects, which are more common in a patchy urban landscape, may also affect EC fluxes (Kotthaus and Grimmond, 2014; Vesala et al., 2008). Kotthaus and Grimmond (2014) also found that turbulent fluxes in the center of London, UK could not be explained by surface cover types and suggested that other effects besides impervious surface cover fractions need to be considered to interpret urban fluxes.

The presented approach can be integrated with any remote sensing and GIS layers that may be relevant to modeling an EC flux, such as urban morphology and Leaf Area Index (LAI). The integration of footprint models with ML and DL is also relevant for other urban fluxes, such as

CO_2 (Crawford and Christen, 2015; Järvi et al., 2012; Kotthaus and Grimmond, 2014; Menzer et al., 2015). Future studies can apply this approach to other cities and fluxes.

3.6 CONCLUSIONS

Urban ET is a key aspect of urban greening efforts worldwide and therefore requires innovative methods to be accurately quantified. We presented a novel approach fusing flux footprint modeling, remote sensing and GIS data, and ML and DL to model urban ET at a half-hourly scale. Flux footprints allow for land cover characteristics to be incorporated into ML and DL algorithms, which is essential to estimating ET in heterogeneous urban terrain. We compared two DL and ML algorithms (1D convolutional neural networks (CNNs) and random forest (RF)). Although RF showed a slightly higher accuracy, the performance of CNNs was also promising and warrants further exploration with other model architectures.

Four predictor scenarios were tested to assess the contribution of remote sensing and GIS data to model performance. We found incorporating remote sensing, GIS, and meteorological predictors to be the best-performing scenario to estimate urban ET. This predictor scenario with RF showed an RMSE of 0.024 mm/ hour and R^2 of 0.84 and an RMSE of 0.017 mm/ hour and R^2 of 0.54 for the more vegetated and less vegetated site, respectively. NDVI and impervious surface fraction emerged as the most important GIS predictors, while solar zenith angle, shortwave downward radiation, wind speed, and air pressure were the most important meteorological predictors.

Future applications of this methodology include gap-filling ET in order to analyze long-term trends, simulating the influence of altering urban land cover on ET, and spatially upscaling ET to the city scale with a high spatial resolution. The presented method can support sustainable urban planning efforts in the face of climate change, including initiatives to manage blue and green water resources and to mitigate the urban heat island effect.

Chapter 4: Mapping hourly urban evapotranspiration using Sentinel-2, open geodata and machine learning

This study is in review:

Vulova, S., Rocha, A.D., Meier, F., Nouri, H., Schulz, C., Soulsby, C., Tetzlaff, D., Kleinschmit, B., 2022. Mapping hourly urban evapotranspiration using Sentinel-2, open geodata and machine learning. Under review: Remote Sensing of Environment.

This is a preprint version of the article.

GRAPHICAL ABSTRACT
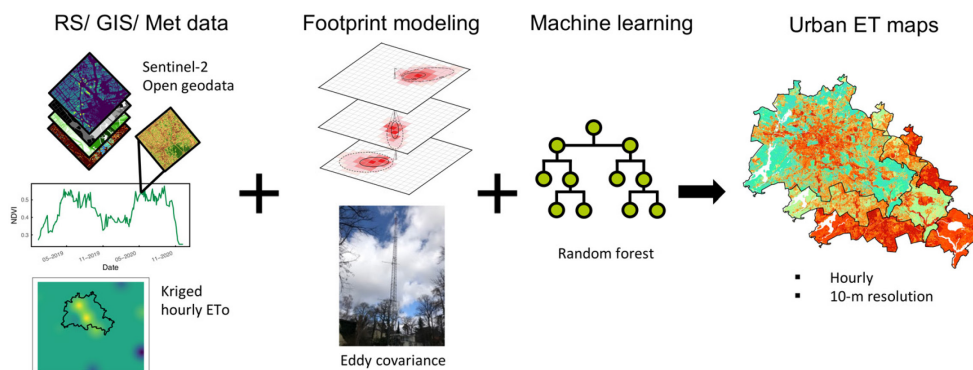
ABSTRACT

The impacts of global change, including extreme heat and water scarcity, are threatening an ever-growing urban world population. Evapotranspiration (ET) mitigates the urban heat island, reducing the effect of heat waves. It can also be used as a proxy for vegetation water use, making it a crucial tool to plan resilient green cities. To optimize the trade-off between health, urban greening and water security, reliable and up-to-date maps of ET for cities are urgently needed. Despite its importance, few studies have mapped urban ET accurately for an entire city in high spatial and temporal resolution. We mapped the ET of Berlin, Germany in high spatial (10-m) and temporal (hourly) resolution for the year of 2019. A novel machine learning (ML) approach combining Sentinel-2 time series, open geodata, and flux footprint modeling was applied. Two eddy flux towers with contrasting surrounding land cover provided the training and testing data. Flux footprint modeling allowed us to incorporate comprehensive land cover types in training the

ML models. Open remote sensing and geodata used as model inputs included Normalized Difference Vegetation Index (NDVI) from Sentinel-2, building height, impervious surface fraction, vegetation fraction, and vegetation height. NDVI was used to indicate vegetation phenology and health, as plant transpiration contributes to the majority of terrestrial ET. Hourly reference ET (ETo) was calculated and used as input to capture the temporal dynamics of the meteorological conditions. Predictions were carried out using random forest (RF) regression. Weighted averages extracted from hourly ET maps using flux footprints were compared to measured ET from the two flux towers. Validation showed that the approach is reliable for mapping urban ET, with a mean R^2 of 0.76 and 0.56 and a mean RMSE of 0.0289 mm and 0.0171 mm at the more vegetated site and the city-center site, respectively. Lastly, the variation of ET between Local Climate Zones (LCZs) was analyzed to support urban planning. This study demonstrated the capacity to map urban ET at an unprecedented high spatial and temporal resolution with a novel methodology, which can be used to support the sustainable management of green infrastructure and water resources in an urbanizing world facing climate change.

4.1 INTRODUCTION

Climate change impacts are fundamentally threatening the way of life of residents in many urban areas, where risks of heat waves, flooding, and water scarcity are exacerbated (Kundzewicz et al., 2014; Li and Bou-Zeid, 2013; Manoli et al., 2019). These hazards are leading to an increase in casualties, adverse health effects, and infrastructure damage (Kundzewicz et al., 2014; Mora et al., 2017). As a growing majority of the world population resides in cities (United Nations, 2019), urban sustainability measures may have the greatest impact on human quality-of-life (Acuto et al., 2018; Gaffin et al., 2012). To meet UN Sustainable Development Goal 11 of making cities more sustainable and resilient, urban planners and policy-makers urgently need to adopt mitigation measures grounded in evidence-based understanding of urban (eco)systems (Bayulken et al., 2021).

Greening cities to adapt to present and future environmental risks is therefore growing in importance worldwide, as epitomized by the frameworks of Water Sensitive Urban Design (WSUD) (Fletcher et al., 2008), Sponge Cities (Gaines, 2016), Nature-Based Solutions (Somarakis et al., 2019), Low-Impact Development (LID) (Qin et al., 2013), and Sustainable Drainage Systems (Zhou, 2014). Evapotranspiration (ET) of green infrastructure plays a central role in all of these concepts (Bayulken et al., 2021; He et al., 2019). Enhancing ET reduces urban heat (Manoli et al., 2019) and associated health burdens and mortality (Cuthbert et al., 2022; Hiemstra et al., 2017), stormwater runoff and flooding risk (Berland et al., 2017; Jongen et al., 2022), cooling demand (Coccolo et al., 2018), and carbon emissions (Scott et al., 2006).

ET is a key indicator for urban sustainability, as it is associated with thermal comfort (Shashua-Bar et al., 2011; Wang et al., 2019), water storage capacity (Jongen et al., 2022), the

irrigation demand of green spaces (Litvak et al., 2017; Saher et al., 2021), and stormwater runoff reduction (Berland et al., 2017; Wan et al., 2021). An improved understanding of urban ET and its spatio-temporal patterns is essential to future-proof cities worldwide against climate risks (Berland et al., 2017; Saher et al., 2021). Sustainability initiatives in urban areas need to be adapted to a neighborhood scale relevant for vulnerable socio-demographic groups (Hölzl et al., 2021). Therefore, for sustainable management of resilient cities, accurate, high-spatial and - temporal resolution maps of ET on a city scale are needed.

However, urban greening programs still require better data and models to assess the value of urban vegetation (Pataki, 2013). In recent years, machine learning (ML) has emerged as one of the most promising methods to extract knowledge from big Earth system data, which comprises a variety of remote sensing and *in situ* sensors (Reichstein et al., 2019). Regression ML algorithms have been especially successful in predicting the spatio-temporal variation of biogeophysical parameters by linking point observations with satellite data (Reichstein et al., 2019). Data-driven methods provide numerous advantages, including the capacity to discover patterns and infer from data without process-based formulation (Kuhn and Johnson, 2013; Reichstein et al., 2019).

Despite the relevance of ET estimation for urban greening, urban ET has rarely been modeled and even more rarely mapped (Nouri et al., 2015; Saher et al., 2021; Vulova et al., 2021). Characterizing ET is always challenging, but in particular in an urban environment due to a high heterogeneity of land cover, plant traits, soil characteristics, and microclimate (Nouri et al., 2013; Saher et al., 2021). Therefore, process-based models have mainly been used to model urban ET, with the most common types being surface energy balance (SEB) models, hydrological models, and urban land-surface models (ULSMs) (Järvi et al., 2011; Karsisto et al., 2016; Nouri et al., 2015; Rafael et al., 2020; Saher et al., 2021; Ward et al., 2016). As the majority of these models have been developed for agricultural and natural areas, their applications in urban areas have been limited in reliability and transferability (Nouri et al., 2015; Saher et al., 2021). Medium-resolution earth observation data, such as Landsat and MODIS, have been integrated into SEB models to improve urban ET mapping (Cong et al., 2017; Faridatul et al., 2020; Jiang and Weng, 2017). However, SEB models still do not perform well in characterizing urban ET, as they exclude anthropogenic heat, assume a single plant species, and are optimized for a regional scale (Saher et al., 2021). Furthermore, physical modeling approaches to quantifying ET developed for urban areas commonly rely on parameters which are impractical to estimate on the city scale at a high resolution, such as soil water potential, wall emissivity, and heat emissions from vehicles (Meili et al., 2020; Saher et al., 2021). For urban planning, regularly updated city-level ET maps at a hyperlocal resolution (10-30 m) (Venter et al., 2020) are needed to assess high-risk areas related to extreme heat and stormwater runoff and to propose integrated solutions, such as street trees and green roofs (Pataki, 2013; Wan et al., 2021). The few studies mapping urban ET so far have

been limited in temporal resolution, extent (Rocha et al., 2022b; Vulova et al., 2021), spatial resolution (Saher et al., 2021; Wang et al., 2016), or accuracy (Chrysoulakis et al., 2018; Grimmond et al., 2010; Rafael et al., 2020; Wan et al., 2021). ML, in combination with open-access remote sensing data and *in situ* measurements, offers a way forward to accurately mapping spatio-temporally dynamic ET in urban areas (Reichstein et al., 2019; Vulova et al., 2021).

Flux towers are increasingly providing direct measurements of ET in cities worldwide (Jongen et al., 2022), opening up the possibility to model urban ET empirically. However, to the best of our knowledge, only one study applied an empirical model to map ET in an urban area so far (Wang et al., 2016). Wang et al. (2016) related a buffer average extracted from two MODIS products (albedo and land surface temperature) around a single flux tower site in Phoenix, USA using ordinary least squares (OLS) regression. This regression model was used to map daily ET at a 500-m spatial resolution (Wang et al., 2016). However, the capacity of mapping ET in a city using a data-driven approach was limited by the lack of ET data from various urban land covers (Wang et al., 2016). To address this issue, Vulova et al. (2021) developed a novel approach to model urban ET using flux footprint modeling to incorporate a variety of land cover into artificial intelligence algorithms. However, this methodology was only applied to model ET at two points in a city (Vulova et al., 2021).

Here, we map urban ET at the city scale at a high spatial (10-m) and temporal (hourly) resolution using a novel data-driven approach. We integrated the influence of heterogeneous land cover on urban ET into ML models by combining open-access remote sensing data and flux footprint modeling (Vulova et al., 2021). City-wide hourly ET maps were generated for the entire year of 2019, which were aggregated to daily, monthly, and annual ET maps. Hourly maps for a 2 km^2 extent around the flux towers for two years (2019-2020) were used for validation with flux tower measurements at two sites in Berlin, Germany. The city-wide ET maps were further analyzed using Local Climate Zones (LCZs), a classification developed to better understand urban heat island (UHI) effects, to explore the implications for urban planning. Our key objectives are to: 1) model and map the spatio-temporal variability of urban ET and 2) assess the impact of LCZs on urban ET.

Such high-resolution, spatially explicit ET products in cities are essential to unraveling the role of vegetation in providing cooling services and the presence of water in increasingly water-scarce urban landscapes. These ET maps can therefore support sustainability measures in cities, especially greening and heat risk adaptation schemes.

4.2 METHODS

4.2.1 EVAPOTRANSPIRATION MAPPING

We mapped the ET of Berlin at a high temporal (hourly) and spatial (10-m) resolution using ML, open data available for most large- and medium-size cities worldwide, including remote sensing imagery, geodata, and meteorological data, and eddy covariance (EC) data. An overview of the methodology is provided in Figure 4.1. We integrated the variability of urban land cover into ML models by estimating the extents of constantly-changing footprints of two flux towers using flux footprint modeling (Kormann and Meixner, 2001). Random Forest (RF) models were trained in one year, while testing and mapping was performed with the other year (e.g., training in 2020 and testing in 2019) and vice versa. Measured ET from the two flux towers was used to estimate the model accuracy. We validated our mapping approach by extracting footprints from hourly ET maps for the entire study period (2019-2020). We then aggregated the city-wide hourly ET maps from 2019 to generate daily, monthly, and annual maps. Finally, as ET varies greatly among different land surfaces, we assessed the variation of ET between LCZs.
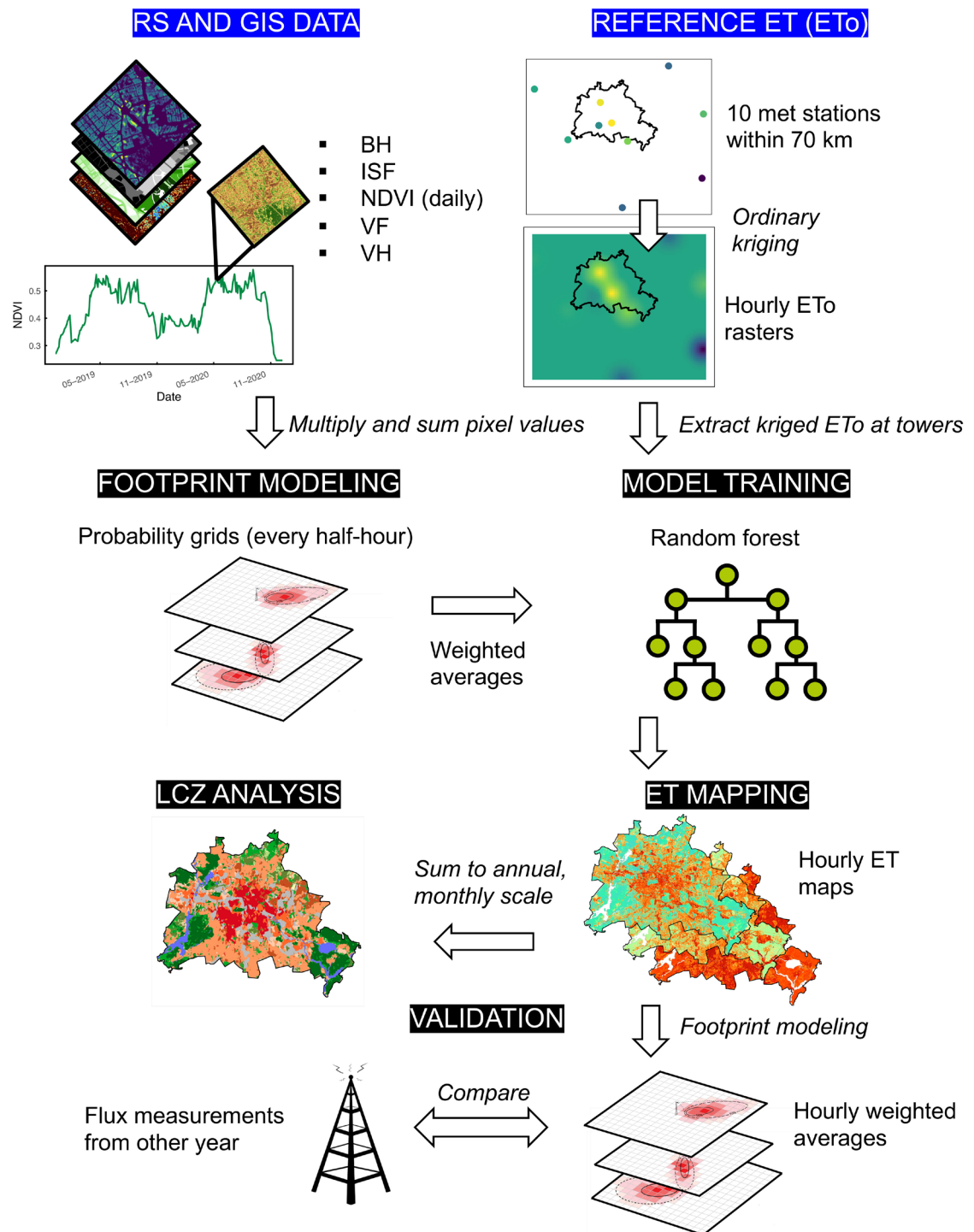
Figure 4.1: Flowchart showing the workflow of the study, including the input data used for ET mapping (in blue), footprint modeling, model training, and validation. Abbreviations can be found in the text. The flowchart is adapted from Vulova et al. (2021), with the footprints diagram adapted from Christen (2016).

4.2.2 STUDY AREA

ET was mapped for the entire city of Berlin, the largest and capital city in Germany with 3.7 million inhabitants and 891 km^2 (Statistical Office of Berlin-Brandenburg, 2019) (Figure 4.2b). The city is characterized by 45% urban green spaces and water bodies. Building density decreases from the city center to the border. Berlin has a temperate oceanic and a humid continental climate with warm summers (Beck et al., 2018) and a flat topography. Berlin has been identified as a forerunner city worldwide in the use of green infrastructure for sustainable water management and heat mitigation (Liu and Jensen, 2018). Daily air temperature (Tair), shortwave downward radiation, precipitation (P), and NDVI during the study period (2019-2020) are shown in Figure C.1 (Appendix C). Both years were warmer than the long-term (1981-2010) mean (9.7 °C), with an average Tair of 11.8 °C and 11.7 °C in 2019 and 2020, respectively (DWD, 2021) and drier than the long-term mean (591 mm) (DWD, 2021), with a P of 520 mm and 479 mm in 2019 and 2020, respectively.

4.2.3 FLUX MEASUREMENTS

Two flux towers, which are part of the Urban Climate Observatory (UCO) Berlin (Scherer et al., 2019), were used for training and testing. The surroundings of the towers contain heterogeneous vegetation (grassland, shrub, and trees) and impervious cover representative of land cover in temperate-zone cities (Gillefalk et al., 2021; Vulova et al., 2021). However, the Rothenburgstrasse (ROTH) tower is surrounded by more green spaces (with 51.1% vegetation surface cover within a 1-km radius) than the TU Berlin Campus Charlottenburg (TUCC) tower (34.4% vegetation surface cover within a 1 km radius). The flux towers and their surroundings are described extensively in Vulova et al. (2021), Jongen et al. (2022), and Gillefalk et al. (2021).

The EC instrumentation, data quality control (QC) and preprocessing are elaborated in Vulova et al. (2021). All half-hourly ET observations (in mm/ h) remaining after QC were used for training, consisting of 16,011 and 16,255 observations (45.7% and 46.3% of the raw EC data) combined from both towers for 2019 and 2020, respectively. The source area contributing to turbulent flux measurements by EC towers, referred to as a flux footprint, constantly changes based upon wind speed, wind direction, and atmospheric stability (Kormann and Meixner, 2001). We used the Kormann and Meixner (2001) model to estimate the footprints for each half hour during the study period. The parametrization of footprint models is described in Vulova et al. (2021) and Duarte Rocha et al. (2022b).

4.2.4 METEOROLOGICAL DATA

Reference ET (ETo) was selected as a predictor to synthesize the meteorological conditions regulating ET (Allen et al., 2005). Using ETo instead of individual meteorological datasets has been shown to model urban ET accurately while being accessible even in data-scarce regions of the world (Vulova et al., 2021). We used open-access hourly meteorological data from the German Meteorological Service (DWD) (DWD, 2021) to calculate ETo. From all the DWD stations within 70 km of Berlin, 10 stations were selected based upon the availability of all the data needed to compute ETo (air temperature, relative humidity, wind speed, and sunshine duration). Solar radiation was estimated from sunshine duration using the extraterrestrial equation from the MeTo package (ra function) (Dettmann and Grimma, 2019) due to only one DWD station within 70 km having solar radiation data. ETo was computed using the hourly ASCE "Standardized Reference Evapotranspiration Equation" for short crops (Allen et al., 2005). For further details on the calculation of ETo, refer to Vulova et al. (2021). ETo was then linearly interpolated to a half-hourly scale for training to match the temporal resolution of the EC data. As spatially continuous rasters of all predictors are necessary for ET mapping, ETo was interpolated with ordinary kriging at a 1 km resolution using the R package "gstat" (Gräler et al., 2016). For training, ETo was extracted from the kriged rasters at the coordinates of the two towers.

4.2.5 REMOTE SENSING AND GIS DATA

As predictors for ET mapping, we used freely-available remote sensing and GIS data (Table 4.1). Normalized Difference Vegetation index (NDVI) (Tucker, 1979) is an indicator of vegetation health, density, and biophysical characteristics which is well-suited for modeling urban ET (Nouri et al., 2015; Saher et al., 2021; Vulova et al., 2021). We used the Google Earth Engine platform (Gorelick et al., 2017) to calculate and download NDVI raster time series from Sentinel-2 Level-2A products at a 10-m resolution (image collection "COPERNICUS/S2_SR"). Atmospheric, terrain, and cirrus correction of Sentinel-2 images was performed using the Sen2Cor atmospheric correction processor. Pixels remaining after cloud masking were interpolated to a daily scale. Urban Atlas 2012 (European Environment Agency, 2018) provided building height at a 10-m resolution. Vegetation height (VH) was provided by the Berlin Digital Environmental Atlas at 1-m resolution (Berlin Senate Department for Urban Development and Housing, 2014). Vegetation fraction (VF) was provided by the Berlin Digital Environmental Atlas (Berlin Senate Department for Urban Development and Housing, 2014). Impervious surface fraction (ISF) was derived from the Urban Atlas 2018 Land Use and Land Cover (LULC) dataset (European Environment Agency, 2018), as explained in Vulova et al. (2020). ISF and VF, which were available as polygons, were rasterized to a 1-m resolution. Finally, all geodata were resampled to a common (10-m) resolution before footprint modeling and ET mapping. Water bodies were excluded from footprint modeling and ET mapping, as there were nearly no water

bodies present in the training regions (footprints). Weighted averages of the surface cover were then computed by multiplying the footprint grids with the raster layers and summing the pixel values on a half-hourly basis. The weighted averages of surface cover extracted with footprint modeling were used to train the ML models, which were then used to predict ET for all pixels within the city border.

Table 4.1: Predictor variables used to model and map evapotranspiration (ET); the data source, provider, and their original spatial and temporal resolutions are listed. In the "Temporal resolution" column, "static" refers to geodata which does not vary temporally (such as impervious surface fraction).

Predictor (acronym)	Unit	Data source	Data provider	Original spatial resolution (m)	Temporal resolution
Building height (BH)	m	Urban Atlas 2012	Copernicus Programme, European Commission (European Environment Agency, 2018)	10	Static
Impervious surface fraction (ISF)	%	Urban Atlas 2018 (LULC)	Copernicus Programme, European Commission (European Environment Agency, 2018)	Polygons*	Static
Normalized Difference Vegetation Index (NDVI)	-	Sentinel-2 Level-2A surface reflectance	European Space Agency (ESA) (ESA, 2015), Google Earth Engine (Gorelick et al., 2017)	10	Daily
Reference evapotranspiration (ETo)	mm/h	DWD	DWD Climate Data Center (CDC) (DWD, 2021)	1000	Hourly
Vegetation fraction (VF)	%	Berlin Digital Environmental Atlas	(Berlin Senate Department for Urban Development and Housing, 2014)	Polygons*	Static
Vegetation height (VH)	m	Berlin Digital Environmental Atlas	(Berlin Senate Department for Urban Development and Housing, 2014)	1	Static

*In the "Spatial resolution" column, "polygons" refers to datasets originally provided as vector data, which were then rasterized to 1-m resolution.

4.2.6 MACHINE LEARNING APPROACH

To map ET, we used the Random Forest (RF) ML algorithm, a supervised regression model which combines classification and regression trees (CART) with bagging (Breiman, 2001). RF

was selected as it can replicate highly nonlinear relationships, is robust against overfitting (Kuhn and Johnson, 2013), and has been shown to perform well in modeling urban ET (Vulova et al., 2021). Detailed descriptions of RF and its implementation using the "caret" R package are provided in Kuhn and Johnson (2013) and Vulova et al. (2021). RF models were trained using half-hourly data (the resolution of EC data and footprint modeling) combined from both towers in order to include a greater variety of combinations of land cover covered by footprints and meteorological conditions.

4.2.7 VALIDATION

We mapped ET on an hourly basis based upon the highest temporal resolution of predictors. For testing, ET was mapped on an hourly basis for 2 km x 2 km (the likely maximum coverage of flux footprints) areas centered around the flux towers for 2019-2020. To perform validation on an hourly basis, the quality-controlled EC data was averaged to an hourly resolution, omitting hours where one or no half-hourly observations were available. Weighted averages from the 2 km x 2 km ET maps were extracted using flux footprints modeled with hourly averaged EC data to compare modeled with measured ET. To evaluate the temporal extrapolation capacity of our models, we used 2020 data for training and 2019 data for testing and vice versa. In total, 6696 and 6771 hourly ET observations combined from both towers were used for testing in 2019 and 2020, respectively. The prediction accuracy was evaluated using root mean square error (RMSE), percent bias (pbias), coefficient of determination (R^2), and normalized root mean square error (NRMSE).

4.2.8 LOCAL CLIMATE ZONE ANALYSIS

LCZs refer to a classification system initially developed to provide a universal framework for UHI studies, consisting of 17 classes of which 10 are urban-specific (Stewart and Oke, 2012). Each LCZ represents a distinct thermal climate regime based on surface structure and land cover (Stewart and Oke, 2012). The main advantage of the LCZ scheme is comparability, where a LCZ in Tokyo, Delhi, or Paris refers to the same urban form and cover (Demuzere et al., 2021; Stewart and Oke, 2012). LCZs are also advantageous for guiding urban planners, architects, and engineers to construct climate-adapted cities due to simple and clear definitions of the design elements of each LCZ (e.g., building height, building density, and vegetation cover ratio) (Stewart and Oke, 2012). LCZ maps are freely-available for global cities, all of Europe, and the United States of America (Demuzere et al., 2021).

To assess and demonstrate the variation of ET for different LCZs across Berlin, we used a LCZ map of Berlin produced by the LCZ Generator (Demuzere et al., 2021) using training data provided by Fenner et al. (2017) (Figure 4.2b). The LCZ Generator allows users to incorporate their local knowledge to map cities into LCZs in an online platform, with >1000 open-access LCZ

maps of cities worldwide already posted by users (Demuzere et al., 2021). The LCZ map is representative for the years 2017-2019 based on the pre-processed earth observation data used to map LCZs (Demuzere et al., 2021). LCZ maps are generally provided at 100-m resolution (Demuzere et al., 2021), as LCZs are defined as "regions of uniform surface cover, structure, material, and human activity that span hundreds of meters to several kilometers in horizontal scale" according to Stewart and Oke (2012). As individual pixels do not comprise an LCZ class, a Gaussian-filtered version of the LCZ map, which reduces granularity, was used. LCZs were converted to polygons to extract summary statistics and produce boxplots from the ET maps. Medians, rather than means, were used to describe differences between LCZs, as the mismatch between the spatial resolution of our ET maps and LCZs (10-m vs. 100-m) introduces more noise in the arithmetic mean.

4.3 RESULTS

4.3.1 MAPPING URBAN EVAPOTRANSPIRATION

Annual estimated ET ranged from 119 mm to 648 mm (mean: 300 mm), with strong spatial patterns dependent on the land cover (Figure 4.2a). Annual ET was highest in Tempelhof Field park, the city's largest open green space (Figure 4.2d), in the NE of the city, which consists of cropland and pasture (Figure 4.2f), and smaller green spaces such as community gardens and parks. Areas of low ET were concentrated in highly impervious areas (Figure 4.2c and e), such as streets, in the city-center and the East.
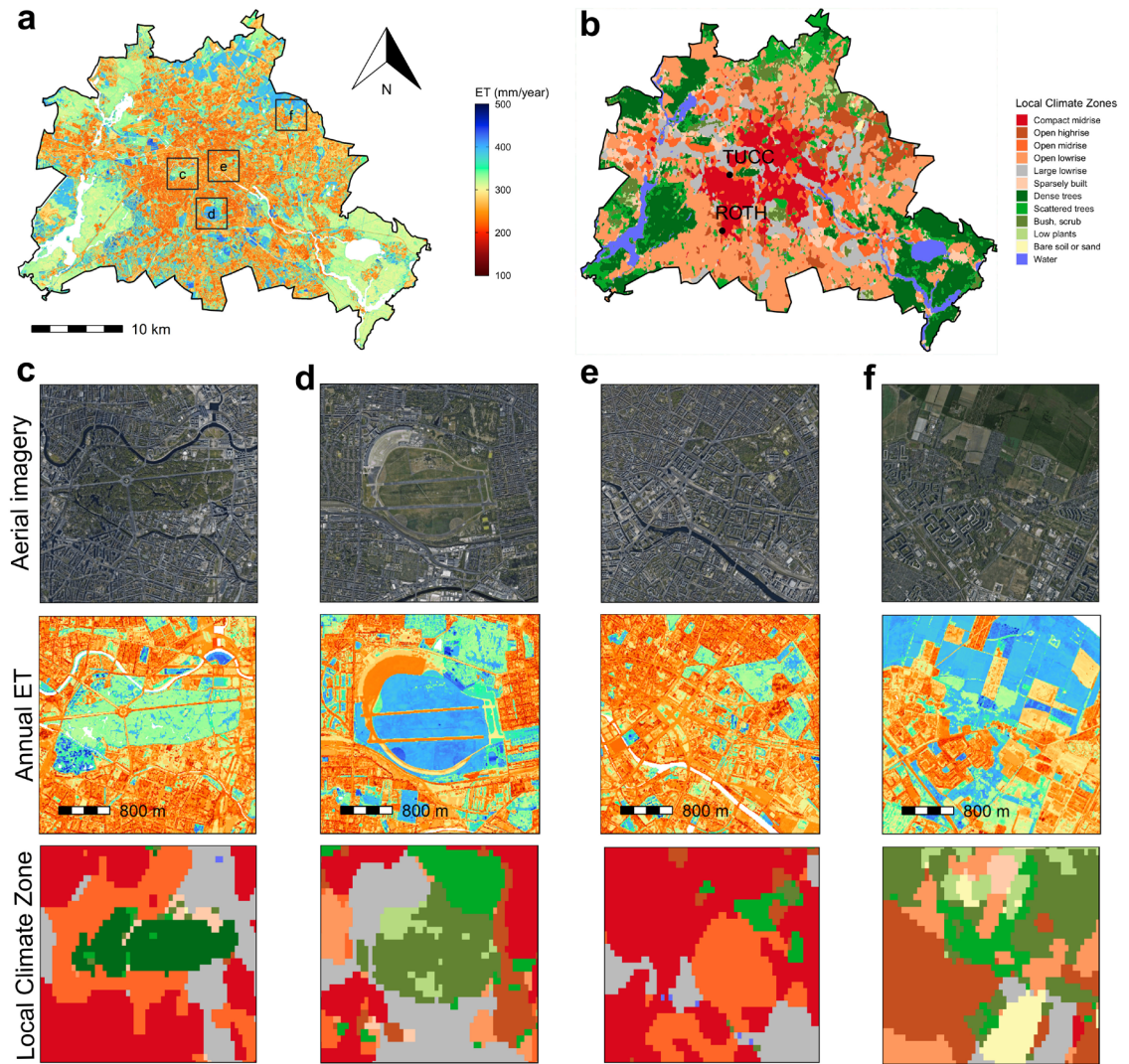
Figure 4.2: Maps of a) annual evapotranspiration (ET) in 2019 and b) Local Climate Zone (LCZ) coverage in Berlin, Germany. Annual ET and LCZs are depicted for zoomed extents around c) a large urban forest (Tiergarten) and d) the largest park, dominated by low vegetation (Tempelhof Field), e) a typical inner-city neighborhood with mid-rise buildings (Karl-Marx-Allee (KMA) neighborhood), and f) an area with croplands and residential buildings on the periphery of the city, with high-resolution aerial imagery (source: Google; ©2022 CNES / Airbus, GeoBasis-DE/BKG, GeoContent, Landsat/ Copernicus, Maxar Technologies) for reference. White colored areas in the annual ET maps represent water bodies, which were excluded from ET modeling.

The estimated monthly ET maps captured the phenology of the urban vegetation (Figure 4.3). Forests in the SE and SW of the city showed very low ET in wintertime, which increased considerably during leaf out (April). Highest ET occurred in June (mean: 44 mm, max: 80 mm), while lowest ET occurred in January (mean: 11 mm, max: 39 mm).
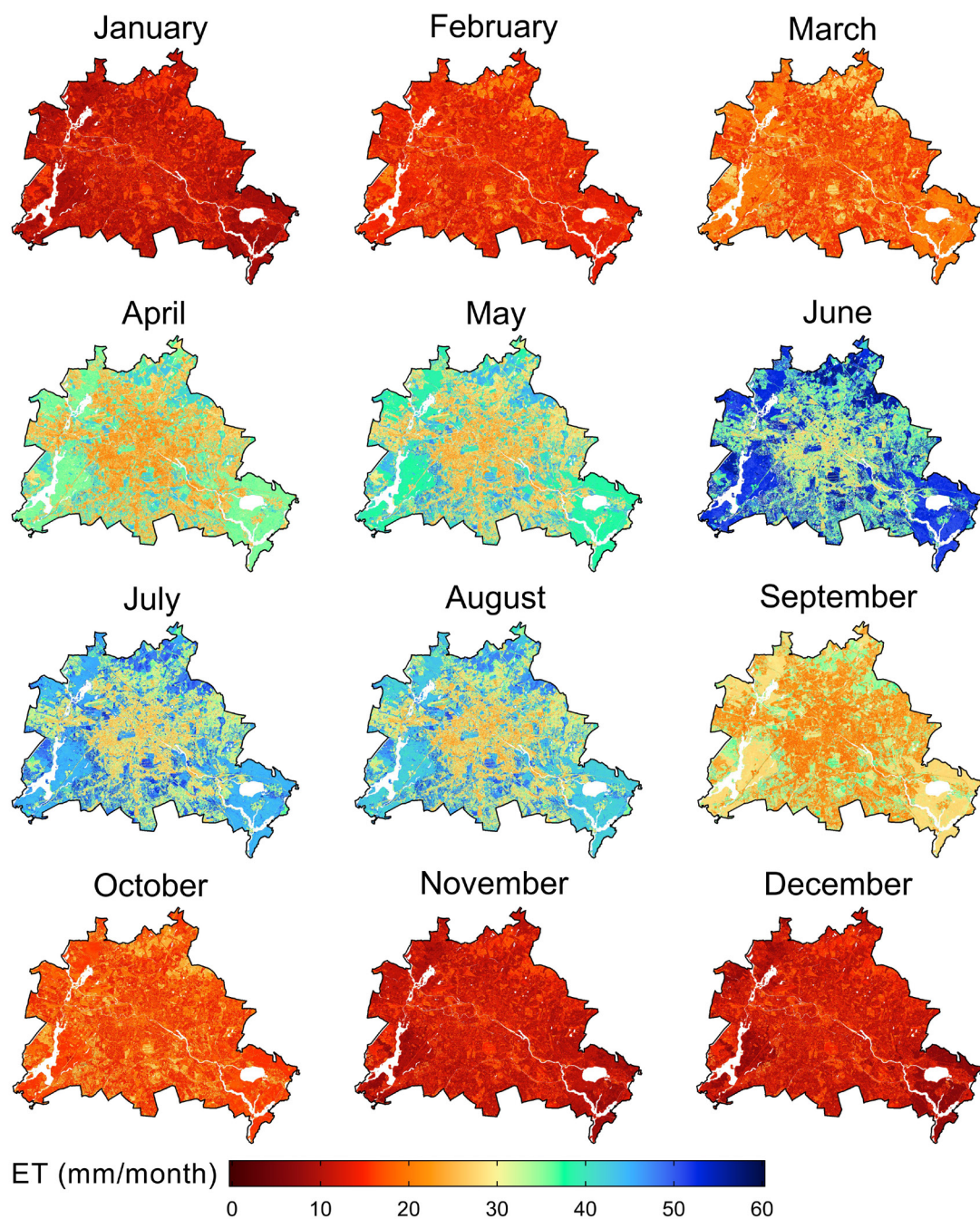
ET (mm/month)

Figure 4.3: Maps of monthly evapotranspiration (mm/ month) in 2019 in Berlin, Germany.

4.3.2 EVALUATION

The performance metrics when testing for each year and flux tower are shown in Table 4.2. On average, validation at the greener site (ROTH) showed a higher R^2 (0.76) than at the less vegetated site (TUCC; 0.56). On the other hand, RMSE was higher at the greener site (0.0289 mm) than at the city-center site (0.0171 mm). The higher R^2 and RMSE when testing at the site surrounded by more vegetation (ROTH) can largely be explained by the higher range and

maximum ET measured at ROTH compared to TUCC (Kuhn and Johnson, 2013; Rocha et al., 2022b; Vulova et al., 2021). Normalized RMSE (NRMSE) allows us to compare the accuracy between the two sites with a metric that removes the influence of the range of measured ET. The difference between NRMSE between the two towers was minimal (11.2% vs. 11.9% at ROTH and TUCC, respectively). ET was overestimated by a greater magnitude at the more built-up site (TUCC; pbias = 28.7%) than at the more vegetated site (ROTH; pbias = 3.1%). Training models using data from both towers led to a higher overestimation of ET at the dense city-center site where lower ET values were observed (TUCC).

Overall, averaged across the two sites, accuracy was slightly higher in 2019 ($R^2 = 0.67$) than in 2020 ($R^2 = 0.65$). ET was on average more overestimated in 2020 (pbias = 21.2%) than in 2019 (pbias = 10.6%). Training in 2019 and predicting in 2020 likely led to the overestimation as the previous year had higher air temperature (Tair) and precipitation (P) in summertime (Tair: 21.5 and 20.3 °C, P: 194 and 126 mm).

Table 4.2: Performance metrics for validation at two flux towers and two years. The performance metrics are root mean square error (RMSE), percent bias (pbias), coefficient of determination (R^2), and normalized root mean square error (NRMSE). The best performance metrics are shown in bold.

Tower (test)	Year (test)	RMSE (mm/h)	pbias (%)	R^2 (-)	NRMSE (%)
ROTH	2019	0.0277	**-0.2**	**0.772**	**11.0**
	2020	0.0301	6.3	0.748	11.3
TUCC	2019	**0.0161**	21.4	0.567	11.3
	2020	0.0181	36.0	0.553	12.4

To better understand the temporal variation of the mapping error, the difference between the modeled and measured ET in mm/h (residuals) was calculated and plotted as a time series (Figure 4.4). At the more vegetated site, residuals were positive in summertime for both testing years, indicating that models were underestimating ET. In contrast, at the dense city-center site, modeled ET was not underestimated in summertime, with near to zero or negative residuals. In April, RF models overestimated ET for both testing years and both sites, likely due to a time lag between the rising NDVI signal in spring and the contribution of plant leaf out to transpiration. At the same flux sites in Berlin, ET was also overestimated in April using a process-based model due to a sharp increase in vegetation greenness, which did not yet correspond to higher ET (Rocha et al., 2022b).
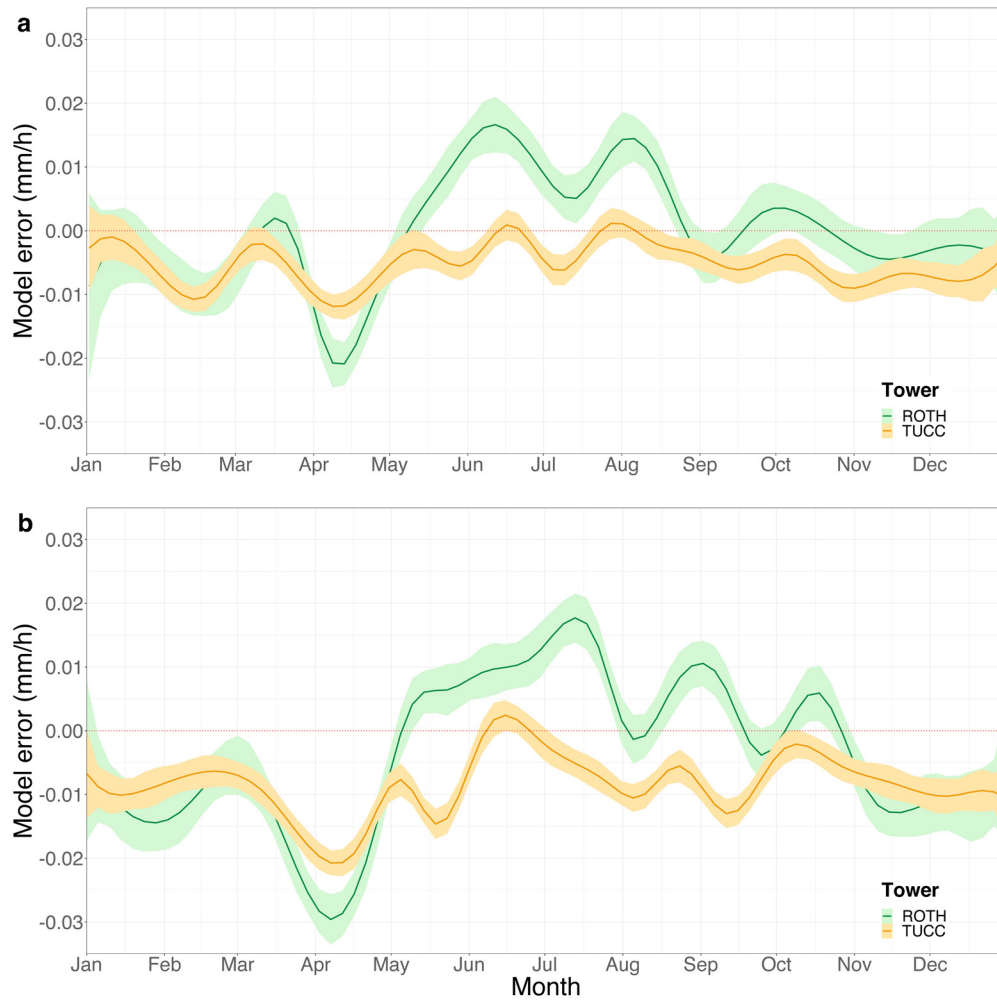
Figure 4.4: Smoothed time series of the hourly model error (observed ET - predicted ET) for the two flux towers when testing in (a) 2019 and (b) 2020. Smoothing function (formula = y~splines::bs(x,20)).

4.3.3 LOCAL CLIMATE ZONE ANALYSIS

We analyzed the variability of urban ET within different LCZ classes both annually and in summertime (Figure 4.5). ET varied highly between both the vegetated and urban LCZs, with the highest and lowest ET estimated for "low plants" and "compact midrise", respectively.

LCZ "low plants" showed the highest median ET both annually (390 mm) and in summertime (50 mm), closely followed by "bush, scrub" for both time periods. "Dense trees" had the lowest median ET (317 mm annually, 46 mm in summertime) of the natural LCZs, other than the non-vegetated "bare soil or sand" LCZ. Our study cannot fully account for interception loss, which may explain the lower estimated ET for trees. EC data during and 4 hours after precipitation events is unreliable and is therefore removed, which does not allow for the evaporation from intercepted rainwater to be characterized (Vulova et al., 2021; Ward et al., 2013).

ET also showed high variation between the urban-specific LCZs. Among the urban LCZ classes, "sparsely built" showed the highest median ET both annually and in summertime (335 mm and 47 mm). "Compact midrise" showed the lowest median ET both annually and in summertime (246 mm and 29 mm).

In a few cases, the ranking of LCZs differed between annual and summer ET. For instance, "large lowrise" showed a higher median ET than "open midrise" annually, while the opposite was the case in summertime. In winter months, ET was higher in "large lowrise", a highly paved LCZ, than in "open midrise" (Appendix C; Figure C.3). In addition, "bare soil or sand" showed higher ET than "open lowrise" annually, while the relationship was reversed in summertime. This discrepancy can also be attributed to the higher ET of "bare soil or sand" than "open lowrise" in wintertime (Appendix C; Figure C.3). Our approach can still capture the influence of interception loss occurring more than 4 hours after rainfall. Thus, these differences can be explained by evaporation from interception loss, which is generally higher in wintertime and over impervious surfaces (Miralles et al., 2020; Ramamurthy and Bou-Zeid, 2014).
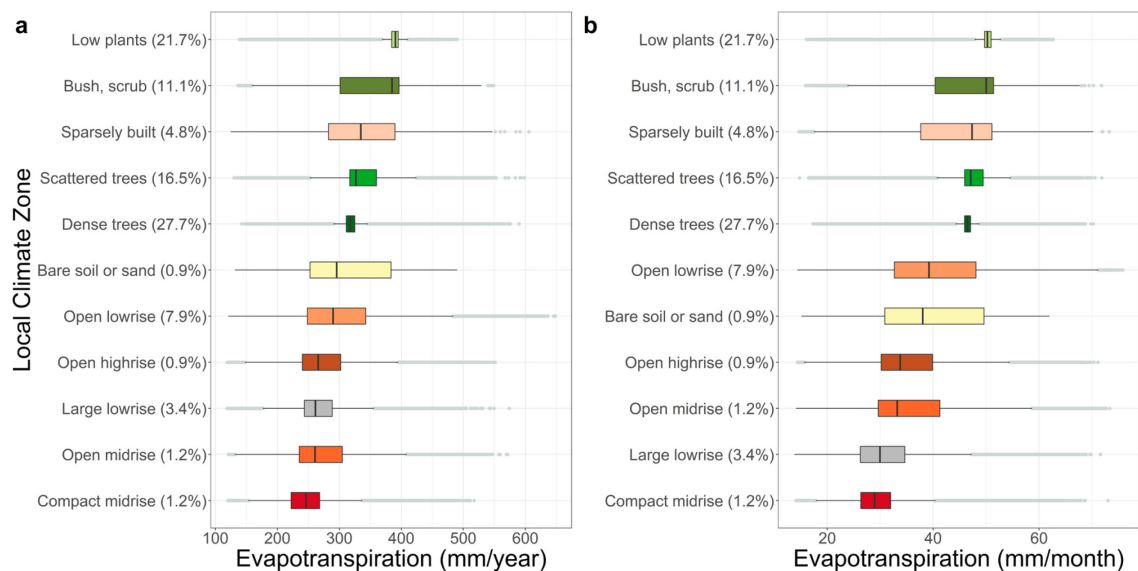
Figure 4.5: Local Climate Zones (LCZs) and modeled (a) annual ET (mm/ year) and (b) monthly ET (mm/ month) averaged in summertime (June, July, and August) over Berlin in 2019. The LCZs are ordered in decreasing order of median summertime ET. For the boxplots, the center line represents the median, the upper and lower hinges represent the 25th and 75th percentile, the upper whisker extends from the hinge to the largest value no further than 1.5 * interquartile range (IQR), the lower whisker extends from the hinge to the smallest value no further than 1.5 * IQR, and the light gray dots represent the outliers (points beyond the whisker range).

4.4 DISCUSSION

Mapping ET is highly relevant to making cities climate-resilient as more of the world population is exposed to extreme heat, flooding, and drought (Kundzewicz et al., 2014; Li and Bou-Zeid, 2013; Mora et al., 2017). In this study, we demonstrated the capacity of an open-source data-driven approach to accurately map urban ET at a high spatial and temporal resolution.

In most cases, urban ET modeling approaches can only be validated by flux tower measurements at one to two points in the city (Chrysoulakis et al., 2018; Karsisto et al., 2016), as was the case in our study. Our modeling approach showed a comparable or higher accuracy than the majority of studies modeling hourly urban ET or LE. Our ET maps were validated at a more vegetated site and a dense city-center site, showing a R^2 of 0.76 and 0.56, respectively. For urban land surface models, the R^2 ranges from 0.017 to 0.79 (Järvi et al., 2011; Karsisto et al., 2016; Rafael et al., 2020; Ward et al., 2016). A micrometeorological approach to mapping LE in three European cities showed R^2 ranging from 0.07 to 0.42 (Chrysoulakis et al., 2018). As in our study, higher R^2 values have been reported for greener sites compared to dense inner-city sites (Rafael et al., 2020; Rocha et al., 2022b; Vulova et al., 2021; Ward et al., 2016), which is likely due to the higher variation of ET at more vegetated sites.

ET maps can support the management of urban green spaces to mitigate heat risk for urban dwellers. An established trade-off exists between the capacity of UGI to provide ecosystem services, in particular evaporative cooling, and the irrigation needed to maintain this cooling capacity (Cuthbert et al., 2022; McCarthy and Pataki, 2010; Nouri et al., 2019). As urban areas are increasingly faced with the double burden of heat waves and droughts, climate change adaptation with UGI may exacerbate water scarcity (Cuthbert et al., 2022; Miller et al., 2020). This trade-off can be visualized on a map showing the ET anomaly between the hottest day in 2019 and the average daily summertime ET (Figure 4.6), with the increase in ET varying with the vegetation health, type, and water availability. In Tiergarten, which is dominated by tree vegetation, ET is ~30% higher than average, enhancing the microclimate where heat risk is highest (Figure 4.6b) (Dugord et al., 2014). In Tempelhof Field, a large open green space, the increase in ET is more differentiated, being higher (30-50 %) in the south of the meadow but lower in the middle (~20%) (Figure 4.6c). The Karl-Marx-Allee (KMA), an inner-city neighborhood (above the river, Figure 4.6d), is a hotspot for thermal stress, while housing a high proportion of elderly residents (Knaus and Haase, 2020). Notably, the ET anomaly in the KMA area (up to 100%) is even higher than in conventional green spaces, likely due to irrigation maintaining the vitality of the vegetation. Figure 4.6e shows a mixture of agricultural fields and residential buildings, with an industrial area showing the highest ET anomaly (>60%). Our mapping approach reveals how the cooling capacity of green spaces responds to drought stress, which can better inform UGI implementation for climate mitigation.
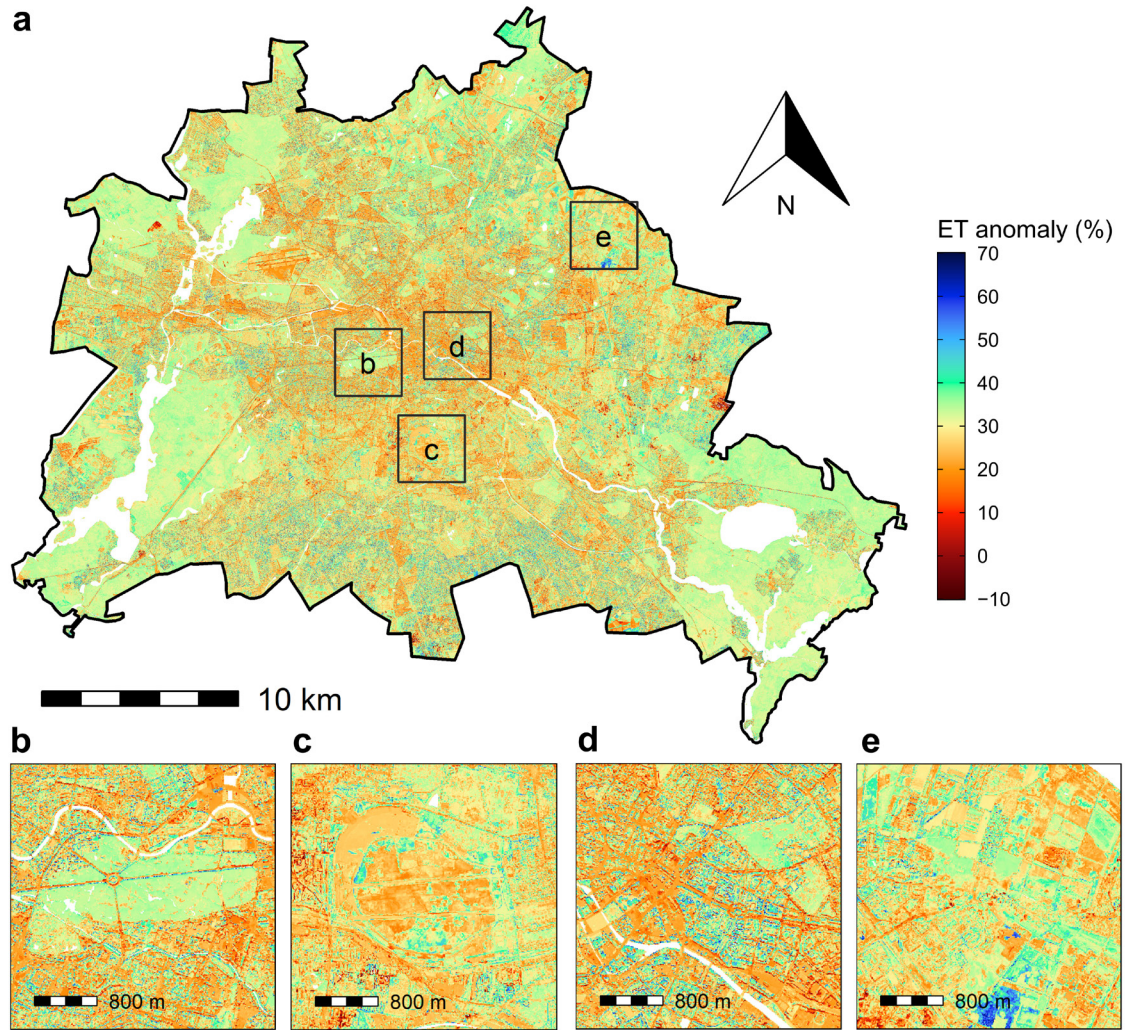
Figure 4.6: Map of the anomaly between the daily evapotranspiration (ET) (mm/ day) of a heatwave date (26 June 2019) and the average daily ET in summertime in 2019. The ET anomaly is depicted for a) the entire city and zoomed extents around b) a large urban forest (Tiergarten) and c) the largest park, dominated by low vegetation (Tempelhof Field), d) a typical inner-city neighborhood with mid-rise buildings (Karl-Marx-Allee neighborhood), and e) an area with croplands and residential buildings. White colored areas represent water bodies, which were excluded from ET modeling.

Grasslands and shrublands, which showed the highest ET in our study, may provide a critical evaporative cooling service. On average, grasslands and shrubs converted more of annual precipitation to ET (72% and 68%) than dense trees (61%). Forests have been commonly assumed to convert more of precipitation into ET due to their higher interception, higher leaf area, and deeper root systems (Teuling, 2018; Williams et al., 2012). However, flux measurements have shown that grasslands consume more precipitation as ET than forests in a global study of 167 sites (Williams et al., 2012), in sites across Europe in summertime (Teuling et al., 2010), and in three sites in northern Germany (Markwitz et al., 2020). Field measurements in semi-arid urban landscapes have confirmed that irrigated turfgrass lawns have substantially higher rates of ET

than trees (Litvak et al., 2014) and reduce air temperatures during both daytime and nighttime (Gómez-Navarro et al., 2021). A remote sensing-based study in Los Angeles, USA also found that increasing turfgrass cover resulted in a greater cooling effect based on land surface temperature (LST) than the same increase in tree fraction (Wetherley et al., 2018). Furthermore, a study investigating the effects of drought on urban vegetation in California using repeat airborne imaging spectroscopy showed that turfgrass recovered more quickly from a multi-year drought than trees despite irrigation restrictions (Miller et al., 2020).

The high ET rate of grasslands may be problematic when precipitation is limited, requiring scarce water resources to maintain ecosystem service provision of these green spaces. Trees remain an essential component of adapting cities to extreme heat due to shading, a lower need for irrigation, and evaporative cooling (Gómez-Navarro et al., 2021; Teuling et al., 2010; Wang et al., 2019; Wong et al., 2021). Interspersing trees and bushes in meadows can be a 'win-win' solution, lowering water demand from grass by shading while maximizing UHI mitigation (Gómez-Navarro et al., 2021).

Nevertheless, ET is more commonly stated to be higher in trees rather than grasslands or shrubs when proposing greenery well-suited to mitigation of urban heat (Gillefalk et al., 2021; Wan et al., 2021; Wong et al., 2021). The lower ET in trees estimated in our study should be cautiously interpreted, as our methodology, like most EC-based studies, cannot estimate interception loss occurring a few hours after rainfall, which is especially high in trees (Gillefalk et al., 2021; Kuhlemann et al., 2021; Vulova et al., 2021; Ward et al., 2013). A paradox has been identified across studies quantifying ET rates, with EC measurements generally showing higher ET rates in grasslands and lysimeters in forests (Teuling, 2018). Forest ET may be underestimated by flux towers due to the role of advection and mesoscale circulation (Teuling, 2018). Teuling (2018) also noted that trees reaching a high ET rate is dependent on their age, height, and a dense forest canopy, which may limit tree ET rates in urban areas. Furthermore, much of the area classified as low plants or bushes in Berlin is in fact cropland or pasture. As agricultural vegetation is irrigated and fertilized to optimize yield, it is generally greener and produces more ET; however, low plants and shrubs in non-irrigated green spaces are unlikely to present such high rates of ET.

Our study also provided insights on the variation of ET between urban LCZs. Among the urban-specific LCZs, "sparsely built" showed the highest ET, followed by "open lowrise" and "open highrise", both annually and in summertime. These findings are in line with previous studies, which found that lower development intensity is associated with higher annual ET (Liu et al., 2010; Wan et al., 2021). "Large lowrise", a mostly paved LCZ, showed relatively higher ET rates in wintertime than more vegetated LCZs (Appendix C; Figure C.3), which can be attributed to higher rates of evaporation from interception loss from impervious surfaces (Miralles

et al., 2020; Ramamurthy and Bou-Zeid, 2014). Our results emphasize the need to map ET at different temporal scales, rather than only annually, to better characterize its influence on heat mitigation and the urban hydrological balance.

One advantage of a data-driven mapping approach is its capacity to uncover patterns not properly characterized by physical models (Reichstein et al., 2019), including the rooting depth, which may be more developed in grasslands than previously assumed (Williams et al., 2012) or the higher interception loss over paved surfaces in wintertime. Another benefit of using a data-driven approach is the flexible integration of diverse and novel datasets (Kuhn and Johnson, 2013). For instance, urban microclimate can be integrated into ET mapping in future work using crowdsourced weather data (Venter et al., 2020; Vulova et al., 2020). The presented methodology can be also used to map other fluxes measured by eddy covariance, such as CO_2 (Kotthaus and Grimmond, 2014), at a high spatio-temporal resolution by combining remote sensing and GIS data, flux footprint modeling, and ML.

LST was not included as a predictor in this study due to the lower temporal (16-day repeat cycle) and spatial resolution (100 m) of thermal infrared (TIR) imagery from the Landsat 8 satellite compared to Sentinel-2 imagery. Assessing the effectiveness of downscaling Sentinel-3 TIR imagery from 1000 m to 10 m resolution using Sentinel-2 imagery in urban areas can allow for LST to be added as a predictor in future implementations (Guzinski and Nieto, 2019). However, since sharpening low-resolution TIR imagery is unlikely to be sufficient in highly heterogeneous areas such as cities (Guzinski and Nieto, 2019), the launch of high-resolution TIR sensors, such as the Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment (TRISHNA) mission (Vidal et al., 2022), will be most beneficial for accurate ET estimation.

There are a few limitations associated with the presented data-driven method. Our approach is driven by the input data, which allows transferability to similar land cover and meteorological conditions to those used in training. On the other hand, re-training with local flux data may be necessary to achieve acceptable accuracy, especially in other climate zones. Our maps underestimated the highest ET in summertime, with underestimation most likely occurring in pure vegetated areas such as forests, due to flux footprints capturing the mixed influence of impervious and vegetated surface cover during training (Figure 4.4). In future work, the variability of ET within a city can be better characterized by incorporating training data from flux towers in other cities in similar climatic zones (Jongen et al., 2022) and in natural areas underrepresented by urban flux towers, such as forests (Markwitz et al., 2020). High ET values may also be underestimated (Appendix C; Figure C.2) as most of the training observations represent low ET (e.g., nighttime and wintertime). To reduce this effect, ML models can be trained separately for summer in daytime in future implementations. The transferability of ET mapping with and

without flux tower data for training can be tested with the increasingly available eddy covariance data in cities worldwide (Jongen et al., 2022) or other ground-truth datasets (Nouri et al., 2013).

Although eddy covariance (EC) is one of the best-suited methods to directly measure urban ET (Nouri et al., 2013), uncertainties characteristic of EC data (Feigenwinter et al., 2018) introduce some inaccuracies into our ET maps. Furthermore, EC systems cannot accurately record fluxes during and a few hours after rainfall events, necessitating the exclusion of this data during quality control (Vulova et al., 2021; Wouters et al., 2015). Training with EC data therefore underestimates evaporation from interception, especially for trees and impervious surfaces (Ward et al., 2013). However, our ET maps are still appropriate for mitigation of thermal stress, as interception loss is not associated with the cooling effect of transpiration (Miralles et al., 2020; Rocha et al., 2022b). Our mapping approach is especially accurate for identifying hotspots for low thermal comfort, which is most relevant for urban planners aiming to mitigate the UHI effect.

4.5 CONCLUSION

Our study successfully mapped ET at the city scale at a high spatio-temporal resolution using a machine learning approach in combination with remote sensing. Urban ET maps contribute essential understanding to urban greening, heat-mitigation, water-saving, and stormwater reduction measures, which potentially buffer the most people against climate change impacts in an increasingly urbanized world (Gaffin et al., 2012). An interaction between increasingly frequent and intense heat waves (Meehl and Tebaldi, 2004) and UHI effects from expanding urbanization (Li and Bou-Zeid, 2013; Manoli et al., 2019) will place unprecedented pressure on the health, livability, and energy use in cities. Our study provides a methodology for mapping ET at a high spatial and temporal resolution to guide urban planning initiatives aiming to reduce heat risks. The effects of drought on the capacity of urban green infrastructure to provide evaporative cooling are still largely unknown (Miller et al., 2020). However, mapping long-term time-series of urban ET in future work can provide insights on balancing water-saving and greening measures. This approach can be adapted by urban planners to map ET under various greening and climate-change scenarios.

DATA AND CODE STATEMENT

The code used for this study is publicly and permanently available on Zenodo at https://doi.org/10.5281/zenodo.6521095. The output data from footprint and machine learning modeling are available on request from the first author. The raw eddy covariance data belongs to the Urban Climate Observatory (UCO) Berlin and are available on request from Dr. Fred Meier (fred.meier@tu-berlin.de). The meteorological (DWD) data are freely available at https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/ and can also

be downloaded using our open code (https://doi.org/10.5281/zenodo.6521095). The source code to compute and download our NDVI time series using Google Earth Engine is also provided in the same Zenodo publication. Urban Atlas products (building height and Land Use and Land Cover (LULC)) are publicly available at https://land.copernicus.eu/local/urban-atlas. The code to derive ISF from Urban Atlas LULC is provided in the Zenodo publication. VH and VF are freely available from the Berlin Digital Environmental Atlas (https://fbinter.stadt-berlin.de/fb/index.jsp). The LCZ map used in this study is freely available at https://lcz-generator.rub.de/factsheets/6d7f501c212dc888e32a3a9a5740ce930addd3bc/6d7f501c212dc888e32a3a9a5740ce930addd3bc_factsheet.html.

ACKNOWLEDGEMENTS

Chapter 5: Synthesis and Conclusions

Worldwide, ever-expanding cities and human settlements are threatened with the exacerbated effects of global and climate change, including heat-related mortalities, water scarcity, flooding, and air pollution (Kundzewicz et al., 2014; Li and Bou-Zeid, 2013; Manoli et al., 2019; Wilby, 2007). Despite their high relevance to the well-being and thermal comfort of urban inhabitants, ET and Tair are still not well-characterized in urban areas (Nouri et al., 2015; Saher et al., 2021; Zhou et al., 2019). The emergence of big data at a high spatial resolution, which includes satellite imagery, volunteered geographic information, and various sensors, has opened up the possibility to develop strategies for climate change mitigation at a neighborhood scale (Milojevic-Dupont and Creutzig, 2021). Extracting information from big Earth system data has been facilitated by artificial intelligence (AI), whose rise is fueled by advances in computational power and computer science theory (Milojevic-Dupont and Creutzig, 2021; Reichstein et al., 2019).

The choice of AI algorithm is less important than the understanding and selection of data to feed into it. Modeling thermal comfort indicators in urban areas, especially with AI approaches, remains a research frontier in its infancy, with some of the very first attempts presented in this thesis. The recent emergence of the necessary data (crowdsourced weather data, urban flux tower data) has opened up the possibility to model Tair and ET using ML regression models. Until recently, however, this opportunity has been hampered by challenges in processing and aligning the data spatially and temporally with remote sensing imagery.

To address the existing research gap, this thesis had two main objectives. The first objective focused on developing workflows to modeling and mapping thermal comfort indicators using AI, remote sensing, and open geodata, combined with in situ sensor data, at a high spatial resolution:

Objective 1: Investigate how accurately thermal comfort indicators in urban areas (ET and Tair) can be modeled at a high spatial resolution with remote sensing imagery, geodata, and meteorological data using ML regression algorithms (Chapters 2-4).

This thesis provided a methodology for spatially predicting nocturnal Tair one day in advance in urban areas, which can be used as a blueprint for a warning system for heat risk for vulnerable residents (Chapter 2). Next, the capacity to model urban ET, a key indicator for climate

change mitigation and greening, accurately at a high temporal (Chapters 3 and 4) and spatial (Chapter 4) resolution was demonstrated using a novel approach which exploited the dynamic source area of eddy covariance measurements. The modeling approaches were tested in Berlin, the largest city in Germany by area (891 km^2) and in the European Union by population (3.7 million) within city limits (Statistical Office of Berlin-Brandenburg, 2019). By assessing the spatial variation of the modeled ET and Tair across Berlin, the second objective was to provide insights for urban planning and policy focused on heat risk mitigation (Chapters 2-4), which is addressed in Section 5.1.3.

The following sections summarize the main findings of this thesis (Section 5.1) and provide a perspective on the remaining challenges and opportunities in this research field (Section 5.2). Specifically, how innovative applications of AI (Section 5.2.1) and remote sensing (Section 5.2.2) could further improve the understanding of Tair and ET in urban areas in the future is discussed. Lastly, Section 5.2.3 provides an outlook for how the insights from this thesis could be integrated into urban planning.

5.1 MAIN FINDINGS

5.1.1 MODELING TAIR IN AN URBAN ENVIRONMENT

Heat extremes in cities are profoundly and detrimentally affecting urban residents and ecosystems, driving heat-related mortalities (Dousset et al., 2011; Scherer et al., 2013), respiratory and cardiovascular health issues (Kovats and Hajat, 2008), enhanced air pollution (Koken et al., 2003; Krug et al., 2019), higher energy use and carbon emissions (Santamouris and Kolokotsa, 2015; Tooke et al., 2014), and unforeseen shifts in urban biodiversity (Turrini and Knop, 2015). With the high human and financial cost of urban heat, mapping hotspots at a high resolution is vital. Thus, the first question of this thesis was:

Research Question I: How can open source remote sensing, GIS and crowdsourced weather data be used to predict the spatial distribution of nocturnal Tair in an urban area? (Chapter 2)

In this thesis, the capacity of spatially modeling Tair by combining crowdsourced weather data, open-source remote sensing data and geodata, and machine learning was demonstrated for nearly the first time (Chapter 2; a similar study by Venter et al. (2020) was conducted in parallel). Tair was predicted for summer nights due to their higher relevance for human health and mortality and more pronounced UHI effect (Christen and Vogt, 2004; Gabriel and Endlicher, 2011; Kovats and Hajat, 2008). Tair was predicted and mapped one day in advance at a 30-m resolution, using Tair deviation from a standard meteorological station located centrally in Berlin. Combining remote sensing and GIS data with crowdsourced data was generally found to be beneficial.

Crowdsourced weather data is certainly useful as a source of training and validation data for Tair modeling; Chapter 2 also tested if integrating kriged crowdsourced data from one day in advance as a predictor would improve the Tair mapping accuracy. While remote sensing and GIS data performed better when Tair was more spatially variable, better reflecting the fine-scale Tair variation due to urban structure and land cover, the crowdsourced Tair data could more accurately predict its distribution one day in the future during stable weather conditions due to temporal autocorrelation. Combining diverse data as predictors allows for models to "hedge their bets" for situations where either the urban form and cover or previous meteorological conditions may play a larger role in determining Tair distribution. Based on independent reference Tair data, however, the scenario using only remote sensing and GIS predictors was most accurate, so integrating crowdsourced weather data from a previous time step is not imperative.

This thesis not only demonstrated that Tair can accurately be modeled with the proposed workflow, but contributed insights on which data and preprocessing were especially important for an optimal model performance. Morphometric parameters, which reflect the 3D structure of the urban environment and can easily be derived from building height data, were especially important in modeling Tair. The land cover of an area surrounding each weather station influences the Tair measurement (Guo and Moore, 1998; Ho et al., 2014; Zakšek and Oštir, 2012); however, which radius can best approximate this influence in an urban environment still remains an open question (Ho et al., 2016; Venter et al., 2020). Therefore, a range of radii (0 - 1500 m) was tested for the spatial averaging of 13 remote sensing and GIS predictors; recommendations for optimal buffers for each type of geodata are provided based on the correlation with the crowdsourced Tair measurements. The workflow demonstrated to model Tair can be implemented in platforms at the intersection of the Smart City and Green City concepts, which could warn residents and policymakers about upcoming or real-time heat stress risk.

5.1.2 MODELING AND MAPPING URBAN ET

To plan greener and more climate-resilient cities, it is critical to characterize ET in urban areas at a scale relevant for urban planning. In this thesis, a novel open-source data-driven approach to modeling urban ET at the city-scale at a high spatio-temporal resolution was developed and validated.

Interfaces play a critical role in the urban water cycle, serving as key sites of fluxes which affect the entire system and its behavior (Gessner et al., 2014). ET is one of the fundamental parameters characterizing the soil-vegetation-atmosphere interface, as it constitutes both an energy flux contributing to the cooling capacity of a surface and a water flux which can facilitate the estimation of available water for human consumption in cities (Damm et al., 2018; Gessner et al., 2014). Thus, this thesis contributed to a better understanding of one of the main urban water

interfaces. Integrative sustainability approaches aiming to simultaneously mitigate UHI and manage water resources more ecologically, such as Nature-Based Solutions (Somarakis et al., 2019), will benefit from an improved characterization of urban ET.

ET is highly dependent not only on meteorological conditions, but also on the underlying land cover, which represents a complex mosaic of materials, vegetation types and 3D structure in an urban environment (Nouri et al., 2015; Rocha et al., 2022b, 2022a). Thus, the capability of modeling urban ET in a data-driven way has mainly been limited by the availability of in situ ET data representing the diversity of urban land cover, with most cities having one to two flux towers (Jongen et al., 2022). At the same time, the interpretation of the signal of ET measured by flux towers has been confounded by the constantly-changing source area (footprint) of eddy covariance measurements over a high urban heterogeneity (Christen et al., 2011; Kotthaus and Grimmond, 2014; Menzer et al., 2015). To simultaneously tackle both of these challenges, a method was developed to incorporate the diverse land cover contributing to urban ET measurements into AI algorithms by combining flux footprint modeling and remote sensing and GIS data (Chapters 3 and 4). Thus, the second research question was:

Research Question II: Is there a benefit to integrating remote sensing and GIS data extracted by flux footprints to model urban ET at a high temporal (half-hourly) resolution? (Chapter 3)

It was hypothesized that integrating remote sensing data by footprint modeling can allow for a better representation of urban ET than only modeling with meteorological data (Chapter 3). Integrating the characteristics of the land cover extracted by flux footprints, which were modeled on a half-hourly basis, into ML and DL algorithms enhanced the modeling accuracy over scenarios only relying on meteorological data. This result demonstrated the capacity of footprint modeling, combined with a geospatial database, to represent the influence of urban land cover on ET variation, paving the way for applying the methodology to upscale ET to the city-scale.

In the next phase, the data-driven approach was further developed to map ET for an entire city at a high spatial (10-m) and temporal (hourly) resolution for one year (Chapter 4). Thus, the third research question focused on the capacity of this methodology to map urban ET:

Research Question III: Can ET be accurately mapped at a high spatial (10-m) and temporal (hourly) resolution in an urban environment with remote sensing imagery, geodata, and meteorological data using ML and flux footprint modeling? (Chapter 4)

Mapping ET required several key modifications to the method introduced in Chapter 3, including using spatially distributed meteorological data, combining the data from the two flux towers to create a general model, and validating the result in a spatially explicit manner. By validating the spatial product with flux footprints, the developed approach was found to be more accurate or similar in accuracy when compared to the few existing approaches used to model hourly urban ET (Chapter 4). The ET maps for Berlin were further assessed using Local Climate Zones (LCZs), a land cover classification designed for UHI and urban planning applications (Stewart and Oke, 2012). The variation of modeled ET based on land cover was found to be in line with previous studies (Liu et al., 2010; Wan et al., 2021). The implications of this thesis's findings for urban planning are discussed in the next section.

5.1.3 IMPLICATIONS FOR URBAN PLANNING

With the world's cities facing accelerating climate and global change effects, it is crucial to supply urban planners with high-resolution spatial data (Milojevic-Dupont and Creutzig, 2021; Pataki, 2013). To move in the direction of sustainable and green cities, this work provided insights on how urban greening can be used to mitigate heat risk based upon high-resolution maps of urban Tair and ET, addressing Objective 2 and Research Question 4:

Objective 2: Assess the spatial variation of the modeled ET and Tair across Berlin to provide insights for urban planning and policy focused on heat risk mitigation

Research Question IV: What is the spatial variation of modeled ET and air temperature across Berlin and what implications does this variation have for urban planners?

In Chapter 4, LCZs were used to investigate the spatial variation of ET in Berlin, indicating that grasslands and shrubs may be especially significant sources of ET. In order to synthesize the Tair and ET results in answering Research Question 4, an additional analysis of the relationship between LCZs and nocturnal Tair in summertime (Chapter 2) was undertaken. The maps of the four study dates modeled in Chapter 2 were averaged and boxplots (Figure 5.1) are used to portray the variation of Tair per LCZ. Comparing the ranking of LCZs for summertime ET (Figure 4.5b) and Tair (Figure 5.1a), there is a good correspondence, with the LCZs with highest ET generally showing lowest nocturnal Tair. Based on the scenario with "all" predictors, "low plants" and "shrubs" are the coolest LCZs at nighttime. Land cover with scattered or dense trees nevertheless promotes higher ET and lower nocturnal Tair than more LCZs with impervious cover. Thermal comfort is most threatened in the "compact midrise" LCZ, which shows lowest ET and highest

nocturnal Tair. More open arrangements of buildings, such as the "open highrise" LCZ concentrated in East Berlin, allow for higher ET and lower Tair due to an abundance of vegetated land cover.
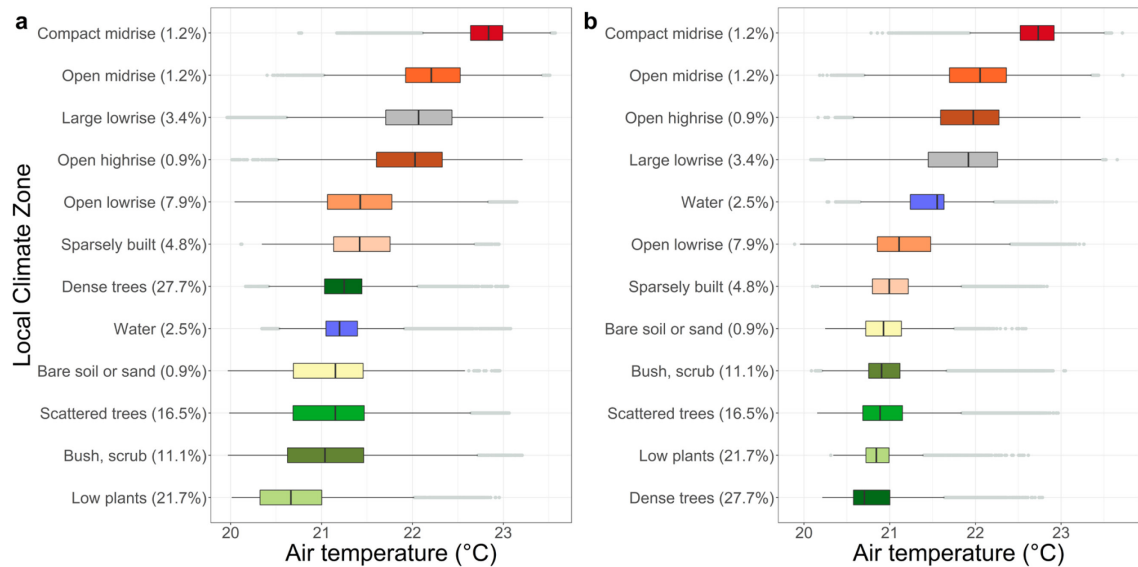
Figure 5.1: Local Climate Zones (LCZs) and nighttime Tair modeled with (a) all predictors and (b) only remote sensing and GIS predictors. The four summertime study dates mapped in Chapter 2 over Berlin using Random Forest were averaged for each predictor scenario. The LCZs are ordered in decreasing order of median Tair. For the boxplots, the center line represents the median, the upper and lower hinges represent the 25th and 75th percentile, the upper whisker extends from the hinge to the largest value no further than 1.5 * interquartile range (IQR), the lower whisker extends from the hinge to the smallest value no further than 1.5 * IQR, and the light gray dots represent the outliers (points beyond the whisker range).

However, the model configuration, such as choice of predictors, can have a large influence on the modeled spatial distribution of thermal comfort indicators. While the "All" scenario, which includes crowdsourced Tair from the day before as a predictor, shows that low plants exhibit the lowest Tair, in the "GIS" scenario dense trees have the lowest Tair (Figure 5.1). The models' incapacity to consistently and accurately characterize Tair over vegetated areas is largely due to the placement of "Netatmo" stations, which are mostly installed in built-up areas and close to building walls (Fenner et al., 2017; Meier et al., 2017). Subsequently, vegetated land cover, especially meadows, showed the lowest accuracy in Chapter 2, with highly sealed urban fabric showing the highest accuracy. This limitation could be overcome if research institutes install a denser network of Netatmo stations in vegetated urban areas, which is feasible considering their low cost (Chapman et al., 2015; Venter et al., 2020). The mapping approaches presented in Chapters 2 and 4 are nevertheless highly relevant for urban planning focused on thermal comfort, as they are most accurate where heat risk is highest.

To compare the spatial distribution of ET and nocturnal Tair, the city-wide maps for a date occurring during a major 2019 European wave (25 June) are shown in Figure 5.2. Despite showing two different phenomena, the two maps are moderately (negatively) correlated (Pearson's r = -0.47). More vegetated areas exhibit higher ET and lower Tair. For instance, in the NE of Berlin, an area dominated by crops and pastures, particularly high ET rates coincide with low Tair. In the impervious areas, mainly concentrated in the city-center, more energy is absorbed from solar radiation during the day, which is released at night, increasing Tair (Christen and Vogt, 2004). A lower sky view factor from buildings blocks the release of longwave radiation and thus slows nighttime cooling (Spronken-Smith and Oke, 1999). In the same dense, highly paved areas, ET is low due to a lack of vegetation for transpiration. Nevertheless, it must be reiterated that ET and Tair act on different spatial and temporal scales. While the majority of ET occurs in daytime, the most pronounced Tair UHI intensity occurs at nighttime (Spronken-Smith and Oke, 1999), which motivated the modeling of average Tair between 22 UTC and 2 UTC in Chapter 2. Furthermore, surface cover has an influence on Tair beyond its boundaries, which decays with distance (Guo and Moore, 1998; Ho et al., 2014; Zakšek and Oštir, 2012). Thus, in Chapter 2, optimal focal buffer radii for each of the remote sensing and GIS predictors were empirically determined and then used to generate Tair maps. ET, on the other hand, is an attribute directly linked to the land surface and is not modeled with spatial aggregation. The blurring effect seen in Figure 5.2a can be partially attributed to the focal buffer approach, as well as the difference in spatial resolution (30-m for Tair; 10-m for ET).
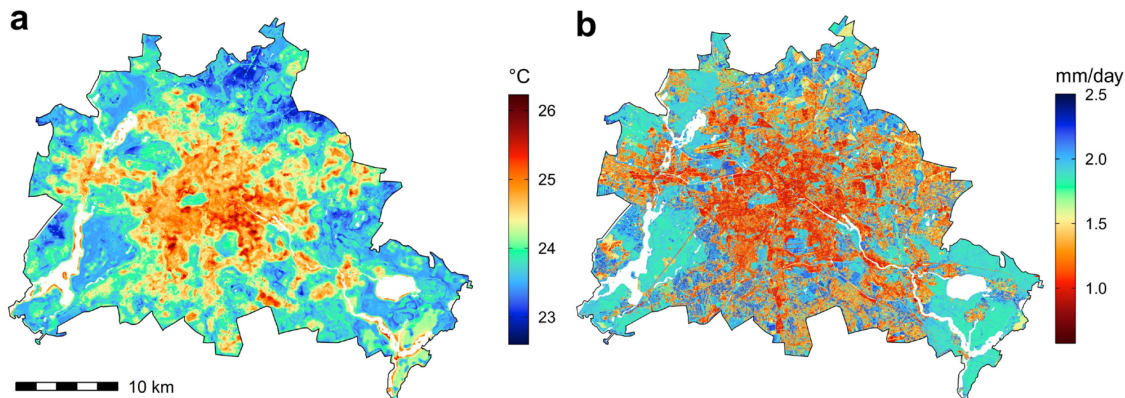
Figure 5.2: Maps of modeled thermal comfort indicators on 25 June 2019: (a) nocturnal Tair (°C) estimated with Random Forests and the "GIS" predictor scenario (Chapter 2) and (b) daily ET (mm/ day) (Chapter 4) in Berlin, Germany. White colored areas in the maps represent water bodies, which were excluded from ET modeling and masked from the Tair map for ease of visual comparison.

A sparse arrangement of trees in grassy parks may be optimal for UHI mitigation while reducing the irrigation requirement for grass by tree shade based on the findings of this thesis and

previous studies (e.g., Gómez-Navarro et al. (2021), Litvak et al. (2014)). Grasslands can also be irrigated to maximize the cooling effect in summertime, with just enough irrigation necessary to enhance ET in daytime without increasing soil moisture excessively, which may reduce cooling at night (Spronken-Smith and Oke, 1999) and increase the city's water footprint (Nouri et al., 2019). Thermal comfort in cities is largely driven by ET and shade in daytime, with a higher dependence on sky view factor and soil moisture in nighttime (Spronken-Smith and Oke, 1999). Open green spaces generally cool nocturnal Tair more than areas with trees (Gómez-Navarro et al., 2021; Spronken-Smith and Oke, 1999; Vaz Monteiro et al., 2016). Trees block radiative loss from the surface at nighttime, with cooling further slowed by wetter conditions (Spronken-Smith and Oke, 1999). In daytime, grasslands may contribute more to evaporative cooling than trees (Chapter 4) (Litvak et al., 2014; Markwitz et al., 2020; Teuling et al., 2010; Williams et al., 2012), while the cooling effect of tree shade remains crucial for thermal comfort (Rocha et al., 2022a; Wong et al., 2021). High-resolution maps of Tair and ET can be used by urban planners to further optimize the balance between nighttime (Chapter 2) and daytime cooling (Chapter 4) by varying vegetation type and irrigation strategies.

5.2 REMAINING CHALLENGES AND FUTURE RESEARCH

An urban data science focused on sustainability has emerged as a solution to addressing climate change across the world, driven by developments in remote sensing, advanced statistical methods, and crowdsourced information (Creutzig et al., 2019). AI combined with big data offers an opportunity to provide spatially precise climate mitigation solutions in urban areas worldwide and support low-carbon urban planning and policy (Milojevic-Dupont and Creutzig, 2021). Most broadly, this thesis constitutes an advance in the use of state-of-the-art technologies (AI, remote sensing, crowdsourcing) to further the understanding necessary for urban sustainability and resilience, with a focus on heat risk mitigation. However, diverse research directions for future studies remain which can address the challenges and knowledge gaps beyond the scope of this work.

This thesis is situated at the forefront of urban climate informatics (Middel et al., 2022), an emerging research field focused on better understanding complex urban climate systems through novel methodological and technological approaches. Four technological trends are driving the advancement of urban climate informatics: novel sensors (e.g., Internet-of-Things), state-of-the-art analytical algorithms and platforms (e.g., AI), new urban datasets (e.g., 3D building data, crowdsourcing), and increasing accessibility of digital infrastructure (e.g., computational power, cloud computing) (Middel et al., 2022). However, in order to meet this field's goal of facilitating climate-sensitive urban design and planning (Middel et al., 2022), substantial work remains in understanding the key indicators of human heat risk. Crowdsourcing of atmospheric parameters has mainly focused on Tair so far (Middel et al., 2022; Venter et al., 2020; Vulova et al., 2020),

but recent studies have also assessed its capacity for characterizing precipitation (Bardossy et al., 2021; de Vos et al., 2017, 2020), air pressure (de Vos et al., 2020), humidity (de Vos et al., 2020) and wind speed (Droste et al., 2020). This crowdsourced atmospheric data can also be combined with remote sensing and GIS data using AI to produce high-resolution spatial products of urban microclimate. Apparent temperature is a measure of the perception of ambient heat by humans combining humidity and Tair (Basu, 2009; Masterton and Richardson, 1979) which has been shown to be even more closely linked to mortality than other weather variables (Zhang et al., 2014). Thus, future studies can focus on mapping apparent temperature using Netatmo data to better characterize the risk of heat exposure. Spatially explicit data of atmospheric parameters derived from crowdsourcing can be integrated into both data-driven and physical models focusing on urban ET, as microclimate effects have not been considered in most approaches used to estimate urban ET (Saher et al., 2021). Furthermore, there is a high potential of crowdsourced weather data combined with satellite imagery (Chapter 2) to be implemented into real-world apps or platforms providing localized warnings of heat exposure in advance when combined with weather forecasts from other models (Nipen et al., 2020). Finally, a more comprehensive understanding of the social dimensions of urban heat risk is needed; heat stress indicators such as Tair and ET should be investigated in relation to socioeconomic status and ethnicity to gain a better understanding of heat-related vulnerability and inequity (Harlan et al., 2006).

This thesis advanced the understanding of how to model (Chapter 3) and map urban ET (Chapter 4) with open-source remote sensing and meteorological data using a data-driven approach. However, urban ET estimation, especially with remote sensing data, is still a major research frontier, with numerous remaining research directions (Nouri et al., 2015; Saher et al., 2021). First of all, the transferability of the modeling approaches proposed in Chapters 3 and 4 needs to be tested in other cities based on their availability of open-source data. The generalizability of ML models can be tested when re-training with local eddy covariance data or not, facilitated by the expanding network of urban flux towers worldwide (Jongen et al., 2022). To better inform urban planners and policy-makers preparing for rapid global change, both ML and process-based urban ET models can be used to model the impact of climate change, urbanization, and water scarcity scenarios on ET (Rocha et al., 2022b, 2022a).

5.2.1 DEVELOPMENTS IN AI

Machine learning, combined with spatially-explicit remote sensing and crowdsourced data, will be increasingly instrumental in guiding urban planning aiming to mitigate the effects of climate change and reduce greenhouse emissions (Creutzig et al., 2019; Milojevic-Dupont and Creutzig, 2021; Rolnick et al., 2023). Expanding the notion of "Smart Cities", which focuses on intelligent urban development through the use of information and communication technologies (Neirotti et al., 2014), to include the concepts of sustainability, resilience, and livability is crucial

(Rolnick et al., 2023). As urban areas will be most severely affected by increasingly frequent heat waves (Li and Bou-Zeid, 2013), characterizing urban microclimate with spatially-explicit data is one of the most relevant applications of ML in cities (Chapter 2) (Creutzig et al., 2019). Better representing ET in urban areas fulfills both the goals of heat risk mitigation and more sustainable water management in cities (Chapters 3 and 4). Optimizing the irrigation needed to maintain the ecosystem services of urban vegetation, including evaporative cooling, can be better accomplished using high-resolution urban ET products (Chapter 4), supporting the concept of the "smart water city" aiming to better manage urban water resources using state-of-the-art technologies (Gessner et al., 2014; Oberascher et al., 2022). The methodology developed in Chapters 3 and 4, which aims to overcome the lack of in situ urban ET data, can also be applied to model other fluxes measured by eddy covariance. As quantifying urban CO_2 emissions will be increasingly important to climate policy (Creutzig et al., 2019), the prototype of modeling urban ET presented in this thesis should also be utilized to model CO_2.

Despite the promise of using AI to generate high-resolution products relevant for sustainable urban planning (Milojevic-Dupont and Creutzig, 2021), as demonstrated in this thesis, this emerging field is still faced with barriers which can be resolved by new directions of research. Some of the main challenges of Earth system science are the need to extract knowledge from a deluge of big data while respecting a physical understanding of nature's laws (Reichstein et al., 2019). While ML methods can flexibly adapt to diverse data, their interpretability is limited and they are not constrained by physical laws. Integrating machine learning and physical modeling ('hybrid modeling') has a high potential for overcoming the limitations of each approach (Kraft et al., 2022; Reichstein et al., 2019) and is a promising future route of using ML to guide climate change mitigation in cities (Milojevic-Dupont and Creutzig, 2021). For instance, ground-truth data scarcity can be overcome by pre-training a neural network with simulated data from a physical model and fine-tuning it with the limited available data, such as urban EC measurements (Roscher et al., 2020a). A simple example of integrating the outputs of physical models is using reference ET (ETo) based on the Penman-Monteith equation as a predictor in ML models for actual ET instead of raw meteorological data (Chapters 3 and 4). Machine learning can also be used to better estimate parameters in physical models based on statistical covariates, such as remote sensing data (Reichstein et al., 2019); this approach has recently been successfully applied to model the global hydrological cycle (Kraft et al., 2022). In parallel to this work, the capacity of a process-based soil-vegetation-atmosphere transfer (SVAT) model to accurately model (Rocha et al., 2022b) and map (Rocha et al., 2022a) urban ET using freely available remote sensing and meteorological data was demonstrated. A process-based tracer-aided ecohydrological model (EcH2O-iso) has also been used to quantify ET in an urban environment (Gillefalk et al., 2022, 2021). Estimating input parameters for these models such as stomatal conductance, interception capacity, and soil moisture in a data-driven way based on high-resolution remote

sensing data can improve their capacity to characterize ET in a complex urban terrain (Gillefalk et al., 2022; Rocha et al., 2022a).

Explainability and interpretability are the next frontier of Earth system modeling using AI, while the overwhelming focus so far has been on optimizing accuracy (Reichstein et al., 2019; Roscher et al., 2020a, 2020b). Explainable AI (XAI) has recently emerged as a research field focused on advancing the transparency, interpretability and explainability of ML and DL algorithms to foster scientific understanding and new discoveries (Roscher et al., 2020a, 2020b). The related fields of informed ML, theory-guided data science, and physics-informed learning aim to integrate domain knowledge, including the physical laws of nature, into the ML pipeline (Roscher et al., 2020b). Decision trees such as Random Forests are, in this respect, advantageous due to the ease of extracting parameter importance, which makes them interpretable (Kuhn and Johnson, 2013; Roscher et al., 2020a). Random Forests also showed the highest accuracy throughout this work (Chapters 2 and 3) and should remain a staple in remote sensing-based regression modeling when the computational cost of more complex algorithms (e.g., deep neural networks) is not justified. One practical application of XAI in remote sensing are saliency maps, which have been used to visualize the pixels and wavelengths of hyperspectral imagery most relevant for classifying plants as diseased (Nagasubramanian et al., 2019). Saliency maps could allow us to gain insights into the interaction between a complex urban landscape and microclimate; for instance, linking crowdsourced images (e.g., Flickr) and Tair (e.g., Netatmo citizen weather stations), saliency maps could show which part of the image (soil, impervious, vegetation) contributes most to the prediction of microclimate. SHAP values are another XAI approach which could be exploited to link urban cover-related predictors to their modeled warming or cooling effect (Marcilio and Eler, 2020).

Deep learning approaches can automatically integrate both temporal and spatial context into prediction (Reichstein et al., 2019). Processes such as ET are affected by the state of the system, which is driven by past events such as drought ('memory effects') (Reichstein et al., 2019). The current distribution of Tair in a city is also highly correlated to its distribution in a previous time step (Chapter 2). In Chapter 3, 1D convolutional neural networks (CNN), a type of DL algorithm which can extract local patterns in a time series, were tested for urban ET modeling. This capacity to exploit temporal patterns may have been hindered by the presence of irregular gaps in the time series due to quality control, exhibited by 1D DNNs' slightly lower accuracy than Random Forests (Chapter 3). Long-short-term-memories (LSTMs), an even more powerful DL model for modeling dynamic time series than 1D CNNs, are promising for future studies modeling urban ET and Tair due to their consideration of memory and lag effects (Chollet and Allaire, 2018; Zhang et al., 2022). Spatially explicit models, such as Integrated Nested Laplace Approximation (INLA), are a promising alternative to deep learning for integrating spatio-temporal dependencies

into modeling (Rocha et al., 2019; Simpson et al., 2012). The use of such spatial models to predict plant traits using remote sensing data is still in its very inception (Rocha et al., 2019), but holds high potential for better modeling the ET of urban vegetation.

To effectively implement machine learning for planning sustainable and resilient cities, a better understanding of the source of the modeling uncertainty is needed. ML-based studies commonly select a data source, often due to its availability for the study area, which is then used for both training and validation. To avoid influencing the accuracy with the spatio-temporal autocorrelation of the data itself, this data are ideally split in a way meant to minimize temporal autocorrelation (e.g., train in one year, test in another) or spatial autocorrelation (train in one location, test in another location) (Roberts et al., 2017). Then, performance metrics are calculated (commonly, RMSE and R^2) and the accuracy is declared to be satisfactory (or not). However, what remains unknown from most studies is how much the uncertainty and bias of the (training and testing) data affected the output and calculated metrics. Given their availability, different sources of reference data can better resolve this open question. In Chapter 2, the influence of validating Tair maps with either crowdsourced weather data (which was also used in training) or traditional weather stations was investigated. Crowdsourced weather data has only recently been used in climatological studies, with its validity for climatological research still widely questioned (Fenner et al., 2017; Meier et al., 2017; Muller et al., 2015; Napoly et al., 2018). Testing the accuracy with traditional reference stations revealed a lower accuracy and a greater overestimation of Tair than testing with crowdsourced data. Most "Netatmo" stations are installed in highly built-up areas and near building walls due to their owners' convenience (Fenner et al., 2017; Meier et al., 2017); thus, this data present more bias and inaccuracy than reference meteorological data even after thorough quality-control.

In situ data on ET in cities is even scarcer than data on Tair. In Chapters 3 and 4, AI models predicting ET are trained and tested using eddy covariance (EC) data. However, EC data also present large uncertainty and require extensive quality-control to be usable (Feigenwinter et al., 2018). Thus, some of the inaccuracy when modeling ET, whether with physical or ML models, is inherently due to the noise of the EC data, and not due to the algorithm (Rocha et al., 2022b). In future work, ET models should be independently validated with alternative sources of ground-truth data, including lysimeters and sap flow sensors (Nouri et al., 2013). Measurements of gas exchange (transpiration, stomatal conductance) in leaves, combined with leaf area index measurements, can also provide a valuable source of validation data for urban ET (Vulova et al., 2019). The nature of the modeling uncertainty should ultimately also be well-communicated to urban planners. While overestimating heat risk due to data bias (see Chapter 2) may be acceptable, underestimating it may lead to inadequate short-term (e.g. warning systems) and long-term (e.g. greening) solutions to protect human well-being.

5.2.2 OUTLOOK ON REMOTE SENSING OF THERMAL COMFORT

Various Earth Observation (EO) sensors hold great potential for modeling urban Tair and ET. In this thesis, LST was used as a predictor of Tair (Chapter 2), but not of ET (Chapters 3-4). As discussed in Chapter 4, though thermal imagery is highly relevant to ET, it was not integrated into urban ET estimation due to the limited temporal and spatial resolution of current TIR sensors. The most commonly used TIR sensor in UHI studies, Landsat, has a relatively high spatial resolution (60-120 m) but its long revisiting cycle (16-day) is further reduced by cloud cover (Zhou et al., 2019). For instance, by selecting cloud-free Landsat 8 imagery over Berlin in summertime from 2015 to 2019 (Chapter 2), only four scenes remained, underscoring the low temporal resolution. The use of Sentinel-2 (Chapter 4) or harmonized Landsat and Sentinel-2 time series (Chapter 3) (Claverie et al., 2018) allowed for a higher temporal resolution. While pixel-wise linear interpolation of a vegetation index (Chapters 3-4) can be justified due to the day-to-day stability of the vegetation state, such interpolation is not applicable for LST, which is highly temporally dynamic (Zhou et al., 2019). Moreover, the spatial resolution of other open-access TIR sensors, such as MODIS and Sentinel-3 (both with 1 km resolution) is insufficient for urban planning applications (Venter et al., 2020; Zhou et al., 2019). The recent or upcoming launch of higher spatial- and temporal-resolution TIR sensors, including the Thermal infraRed Imaging Satellite for High-resolution Natural resource Assessment (TRISHNA) mission (Vidal et al., 2022) and the ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) mission (Anderson et al., 2021), can allow for TIR imagery to be integrated into the mapping of thermal comfort indicators in cities.

Other remotely sensed data which could enhance urban Tair and ET modeling include NASA Global Ecosystem Dynamics Investigation (GEDI), a LiDAR instrument measuring 3D vegetation structure and aboveground biomass (Dubayah et al., 2020), and Sentinel-1, a SAR sensor (Chintala et al., 2022). Improvements in radiometric calibration of nano-satellite imagery, such as the CubeSat constellation, could allow for even higher spatio-temporal resolution (3-5 m, daily) mapping of thermal comfort indicators (Aragon et al., 2021; Leach et al., 2019). Furthermore, nighttime lights imagery, such as NASA's daily Black Marble product (Román et al., 2018), can be used as a proxy of anthropogenic heat, a major driver of ET variation in urban environments which is unaccounted for in most ET models (Cong et al., 2017; Saher et al., 2021). The relationship between chlorophyll fluorescence, an indicator of plant photosynthetic activity, and transpiration, has not been investigated at large scale due to the limited existing instrumentation (Damm et al., 2018). Fluorescence spectroscopy may become a viable source of data through upcoming missions such as the Fluorescence Explorer (FLEX) (Drusch et al., 2017).

An understanding of entire components and processes contributing to urban ET is still missing, which can be improved by analyzing remote sensing data in conjunction with various in

situ data. Soil sealing is the central driver of the challenges associated with urbanization, including the UHI effect and flooding, but an understanding of the hydrological balance of paved surfaces, in particular of evaporation, is still lacking (Gessner et al., 2014; Timm et al., 2018). This thesis contributed to filling this gap by estimating ET for an entire city, including impervious surfaces, using a data-driven approach in Chapter 4. Lysimeter measurements can provide ground-truth data for modeling evaporation from different pavement types (Aljoumani et al., 2022). In order to characterize the thermal dynamics and evaporation of paved surfaces, a high spatio-temporal time series of thermal imagery has been collected using a fixed platform over two lysimeters in an ongoing study (Pipkins et al., 2017). Coupling continuous in situ and remote sensing measurements holds great potential for better modeling the evaporation of sealed surfaces. The evaporation and cooling potential of urban water bodies is still largely uncertain (Jacobs et al., 2020; Wang and Ouyang, 2021). To plan for cooler cities with blue infrastructure, further studies combining field-based, remote sensing, and modeling methods are needed to assess the contribution of water bodies to urban cooling and ET. Finally, interception loss from precipitation in an urban environment, which is highly variable due to a complex topography, should be investigated in future work. Eddy covariance measurements cannot separate ET into the different sources of water vapor: plant transpiration, soil evaporation, interception loss, and anthropogenic emissions (Ramamurthy and Bou-Zeid, 2014). Terrestrial and airborne laser scanning is one promising tool that can better characterize urban water storage capacity and potential interception loss (Baptista et al., 2018; Sampson et al., 2012).

Higher-resolution remote sensing imagery collected from drones and other platforms can provide a crucial link between in situ ET measurements and satellite imagery (Feng et al., 2022; Qin et al., 2022; Saher et al., 2021). In a preliminary study, UAV flights with thermal and multispectral cameras were conducted over an urban garden in Berlin on a monthly basis throughout the growing season of 2019 (Vulova et al., 2019) (Figure 5.3). Diverse field data was collected simultaneously as the UAV flights, including transpiration, stomatal conductance, leaf area index, soil moisture, flux tower measurements, and sap flow (Kuhlemann et al., 2021; Vulova et al., 2019). This unique dataset can facilitate an understanding of how well different vegetation indices and models can approximate the spatio-temporal variation of urban ET. The spatial resolution of UAV data (~5 cm) is well-suited for investigating the effects of shade on ET and microclimate, which are currently largely unknown (Saher et al., 2021) (Figure 5.3). Two field days with hourly flights also captured the diurnal ET variation of urban vegetation at a fine scale for the first time.
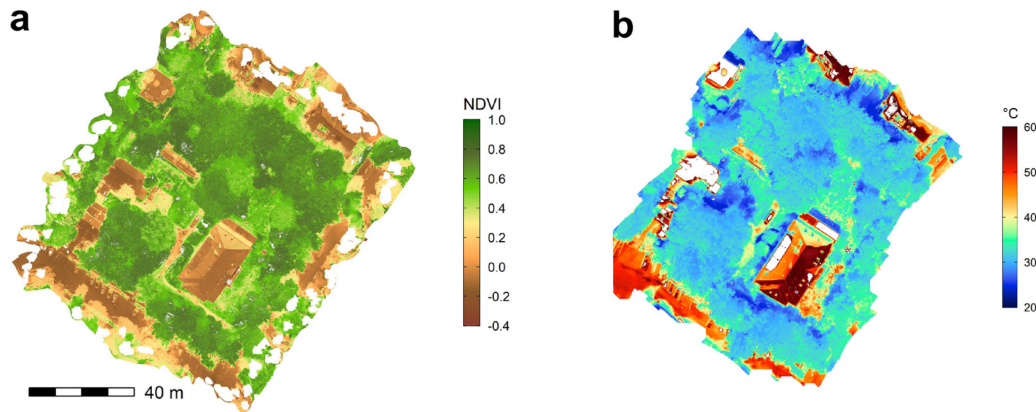
Figure 5.3: UAV imagery from an overflight on 14 June 2019 from an ongoing study aiming to characterize urban ET and microclimate in a research garden in Berlin, Germany (Vulova et al., 2019): (a) normalized difference vegetation index (NDVI) and (b) land surface temperature (LST).

Although UAVs could prove to be highly beneficial for urban ET estimation (Saher et al., 2021), restrictions on UAV flights in many cities pose a challenge for this avenue of research. Proximal sensing, referring to mounting camera systems on fixed platforms or vehicles, is a promising alternative to monitor urban phenology and other plant traits (Rumora et al., 2021). Mounting thermal and multispectral cameras which continuously and regularly take photos on top of an urban flux tower, an example of stationary sensing (Anjomshoaa et al., 2018), could improve our understanding of the link between remote sensing time series and ET measurements. "Drive-by sensing" is an emerging approach within the "smart city" paradigm (Anjomshoaa et al., 2018) that can facilitate a continuous and cost-effective monitoring of the phenology, vitality, and ecosystem services of street trees (Filippa et al., 2018). Mounting multispectral and thermal cameras on existing vehicle fleets such as trams or garbage trucks could allow for the state of urban vegetation to regularly be captured at an unprecedented high spatial and temporal resolution (Anjomshoaa et al., 2018; Filippa et al., 2018). To model the ecosystem services of urban vegetation, including evaporative cooling, drive-by sensing should be coupled with an expansion of continuous ecophysiological measurements in urban trees: e.g. sap flow and other sensors which can be integrated into an Internet-of-Things (IoT) network for monitoring urban green infrastructure in real-time (Matasov et al., 2020).

5.2.3 FUTURE APPLICATIONS IN URBAN PLANNING

The methodologies for modeling thermal comfort indicators developed in this thesis can be integrated into urban planning processes aimed at UHI mitigation and the enhancement of human health and well-being. AI in combination with remote sensing imagery, crowdsourced data, and other geodata makes it possible to generate spatially explicit products of parameters relevant to

climate mitigation in cities (Milojevic-Dupont and Creutzig, 2021) - Tair and ET, in the case of this thesis. High-resolution, localized information on thermal comfort can be integrated into national or European-wide heat risk governance plans (Lass et al., 2011). Identifying the areas of a city with highest heat risk is a crucial first step for decision-makers to target mitigation interventions. Such interventions can include the addition of vegetated surfaces in cities, as epitomized by the Sponge City (He et al., 2019) or similar frameworks, or the optimization of the plant species and amount of irrigation in existing green spaces. Heat exposure risk can even be predicted in advance (as demonstrated for Tair in this thesis), and therefore can be integrated into early-warning systems such as smartphone apps, an increasingly important form of 'risk communication' (Lass et al., 2011). Such platforms will allow urban residents to modify their own behavior to reduce heat stress, while also indicating an anticipated strain on energy infrastructure for the affected stakeholders. Policymakers need to develop actionable solutions to heat-related risks, which require an understanding of the impact of anthropogenic and environmental changes in the urban environment (Milojevic-Dupont and Creutzig, 2021). ML approaches integrating various geodata allow for the simulation of the effects of both climate change scenarios and urban planning and policy decisions such as land cover conversion on thermal comfort. For instance, Zurich is already using a Digital Twin of the city integrating 3D spatial data to assess the potential impact of urban planning decisions on the urban climate (Schrotter and Hürzeler, 2020). While this thesis focuses largely on thermal comfort, a greater understanding of the spatio-temporal dynamics of urban ET will allow for a smarter management of water resources in a time when drought is increasingly threatening the ecosystem services provided by urban vegetation (Miller et al., 2020), including its cooling capacity. Ultimately, emerging technologies, including novel remote and in situ sensors, volunteered geographic information, and AI, can revolutionize our understanding of urban ecosystems and allow for the world's growing cities to better plan for human well-being, thermal comfort, and sustainable water resource management.

Bibliography

Acuto, M., Parnell, S., Seto, K.C., 2018. Building a global urban science. Nat. Sustain. https://doi.org/10.1038/s41893-017-0013-9

Aljoumani, B., Sanchez-Espigares, J.A., Kluge, B., Wessolek, G., Kleinschmit, B., 2022. Analyzing Temporal Trends of Urban Evaporation Using Generalized Additive Models. Land 11, 508. https://doi.org/10.3390/land11040508

Allaire, J.J., 2018. tfruns: Training Run Tools for "TensorFlow" (R package version 1.4) [WWW Document]. URL https://cran.r-project.org/package=tfruns

Allaire, J.J., Chollet, F., 2020. keras: R Interface to "Keras" (R package version 2.3.0.0.9000) [WWW Document]. URL https://keras.rstudio.com

Allen, L., Lindberg, F., Grimmond, C.S.B., 2011. Global to city scale urban anthropogenic heat flux: Model and variability. Int. J. Climatol. 31, 1990–2005. https://doi.org/10.1002/joc.2210

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration: Guidelines for computing crop requirements. FAO Irrigation and drainage paper 56, FAO - Food and Agriculture Organization of the United Nations. Rome, Italy. https://doi.org/10.1016/j.eja.2010.12.001

Allen, R.G., Smith, M., Perrier, A., Pereira, L.S., 1994. An Update for the Definition of Reference Evapotranspiration. ICID Bull. 43, 1–34.

Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E., Snyder, R.L., 2005. The ASCE Standardized Reference Evapotranspiration Equation. American Society of Civil Engineers. https://doi.org/10.1061/9780784408056

Anderson, M.C., Yang, Yang, Xue, J., Knipper, K.R., Yang, Yun, Gao, F., Hain, C.R., Kustas, W.P., Cawse-Nicholson, K., Hulley, G., Fisher, J.B., Alfieri, J.G., Meyers, T.P., Prueger, J., Baldocchi, D.D., Rey-Sanchez, C., 2021. Interoperability of ECOSTRESS and Landsat for mapping evapotranspiration time series at sub-field scales. Remote Sens. Environ. 252, 112189. https://doi.org/10.1016/j.rse.2020.112189

Anjomshoaa, A., Duarte, F., Rennings, D., Matarazzo, T.J., Desouza, P., Ratti, C., 2018. City Scanner: Building and Scheduling a Mobile Sensing Platform for Smart City Services. IEEE Internet Things J. 5, 4567–4579. https://doi.org/10.1109/JIOT.2018.2839058

Aragon, B., Ziliani, M.G., Houborg, R., Franz, T.E., McCabe, M.F., 2021. CubeSats deliver new insights into agricultural water use at daily and 3 m resolutions. Sci. Rep. 11. https://doi.org/10.1038/s41598-021-91646-w

Azevedo, J., Chapman, L., Muller, C., 2016. Quantifying the Daytime and Night-Time Urban Heat Island in Birmingham, UK: A Comparison of Satellite Derived Land Surface Temperature and High Resolution Air Temperature Observations. Remote Sens. 8, 153. https://doi.org/10.3390/rs8020153

Baptista, M., Livesley, S., G. Parmehr, E., Neave, M., Amati, M., 2018. Terrestrial Laser Scanning to Predict Canopy Area Metrics, Water Storage Capacity, and Throughfall Redistribution in Small Trees. Remote Sens. 10, 1958. https://doi.org/10.3390/rs10121958

Bardossy, A., Seidel, J., El Hachem, A., 2021. The use of personal weather station observations to improve precipitation estimation and interpolation. Hydrol. Earth Syst. Sci. 25, 583–601. https://doi.org/10.5194/hess-25-583-2021

Bastiaanssen, W.G.M., Pelgrum, H., Wang, J., Ma, Y., Moreno, J.F., Roerink, G.J., Van Der Wal, T., 1998. A remote sensing surface energy balance algorithm for land (SEBAL): 2. Validation. J. Hydrol. 212–213, 213–229. https://doi.org/10.1016/S0022-1694(98)00254-6

Basu, R., 2009. High ambient temperature and mortality: A review of epidemiologic studies from 2001 to 2008. Environ. Heal. A Glob. Access Sci. Source. https://doi.org/10.1186/1476-069X-8-40

Bayulken, B., Huisingh, D., Fisher, P.M.J., 2021. How are nature based solutions helping in the greening of cities in the context of crises such as climate change and pandemics? A comprehensive review. J. Clean. Prod. 288. https://doi.org/10.1016/j.jclepro.2020.125569

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5. https://doi.org/10.1038/sdata.2018.214

Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N., Santos, A., 2012. Estimating air surface temperature in Portugal using MODIS LST data. Remote Sens. Environ. 124, 108–121. https://doi.org/10.1016/j.rse.2012.04.024

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B 57, 289–300. https://doi.org/https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berland, A., Shiflett, S.A., Shuster, W.D., Garmestani, A.S., Goddard, H.C., Herrmann, D.L., Hopton, M.E., 2017. The role of trees in urban stormwater management. Landsc. Urban Plan. https://doi.org/10.1016/j.landurbplan.2017.02.017

Berlin Senate Department for Urban Development and Housing, 2014. Building and Vegetation Heights (Edition 2014) [WWW Document]. Berlin Environ. Atlas. URL https://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/ed610_05.htm (accessed

8.14.19).

Besir, A.B., Cuce, E., 2018. Green roofs and facades: A comprehensive review. Renew. Sustain. Energy Rev. 82, 915–939. https://doi.org/10.1016/j.rser.2017.09.106

Boegh, E., Poulsen, R.N., Butts, M., Abrahamsen, P., Dellwik, E., Hansen, S., Hasager, C.B., Ibrom, A., Loerup, J.K., Pilegaard, K., Soegaard, H., 2009. Remote sensing based evapotranspiration and runoff modeling of agricultural, forest and urban flux sites in Denmark: From field to macro-scale. J. Hydrol. 377, 300–316. https://doi.org/10.1016/j.jhydrol.2009.08.029

Boegh, E., Thorsen, M., Butts, M.B., Hansen, S., Christiansen, J.S., Abrahamsen, P., Hasager, C.B., Jensen, N.O., Van Der Keur, P., Refsgaard, J.C., Schelde, K., Soegaard, H., Thomsen, A., 2004. Incorporating remote sensing data in physically based distributed agro-hydrological modelling. J. Hydrol. 287, 279–299. https://doi.org/10.1016/j.jhydrol.2003.10.018

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chapman, L., Bell, C., Bell, S., 2017. Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations. Int. J. Climatol. 37, 3597–3605. https://doi.org/10.1002/joc.4940

Chapman, L., Muller, C.L., Young, D.T., Warren, E.L., Grimmond, C.S.B., Cai, X.M., Ferranti, E.J.S., 2015. The Birmingham Urban Climate Laboratory: An Open Meteorological Test Bed and Challenges of the Smart City. Bull. Am. Meteorol. Soc. 96, 1545–1560. https://doi.org/10.1175/BAMS-D-13-00193.1

Chintala, S., Harmya, T.S., Kambhammettu, B.V.N.P., Moharana, S., Duvvuri, S., 2022. Modelling high-resolution Evapotranspiration in fragmented croplands from the constellation of Sentinels. Remote Sens. Appl. Soc. Environ. 26, 100704. https://doi.org/10.1016/j.rsase.2022.100704

Chollet, F., Allaire, J.J., 2018. Deep Learning with R. Manning Publications Co., Shelter Island, NY.

Christen, A., 2016. Gridded Turbulent Source Area [WWW Document]. GitHub. URL https://github.com/achristen/Gridded-Turbulent-Source-Area (accessed 10.26.20).

Christen, A., Coops, N.C., Crawford, B.R., Kellett, R., Liss, K.N., Olchovski, I., Tooke, T.R., Van Der Laan, M., Voogt, J.A., 2011. Validation of modeled carbon-dioxide emissions from an urban neighborhood with direct eddy-covariance measurements. Atmos. Environ. 45, 6057–6069. https://doi.org/10.1016/j.atmosenv.2011.07.040

Christen, A., Vogt, R., 2004. Energy and radiation balance of a central European City. Int. J. Climatol. 24, 1395–1421. https://doi.org/10.1002/joc.1074

Chrysoulakis, N., Grimmond, S., Feigenwinter, C., Lindberg, F., Gastellu-Etchegorry, J.-P., Marconcini, M., Mitraka, Z., Stagakis, S., Crawford, B., Olofson, F., Landier, L., Morrison, W., Parlow, E., 2018. Urban energy exchanges monitoring from space. Sci. Rep. 8, 11498. https://doi.org/10.1038/s41598-018-29873-x

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002

Coccolo, S., Kämpf, J., Mauree, D., Scartezzini, J.L., 2018. Cooling potential of greening in the urban environment, a step further towards practice. Sustain. Cities Soc. 38, 543–559. https://doi.org/10.1016/j.scs.2018.01.019

Cong, Z.T., Shen, Q.N., Zhou, L., Sun, T., Liu, J.H., 2017. Evapotranspiration estimation considering anthropogenic heat based on remote sensing in urban area. Sci. China Earth Sci. 60, 659–671. https://doi.org/10.1007/s11430-016-0216-3

Crawford, B., Christen, A., 2015. Spatial source attribution of measured urban eddy covariance CO2 fluxes. Theor. Appl. Climatol. 119, 733–755. https://doi.org/10.1007/s00704-014-1124-0

Creutzig, F., Lohrey, S., Bai, X., Baklanov, A., Dawson, R., Dhakal, S., Lamb, W.F., McPhearson, T., Minx, J., Munoz, E., Walsh, B., 2019. Upscaling urban data science for global climate solutions. Glob. Sustain. 2. https://doi.org/10.1017/sus.2018.16

Cuthbert, M.O., Rau, G.C., Ekström, M., O'Carroll, D.M., Bates, A.J., 2022. Global climate-driven trade-offs between the water retention and cooling benefits of urban greening. Nat. Commun. 13. https://doi.org/10.1038/s41467-022-28160-8

Damm, A., Paul-Limoges, E., Haghighi, E., Simmer, C., Morsdorf, F., Schneider, F.D., van der Tol, C., Migliavacca, M., Rascher, U., 2018. Remote sensing of plant-water relations: An overview and future perspectives. J. Plant Physiol. 227, 3–19. https://doi.org/10.1016/j.jplph.2018.04.012

De Ridder, K., Lauwaet, D., Maiheu, B., 2015. UrbClim - A fast urban boundary layer climate model. Urban Clim. 12, 21–48. https://doi.org/10.1016/j.uclim.2015.01.001

de Vos, L., Leijnse, H., Overeem, A., Uijlenhoet, R., 2017. The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. Hydrol. Earth Syst. Sci. 21, 765–777. https://doi.org/10.5194/hess-21-765-2017

de Vos, L.W., Droste, A.M., Zander, M.J., Overeem, A., Leijnse, H., Heusinkveld, B.G.,

Steeneveld, G.J., Uijlenhoet, R., 2020. Hydrometeorological monitoring using opportunistic sensing networks in the Amsterdam metropolitan area. Bull. Am. Meteorol. Soc. 101, E167–E185. https://doi.org/10.1175/BAMS-D-19-0091.1

Demuzere, M., Kittner, J., Bechtel, B., 2021. LCZ Generator: A Web Application to Create Local Climate Zone Maps. Front. Environ. Sci. 9. https://doi.org/10.3389/fenvs.2021.637455

Dettmann, U., Grimma, R., 2019. MeTo: Meteorological Tools (R package version 0.1.0) [WWW Document]. URL https://cran.r-project.org/package=MeTo

Dousset, B., Gourmelon, F., Laaidi, K., Zeghnoun, A., Giraudet, E., Bretin, P., Mauri, E., Vandentorren, S., 2011. Satellite monitoring of summer heat waves in the Paris metropolitan area. Int. J. Climatol. 31, 313–323. https://doi.org/10.1002/joc.2222

Droste, A.M., Heusinkveld, B.G., Fenner, D., Steeneveld, G., 2020. Assessing the potential and application of crowdsourced urban wind data. Q. J. R. Meteorol. Soc. 146, 2671–2688. https://doi.org/10.1002/qj.3811

Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., Kraft, S., Middleton, E.M., Miglietta, F., Mohammed, G., Nedbal, L., Rascher, U., Schuttemeyer, D., Verhoef, W., 2017. The FLuorescence EXplorer Mission Concept-ESA's Earth Explorer 8. IEEE Trans. Geosci. Remote Sens. 55, 1273–1284. https://doi.org/10.1109/TGRS.2016.2621820

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P.L., Qi, W., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Sci. Remote Sens. 1, 100002. https://doi.org/10.1016/j.srs.2020.100002

Dugord, P.A., Lauf, S., Schuster, C., Kleinschmit, B., 2014. Land use patterns, temperature distribution, and potential heat stress risk - The case study Berlin, Germany. Comput. Environ. Urban Syst. 48, 86–98. https://doi.org/10.1016/j.compenvurbsys.2014.07.005

Dwarakish, G.S., Ganasri, B.P., De Stefano, L., 2015. Impact of land use change on hydrological systems: A review of current modeling approaches. Cogent Geosci. 1, 1115691. https://doi.org/10.1080/23312041.2015.1115691

DWD, 2022. Vieljährige Mittelwerte [WWW Document]. URL https://www.dwd.de/DE/leistungen/klimadatendeutschland/vielj_mittelwerte.html (accessed 8.9.22).

DWD, 2021. DWD Climate Data Center (CDC) [WWW Document]. URL https://opendata.dwd.de/climate_environment/CDC/observations_germany/ (accessed 6.15.21).

DWD, 2020. Vieljährige Mittelwerte [WWW Document]. URL https://www.dwd.de/DE/leistungen/klimadatendeutschland/vielj_mittelwerte.html (accessed 11.4.20).

EEA, 2022a. Economic losses and fatalities from weather- and climate-related events in Europe. https://doi.org/10.2800/530599

EEA, 2022b. Who benefits from nature in cities? Social inequalities in access to urban green and blue spaces across Europe. https://doi.org/10.2800/041176

ESA, 2015. Sentinel-2 User Handbook.

European Environment Agency, 2018. Urban Atlas [WWW Document]. URL https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-urban-atlas (accessed 5.14.20).

European Union, 2016. Mapping Guide v4.7 for a European Urban Atlas.

Faridatul, M.I., Wu, B., Zhu, X., Wang, S., 2020. Improving remote sensing based evapotranspiration modelling in a heterogeneous urban environment. J. Hydrol. 581. https://doi.org/10.1016/j.jhydrol.2019.124405

Feichtinger, M., de Wit, R., Goldenits, G., Kolejka, T., Hollósi, B., Žuvela-Aloise, M., Feigl, J., 2020. Case-study of neighborhood-scale summertime urban air temperature for the City of Vienna using crowd-sourced data. Urban Clim. 32, 1–12. https://doi.org/10.1016/j.uclim.2020.100597

Feigenwinter, C., Vogt, R., Parlow, E., Lindberg, F., Marconcini, M., Frate, F. Del, Chrysoulakis, N., 2018. Spatial Distribution of Sensible and Latent Heat Flux in the City of Basel (Switzerland). IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 2717–2723. https://doi.org/10.1109/JSTARS.2018.2807815

Feng, L., Liu, Y., Zhou, Y., Yang, S., 2022. A UAV-derived thermal infrared remote sensing three-temperature model and estimation of various vegetation evapotranspiration in urban micro-environments. Urban For. Urban Green. 69, 127495. https://doi.org/10.1016/j.ufug.2022.127495

Fenner, D., Holtmann, A., Krug, A., Scherer, D., 2019. Heat waves in Berlin and Potsdam, Germany – Long-term trends and comparison of heat wave definitions from 1893 to 2017. Int. J. Climatol. 39, 2422–2437. https://doi.org/10.1002/joc.5962

Fenner, D., Meier, F., Bechtel, B., Otto, M., Scherer, D., 2017. Intra and inter "local climate zone" variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. Meteorol. Zeitschrift 26, 525–547. https://doi.org/10.1127/metz/2017/0861

Fenner, D., Meier, F., Scherer, D., Polze, A., 2014. Spatial and temporal air temperature variability in Berlin, Germany, during the years 2001-2010. Urban Clim. 10, 308–331. https://doi.org/10.1016/j.uclim.2014.02.004

Ferreira, L.B., da Cunha, F.F., 2020. New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning. Agric. Water Manag. 234, 106113. https://doi.org/10.1016/j.agwat.2020.106113

Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Sonnentag, O., Humphreys, E., Hufkens, K., Ryu, Y., Verfaillie, J., Morra di Cella, U., Richardson, A.D., 2018. NDVI derived from near-infrared-enabled digital cameras: Applicability across different plant functional types. Agric. For. Meteorol. 249, 275–285. https://doi.org/10.1016/j.agrformet.2017.11.003

Fletcher, T.D., Deletic, A., Mitchell, V.G., Hatt, B.E., 2008. Reuse of Urban Runoff in Australia: A Review of Recent Advances and Remaining Challenges. J. Environ. Qual. 37, S-116-S-127. https://doi.org/10.2134/jeq2007.0411

Foken, T., 2016. Angewandte Meteorologie: Mikrometeorologische Methoden. Springer Spektrum, New York, United States. https://doi.org/10.1007/978-3-642-25525-0

Foltýnová, L., Fischer, M., McGloin, R.P., 2020. Recommendations for gap-filling eddy covariance latent heat flux measurements using marginal distribution sampling. Theor. Appl. Climatol. 139, 677–688. https://doi.org/10.1007/s00704-019-02975-w

Gabriel, K.M.A., Endlicher, W.R., 2011. Urban and rural mortality rates during heat waves in Berlin and Brandenburg, Germany. Environ. Pollut. 159, 2044–2050. https://doi.org/10.1016/j.envpol.2011.01.016

Gaffin, S.R., Rosenzweig, C., Kong, A.Y.Y., 2012. Adapting to climate change through urban green infrastructure. Nat. Clim. Chang. https://doi.org/10.1038/nclimate1685

Gaines, J.M., 2016. Flooding: Water potential. Nature 531, S54–S55. https://doi.org/10.1038/531S54a

Gessner, M.O., Hinkelmann, R., Nützmann, G., Jekel, M., Singer, G., Lewandowski, J., Nehls, T., Barjenbruch, M., 2014. Urban water interfaces. J. Hydrol. 514, 226–232. https://doi.org/10.1016/j.jhydrol.2014.04.021

Gillefalk, M., Tetzlaff, D., Hinkelmann, R., Kuhlemann, L.M., Smith, A., Meier, F., Maneta, M.P., Soulsby, C., 2021. Quantifying the effects of urban green space on water partitioning and ages using an isotope-based ecohydrological model. Hydrol. Earth Syst. Sci. 25, 3635–3652. https://doi.org/10.5194/hess-25-3635-2021

Gillefalk, M., Tetzlaff, D., Marx, C., Smith, A., Meier, F., Hinkelmann, R., Soulsby, C., 2022.

Estimates of water partitioning in complex urban landscapes with isotope-aided ecohydrological modelling. Hydrol. Process. 36, e14532. https://doi.org/10.1002/hyp.14532

Gómez-Navarro, C., Pataki, D.E., Pardyjak, E.R., Bowling, D.R., 2021. Effects of vegetation on the spatial and temporal variation of microclimate in the urbanized Salt Lake Valley. Agric. For. Meteorol. 296. https://doi.org/10.1016/j.agrformet.2020.108211

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202. https://doi.org/10.1016/j.rse.2017.06.031

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. R J. 8. https://doi.org/10.32614/rj-2016-014

Grassmann, T., Napoly, A., Meier, F., Fenner, D., 2018. CrowdQC v1.2.0 - Quality Control for Crowdsourced Data From CWS [WWW Document]. https://doi.org/10.14279/depositonce-6740.3

Greenwell, B., Boehmke, B., Gray, B., 2020. vip: Variable Importance Plots (R package version 0.2.2) [WWW Document]. URL https://cran.r-project.org/package=vip

Grimmond, C.S.B., Blackett, M., Best, M.J., Barlow, J., Baik, J.J., Belcher, S.E., Bohnenstengel, S.I., Calmet, I., Chen, F., Dandou, A., Fortuniak, K., Gouvea, M.L., Hamdi, R., Hendry, M., Kawai, T., Kawamoto, Y., Kondo, H., Krayenhoff, E.S., Lee, S.H., Loridan, T., Martilli, A., Masson, V., Miao, S., Oleson, K., Pigeon, G., Porson, A., Ryu, Y.H., Salamanca, F., Shashua-Bar, L., Steeneveld, G.J., Tombrou, M., Voogt, J., Young, D., Zhang, N., 2010. The international urban energy balance models comparison project: First results from phase 1. J. Appl. Meteorol. Climatol. 49, 1268–1292. https://doi.org/10.1175/2010JAMC2354.1

Grimmond, C.S.B., Oke, T.R., 1999. Aerodynamic Properties of Urban Areas Derived from Analysis of Surface Form. J. Appl. Meteorol. 38, 1262–1292. https://doi.org/https://doi.org/10.1175/1520-0450(1999)038<1262:APOUAD>2.0.CO;2

Gunawardena, K.R., Wells, M.J., Kershaw, T., 2017. Utilising green and bluespace to mitigate urban heat island intensity. Sci. Total Environ. 584–585, 1040–1055. https://doi.org/10.1016/j.scitotenv.2017.01.158

Guo, L.J., Moore, J.M.M., 1998. Pixel block intensity modulation: Adding spatial detail to tm band 6 thermal imagery. Int. J. Remote Sens. 19, 2477–2491. https://doi.org/10.1080/014311698214578

Guzinski, R., Nieto, H., 2019. Evaluating the feasibility of using Sentinel-2 and Sentinel-3 satellites for high-resolution evapotranspiration estimations. Remote Sens. Environ. 221, 157–172. https://doi.org/10.1016/j.rse.2018.11.019

Haidar, A., Verma, B., 2018. Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network. IEEE Access 6, 69053–69063. https://doi.org/10.1109/ACCESS.2018.2880044

Harlan, S.L., Brazel, A.J., Prashad, L., Stefanov, W.L., Larsen, L., 2006. Neighborhood microclimates and vulnerability to heat stress. Soc. Sci. Med. 63, 2847–2863. https://doi.org/10.1016/j.socscimed.2006.07.030

Hastie, T., Friedman, J., Tibshirani, R., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-21606-5

He, B.J., Zhu, J., Zhao, D.X., Gou, Z.H., Qi, J. Da, Wang, J., 2019. Co-benefits approach: Opportunities for implementing sponge city and urban heat island mitigation. Land use policy 86, 147–157. https://doi.org/10.1016/j.landusepol.2019.05.003

Hiemstra, J.A., Saaroni, H., Amorim, J.H., 2017. The Urban Heat Island: Thermal Comfort and the Role of Urban Greening. Springer, Cham, pp. 7–19. https://doi.org/10.1007/978-3-319-50280-9_2

Hiemstra, P.H., Pebesma, E.J., Twenhöfel, C.J.W., Heuvelink, G.B.M., 2009. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. Comput. Geosci. 35, 1711–1721. https://doi.org/https://doi.org/10.1016/j.cageo.2008.10.011

Ho, H.C., Knudby, A., Sirovyak, P., Xu, Y., Hodul, M., Henderson, S.B., 2014. Mapping maximum urban air temperature on hot summer days. Remote Sens. Environ. 154, 38–45. https://doi.org/10.1016/j.rse.2014.08.012

Ho, H.C., Knudby, A., Xu, Y., Hodul, M., Aminipouri, M., 2016. A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area. Sci. Total Environ. 544, 929–938. https://doi.org/10.1016/j.scitotenv.2015.12.021

Holl, D., Pfeiffer, E.M., Kutzbach, L., 2020. Comparison of eddy covariance CO2 and CH4 fluxes from mined and recently rewetted sections in a northwestern German cutover bog. Biogeosciences 17, 2853–2874. https://doi.org/10.5194/bg-17-2853-2020

Hölzl, S.E., Veskov, M., Scheibner, T., Le, T.T., Kleinschmit, B., 2021. Vulnerable socioeconomic groups are disproportionately exposed to multiple environmental burden in Berlin - implications for planning. Int. J. Urban Sustain. Dev. 13, 334–350. https://doi.org/10.1080/19463138.2021.1904246

Huang, K., Li, X., Liu, X., Seto, K.C., 2019. Projecting global urban land expansion and heat

island intensification through 2050. Environ. Res. Lett. 14, 114037. https://doi.org/10.1088/1748-9326/ab4b71

Izrailev, S., 2014. tictoc: Functions for timing R scripts, as well as implementations of Stack and List structures (R package version 1.0) [WWW Document]. URL https://cran.r-project.org/package=tictoc

Jacobs, C., Klok, L., Bruse, M., Cortesão, J., Lenzholzer, S., Kluck, J., 2020. Are urban water bodies really cooling? Urban Clim. 32, 100607. https://doi.org/10.1016/j.uclim.2020.100607

Järvi, L., Grimmond, C.S.B., Christen, A., 2011. The Surface Urban Energy and Water Balance Scheme (SUEWS): Evaluation in Los Angeles and Vancouver. J. Hydrol. 411, 219–237. https://doi.org/10.1016/j.jhydrol.2011.10.001

Järvi, L., Nordbo, A., Junninen, H., Riikonen, A., Moilanen, J., Nikinmaa, E., Vesala, T., 2012. Seasonal and annual variation of carbon dioxide surface fluxes in Helsinki, Finland, in 2006-2010. Atmos. Chem. Phys. 12, 8475–8489. https://doi.org/10.5194/acp-12-8475-2012

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT) [WWW Document]. URL http://srtm.csi.cgiar.org

Jasechko, S., Sharp, Z.D., Gibson, J.J., Birks, S.J., Yi, Y., Fawcett, P.J., 2013. Terrestrial water fluxes dominated by transpiration. Nature 496, 347–350. https://doi.org/10.1038/nature11983

Jennings, D.B., Jarnagin, S.T., Ebert, D.W., 2004. A modeling approach for estimating watershed impervious surface area from National Land Cover Data 92. Photogramm. Eng. Remote Sensing 70, 1295–1307. https://doi.org/10.14358/PERS.70.11.1295

Jiang, Y., Weng, Q., 2017. Estimation of hourly and daily evapotranspiration and soil moisture using downscaled LST over various urban surfaces. GIScience Remote Sens. 54, 95–117. https://doi.org/10.1080/15481603.2016.1258971

Jongen, H.J., Steeneveld, G.J., Beringer, J., Christen, A., Chrysoulakis, N., Fortuniak, K., Hong, J., Hong, J.W., Jacobs, C.M.J., Järvi, L., Meier, F., Pawlak, W., Roth, M., Theeuwes, N.E., Velasco, E., Vogt, R., Teuling, A.J., 2022. Urban Water Storage Capacity Inferred From Observed Evapotranspiration Recession. Geophys. Res. Lett. 49. https://doi.org/10.1029/2021GL096069

Kanda, M., Inagaki, A., Miyamoto, T., Gryschka, M., Raasch, S., 2013. A New Aerodynamic Parametrization for Real Urban Surfaces. Boundary-Layer Meteorol. 148, 357–377. https://doi.org/10.1007/s10546-013-9818-x

Karsisto, P., Fortelius, C., Demuzere, M., Grimmond, C.S.B., Oleson, K.W., Kouznetsov, R., Masson, V., Järvi, L., 2016. Seasonal surface urban energy balance and wintertime stability simulated using three land-surface models in the high-latitude city Helsinki. Q. J. R. Meteorol. Soc. 142, 401–417. https://doi.org/10.1002/qj.2659

Kent, C.W., Grimmond, S., Gatey, D., 2017. Aerodynamic roughness parameters in cities: Inclusion of vegetation. J. Wind Eng. Ind. Aerodyn. 169, 168–176. https://doi.org/10.1016/j.jweia.2017.07.016

Kim, Y., Johnson, M.S., Knox, S.H., Black, T.A., Dalmagro, H.J., Kang, M., Kim, J., Baldocchi, D., 2020. Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis. Glob. Chang. Biol. 26, 1499–1518. https://doi.org/10.1111/gcb.14845

Kingma, D.P., Ba, J.L., 2015. Adam: A Method for Stochastic Optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. San Diego, California, United States, pp. 1–15.

Kljun, N., Rotach, M.W., Schmid, H.P., 2002. A three-dimensional backward lagrangian footprint. Boundary-Layer Meteorol 103, 205–226.

Knaus, M., Haase, D., 2020. Green roof effects on daytime heat in a prefabricated residential neighbourhood in Berlin, Germany. Urban For. Urban Green. 53. https://doi.org/10.1016/j.ufug.2020.126738

Koken, P.J.M., Piver, W.T., Ye, F., Elixhauser, A., Olsen, L.M., Portier, C.J., 2003. Temperature, air pollution, and hospitalization for cardiovascular diseases among elderly people in Denver. Environ. Health Perspect. 111, 1312–1317. https://doi.org/10.1289/ehp.5957

Kordowski, K., Kuttler, W., 2010. Carbon dioxide fluxes over an urban park area. Atmos. Environ. 44, 2722–2730. https://doi.org/10.1016/j.atmosenv.2010.04.039

Kormann, R., Meixner, F.X., 2001. An analytical footprint model for non-neutral stratification. Boundary-Layer Meteorol. 99, 207–224. https://doi.org/10.1023/A:1018991015119

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kotthaus, S., Grimmond, C.S.B., 2014. Energy exchange in a dense urban environment - Part II: Impact of spatial heterogeneity of the surface. Urban Clim. 10, 281–307. https://doi.org/10.1016/j.uclim.2013.10.001

Kotthaus, S., Grimmond, C.S.B., 2012. Identification of Micro-scale Anthropogenic CO_2, heat and moisture sources - Processing eddy covariance fluxes for a dense urban environment.

Atmos. Environ. 57, 301–316. https://doi.org/10.1016/j.atmosenv.2012.04.024

Kovats, R.S., Hajat, S., 2008. Heat Stress and Public Health: A Critical Review. Annu. Rev. Public Health 29, 41–55. https://doi.org/10.1146/annurev.publhealth.29.020907.090843

Kraft, B., Jung, M., Körner, M., Koirala, S., Reichstein, M., 2022. Towards hybrid modeling of the global hydrological cycle. Hydrol. Earth Syst. Sci. 26, 1579–1614. https://doi.org/10.5194/hess-26-1579-2022

Krug, A., Fenner, D., Holtmann, A., Scherer, D., 2019. Occurrence and coupling of heat and ozone events and their relation to mortality rates in Berlin, Germany, between 2000 and 2014. Atmosphere (Basel). 10. https://doi.org/10.3390/atmos10060348

Kuhlemann, L.-M., 2022. Using stable isotopes for multi-scale assessment of ecohydrology in drought-affected urban water systems. Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät. https://doi.org/http://dx.doi.org/10.18452/23926

Kuhlemann, L.M., Tetzlaff, D., Smith, A., Kleinschmit, B., Soulsby, C., 2021. Using soil water isotopes to infer the influence of contrasting urban green space on ecohydrological partitioning. Hydrol. Earth Syst. Sci. 25, 927–943. https://doi.org/10.5194/hess-25-927-2021

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J. Stat. Softw. 28, 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling, 1st ed. ed. Springer, New York. https://doi.org/10.1007/978-1-4614-6849-3

Kundzewicz, Z.W., Kanae, S., Seneviratne, S.I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L.M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G.R., Kron, W., Benito, G., Honda, Y., Takahashi, K., Sherstyukov, B., 2014. Flood risk and climate change: global and regional perspectives. Hydrol. Sci. J. 59, 1–28. https://doi.org/10.1080/02626667.2013.857411

Lass, W., Haas, A., Hinkel, J., Jaeger, C., 2011. Avoiding the avoidable: Towards a European heat waves risk governance. Int. J. Disaster Risk Sci. 2, 1–14. https://doi.org/10.1007/s13753-011-0001-z

Leach, N., Coops, N.C., Obrknezev, N., 2019. Normalization method for multi-sensor high spatial and temporal resolution satellite imagery with radiometric inconsistencies. Comput. Electron. Agric. 164, 104893. https://doi.org/10.1016/j.compag.2019.104893

Li, D., Bou-Zeid, E., 2013. Synergistic Interactions between Urban Heat Islands and Heat Waves: The Impact in Cities Is Larger than the Sum of Its Parts. J. Appl. Meteorol. Climatol. 52, 2051–2064. https://doi.org/10.1175/JAMC-D-13-02.1

Liang, S., 2001. Narrowband to broadband conversions of land surface albedo I algorithms. Remote Sens. Environ. 76, 213–238. https://doi.org/10.1016/S0034-4257(00)00205-4

Lindberg, F., Grimmond, C.S.B., Gabey, A., Huang, B., Kent, C.W., Sun, T., Theeuwes, N.E., Järvi, L., Ward, H.C., Capel-Timms, I., Chang, Y., Jonsson, P., Krave, N., Liu, D., Meyer, D., Olofson, K.F.G., Tan, J., Wästberg, D., Xue, L., Zhang, Z., 2018. Urban Multi-scale Environmental Predictor (UMEP): An integrated tool for city-based climate services. Environ. Model. Softw. 99, 70–87. https://doi.org/10.1016/j.envsoft.2017.09.020

Litvak, E., Bijoor, N.S., Pataki, D.E., 2014. Adding trees to irrigated turfgrass lawns may be a water-saving measure in semi-arid environments. Ecohydrology 7, 1314–1330. https://doi.org/10.1002/eco.1458

Litvak, E., McCarthy, H.R., Pataki, D.E., 2017. A method for estimating transpiration of irrigated urban trees in California. Landsc. Urban Plan. 158, 48–61. https://doi.org/10.1016/j.landurbplan.2016.09.021

Liu, L., Jensen, M.B., 2018. Green infrastructure for sustainable urban water management: Practices of five forerunner cities. Cities 74, 126–133. https://doi.org/10.1016/j.cities.2017.11.013

Liu, W., Hong, Y., Khan, S.I., Huang, M., Vieux, B., Caliskan, S., Grout, T., 2010. Actual evapotranspiration estimation for different land use and land cover in urban regions using Landsat 5 data. J. Appl. Remote Sens. 4, 041873. https://doi.org/10.1117/1.3525566

Manoli, G., Fatichi, S., Schläpfer, M., Yu, K., Crowther, T.W., Meili, N., Burlando, P., Katul, G.G., Bou-Zeid, E., 2019. Magnitude of urban heat islands largely explained by climate and population. Nature 573, 55–60. https://doi.org/10.1038/s41586-019-1512-9

Marcilio, W.E., Eler, D.M., 2020. From explanations to feature selection: Assessing SHAP values as feature selection mechanism, in: Proceedings - 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2020. Institute of Electrical and Electronics Engineers Inc., pp. 340–347. https://doi.org/10.1109/SIBGRAPI51738.2020.00053

Marconcini, M., Heldens, W., Del Frate, F., Latini, D., Mitraka, Z., Lindberg, F., 2017. EO-based Products in Support of Urban Heat Fluxes Estimation, in: 2017 Joint Urban Remote Sensing Event (JURSE). IEEE, Dubai, United Arab Emirates, pp. 1–4. https://doi.org/10.1109/JURSE.2017.7924592

Markwitz, C., Knohl, A., Siebicke, L., 2020. Evapotranspiration over agroforestry sites in Germany. Biogeosciences 17, 5183–5208. https://doi.org/10.5194/bg-17-5183-2020

Masek, J.G., Skakun, S., Ju, J., Claverie, M., Roger, J.C., Vermote, E., Franch, B., Dungan, J.L., 2018. Harmonized Landsat Sentinel-2 (HLS) Product User's Guide (version 1.4).

Masson, V., Heldens, W., Bocher, E., Bonhomme, M., Bucher, B., Burmeister, C., de Munck, C., Esch, T., Hidalgo, J., Kanani-Sühring, F., Kwok, Y.T., Lemonsu, A., Lévy, J.P., Maronga, B., Pavlik, D., Petit, G., See, L., Schoetter, R., Tornay, N., Votsis, A., Zeidler, J., 2020. City-descriptive input data for urban climate models: Model requirements, data sources and challenges. Urban Clim. https://doi.org/10.1016/j.uclim.2019.100536

Masterton, J., Richardson, F., 1979. Humidex: a method of quantifying human discomfort due to excessive heat and humidity. Environment Canada, Atmospheric Environment, Downsview, Ontario, Canada.

Matasov, V., Marchesini, L.B., Yaroslavtsev, A., Sala, G., Fareeva, O., Seregin, I., Castaldi, S., Vasenev, V., Valentini, R., 2020. IoT monitoring of urban tree ecosystem services: Possibilities and challenges. Forests 11, 775. https://doi.org/10.3390/F11070775

McCarthy, H.R., Pataki, D.E., 2010. Drivers of variability in water use of native and non-native urban trees in the greater Los Angeles area. Urban Ecosyst. 13, 393–414. https://doi.org/10.1007/s11252-010-0127-6

Meehl, G.A., Tebaldi, C., 2004. More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. Science (80-.). 305, 994–997. https://doi.org/DOI: 10.1126/science.1098704

Meier, F., Fenner, D., Grassmann, T., Otto, M., Scherer, D., 2017. Crowdsourcing air temperature from citizen weather stations for urban climate research. Urban Clim. 19, 170–191. https://doi.org/10.1016/j.uclim.2017.01.006

Meili, N., Manoli, G., Burlando, P., Bou-Zeid, E., Chow, W.T.L., Coutts, A.M., Daly, E., Nice, K.A., Roth, M., Tapper, N.J., Velasco, E., Vivoni, E.R., Fatichi, S., 2020. An urban ecohydrological model to quantify the effect of vegetation on urban climate and hydrology (UT&C v1.0). Geosci. Model Dev. 13, 335–362. https://doi.org/10.5194/gmd-13-335-2020

Menzer, O., Meiring, W., Kyriakidis, P.C., McFadden, J.P., 2015. Annual sums of carbon dioxide exchange over a heterogeneous urban landscape through machine learning based gap-filling. Atmos. Environ. 101, 312–327. https://doi.org/10.1016/j.atmosenv.2014.11.006

Microsoft Corporation, Weston, S., 2019. doSNOW: Foreach Parallel Adaptor for the "snow" Package (R package version 1.0.18) [WWW Document]. URL https://cran.r-project.org/package=doSNOW

Middel, A., Nazarian, N., Demuzere, M., Bechtel, B., 2022. Urban Climate Informatics: An Emerging Research Field. Front. Environ. Sci. 10. https://doi.org/10.3389/fenvs.2022.867434

Miller, D.L., Alonzo, M., Roberts, D.A., Tague, C.L., McFadden, J.P., 2020. Drought response

of urban trees and turfgrass using airborne imaging spectroscopy. Remote Sens. Environ. 240. https://doi.org/10.1016/j.rse.2020.111646

Milojevic-Dupont, N., Creutzig, F., 2021. Machine learning for geographically differentiated climate change mitigation in urban areas. Sustain. Cities Soc. 64, 102526. https://doi.org/10.1016/j.scs.2020.102526

Miralles, D.G., Brutsaert, W., Dolman, A.J., Gash, J.H., 2020. On the Use of the Term "Evapotranspiration." Water Resour. Res. 56. https://doi.org/10.1029/2020WR028055

Mirzaei, P.A., Haghighat, F., 2010. Approaches to study Urban Heat Island - Abilities and limitations. Build. Environ. 45, 2192–2201. https://doi.org/10.1016/j.buildenv.2010.04.001

Moffat, A.M., Papale, D., Reichstein, M., Hollinger, D.Y., Richardson, A.D., Barr, A.G., Beckstein, C., Braswell, B.H., Churkina, G., Desai, A.R., Falge, E., Gove, J.H., Heimann, M., Hui, D., Jarvis, A.J., Kattge, J., Noormets, A., Stauch, V.J., 2007. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric. For. Meteorol. 147, 209–232. https://doi.org/10.1016/j.agrformet.2007.08.011

Moncrieff, J.B., Massheder, J.M., De Bruin, H., Elbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S., Soegaard, H., Verhoef, A., 1997. A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. J. Hydrol. 188–189, 589–611. https://doi.org/10.1016/S0022-1694(96)03194-0

Mora, C., Dousset, B., Caldwell, I.R., Powell, F.E., Geronimo, R.C., Bielecki, C.R., Counsell, C.W.W., Dietrich, B.S., Johnston, E.T., Louis, L. V., Lucas, M.P., Mckenzie, M.M., Shea, A.G., Tseng, H., Giambelluca, T.W., Leon, L.R., Hawkins, E., Trauernicht, C., 2017. Global risk of deadly heat. Nat. Clim. Chang. 7, 501–506. https://doi.org/10.1038/nclimate3322

Mostovoy, G. V., King, R.L., Reddy, K.R., Kakani, V.G., Filippova, M.G., 2006. Statistical estimation of daily maximum and minimum air temperatures from MODIS LST data over the State of Mississippi. GIScience Remote Sens. 43, 78–110. https://doi.org/10.2747/1548-1603.43.1.78

Muller, C.L., Chapman, L., Grimmond, C.S.B., Young, D.T., Cai, X., 2013. Sensors and the city: A review of urban meteorological networks. Int. J. Climatol. 33, 1585–1600. https://doi.org/10.1002/joc.3678

Muller, C.L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., Leigh, R.R., 2015. Crowdsourcing for climate and atmospheric sciences: Current status and future potential. Int. J. Climatol. 35, 3185–3203. https://doi.org/10.1002/joc.4210

Nagasubramanian, K., Jones, S., Singh, A.K., Sarkar, S., Singh, A., Ganapathysubramanian, B., 2019. Plant disease identification using explainable 3D deep learning on hyperspectral

images. Plant Methods 15, 1–10. https://doi.org/10.1186/s13007-019-0479-8

Napoly, A., Grassmann, T., Meier, F., Fenner, D., 2018. Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data. Front. Earth Sci. 6, 1–16. https://doi.org/10.3389/feart.2018.00118

Neirotti, P., De Marco, A., Cagliano, A.C., Mangano, G., Scorrano, F., 2014. Current trends in smart city initiatives: Some stylised facts. Cities 38, 25–36. https://doi.org/10.1016/j.cities.2013.12.010

Nguyen, T.T., Ngo, H.H., Guo, W., Wang, X.C., Ren, N., Li, G., Ding, J., Liang, H., 2019. Implementation of a specific urban water management - Sponge City. Sci. Total Environ. 652, 147–162. https://doi.org/10.1016/j.scitotenv.2018.10.168

Nichol, J., 2005. Remote sensing of urban heat islands by day and night. Photogramm. Eng. Remote Sensing 71, 613–621. https://doi.org/10.14358/PERS.71.5.613

Nichol, J.E., Fung, W.Y., Lam, K. se, Wong, M.S., 2009. Urban heat island diagnosis using ASTER satellite images and "in situ" air temperature. Atmos. Res. 94, 276–284. https://doi.org/10.1016/j.atmosres.2009.06.011

Nipen, T.N., Seierstad, I.A., Lussana, C., Kristiansen, J., Hov, Ø., 2020. Adopting citizen observations in operational weather prediction. Bull. Am. Meteorol. Soc. 101, E43–E57. https://doi.org/10.1175/BAMS-D-18-0237.1

Nordbo, A., Järvi, L., Vesala, T., 2012. Revised eddy covariance flux calculation methodologies - effect on urban energy balance. Tellus, Ser. B Chem. Phys. Meteorol. 64. https://doi.org/10.3402/tellusb.v64i0.18184

Norman, J.M., Kustas, W.P., Humes, K.S., 1995. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. Agric. For. Meteorol. 77, 263–293. https://doi.org/10.1016/0168-1923(95)02265-Y

Nouri, H., Beecham, S., Anderson, S., Hassanli, A.M., Kazemi, F., 2015. Remote sensing techniques for predicting evapotranspiration from mixed vegetated surfaces. Urban Water J. 12, 380–393. https://doi.org/10.1080/1573062X.2014.900092

Nouri, H., Beecham, S., Kazemi, F., Hassanli, A.M., 2013. A review of ET measurement techniques for estimating the water requirements of urban landscape vegetation. Urban Water J. 10, 247–259. https://doi.org/10.1080/1573062X.2012.726360

Nouri, H., Chavoshi Borujeni, S., Hoekstra, A.Y., 2019. The blue water footprint of urban green spaces: An example for Adelaide, Australia. Landsc. Urban Plan. 190, 103613. https://doi.org/10.1016/j.landurbplan.2019.103613

Oberascher, M., Rauch, W., Sitzenfrei, R., 2022. Towards a smart water city: A comprehensive

review of applications, data requirements, and communication technologies for integrated management. Sustain. Cities Soc. 76, 103442. https://doi.org/10.1016/j.scs.2021.103442

Oke, T.R., 1982. The energetic basis of the urban heat island (Symons Memorial Lecture, 20 May 1980). Q. Journal, R. Meteorol. Soc. 108, 1–24.

Oke, T.R., Johnson, G.T., Steyn, D.G., Watson, I.D., 1991. Simulation of surface urban heat islands under "ideal" conditions at night part 2: Diagnosis of causation. Boundary-Layer Meteorol. 56, 339–358. https://doi.org/10.1007/BF00119211

Oke, T.R., Maxwell, G.B., 1975. Urban heat island dynamics in Montreal and Vancouver. Atmos. Environ. 9, 191–200. https://doi.org/10.1016/0004-6981(75)90067-0

Oke, T.R., Mills, G., Christen, A., Voogt, J.A., 2017. Urban Climates. Cambridge University Press, Cambridge, UK. https://doi.org/https://doi.org/10.1017/9781139016476

Olmedo, G.F., Ortega-Farías, S., de la Fuente-Sáiz, D., Fonseca-Luengo, D., Fuentes-Peñailillo, F., 2016. water: Tools and Functions to Estimate Actual Evapotranspiration Using Land Surface Energy Balance Models in R. R J. 8, 352–370. https://doi.org/10.32614/rj-2016-051

Papale, D., Valentini, R., 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Glob. Chang. Biol. 9, 525–535. https://doi.org/10.1046/j.1365-2486.2003.00609.x

Parlow, E., Vogt, R., Feigenwinter, C., 2014. The urban heat island of Basel - Seen from different perspectives. Erde 145, 96–110. https://doi.org/10.12854/erde-145-8

Pataki, D.E., 2013. City trees: Urban greening needs better data. Nature. https://doi.org/10.1038/502624d

Pataki, D.E., McCarthy, H.R., Litvak, E., Pincetl, S., 2011. Transpiration of urban forests in the Los Angeles metropolitan area. Ecol. Appl. 21, 661–677. https://doi.org/10.1890/09-1717.1

Pipkins, K., Kleinschmit, B., Wessolek, G., 2017. A spatial-temporal method for assessing the energy balance dynamics of partially sealed surfaces, Geophysical Research Abstracts.

Probst, P., Boulesteix, A.L., 2018. To tune or not to tune the number of trees in random forest. J. Mach. Learn. Res. 18, 1–18.

Qin, H. peng, Li, Z. xi, Fu, G., 2013. The effects of low impact development on urban flooding under different rainfall characteristics. J. Environ. Manage. 129, 577–585. https://doi.org/10.1016/j.jenvman.2013.08.026

Qin, L., Yan, C., Yu, L., Chai, M., Wang, B., Hayat, M., Shi, Z., Gao, H., Jiang, X., Xiong, B., Mao, P., Qiu, G.Y., 2022. High-resolution spatio-temporal characteristics of urban evapotranspiration measured by unmanned aerial vehicle and infrared remote sensing.

Build. Environ. 222, 109389. https://doi.org/10.1016/j.buildenv.2022.109389

Quanz, J.A., 2018. Impact of spatial heterogeneity on energy exchange in an urban environment in Berlin, Germany. Technische Universität Berlin.

Quanz, J.A., Ulrich, S., Fenner, D., Holtmann, A., Eimermacher, J., 2018. Micro-scale variability of air temperature within a local climate zone in Berlin, Germany, during summer. Climate 6. https://doi.org/10.3390/cli6010005

R Core Team, 2020. R: A Language and Environment for Statistical Computing [WWW Document]. R Found. Stat. Comput. URL https://www.r-project.org/

Rafael, S., Martins, H., Matos, M.J., Cerqueira, M., Pio, C., Lopes, M., Borrego, C., 2020. Application of SUEWS model forced with WRF: Energy fluxes validation in urban and suburban Portuguese areas. Urban Clim. 33. https://doi.org/10.1016/j.uclim.2020.100662

Ramamurthy, P., Bou-Zeid, E., 2014. Contribution of impervious surfaces to urban evaporation. Water Resour. Res. 50, 2889–2902. https://doi.org/10.1002/2013WR013909

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ren, C., Cai, M., Li, X., Shi, Y., See, L., 2020. Developing a rapid method for 3-dimensional urban morphology extraction using open-source data. Sustain. Cities Soc. 53, 101962. https://doi.org/10.1016/j.scs.2019.101962

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography (Cop.). 40, 913–929. https://doi.org/10.1111/ecog.02881

Rocha, A.D., Groen, T.A., Skidmore, A.K., 2019. Spatially-explicit modelling with support of hyperspectral data can improve prediction of plant traits. Remote Sens. Environ. 231, 111200. https://doi.org/10.1016/j.rse.2019.05.019

Rocha, A.D., Vulova, S., Meier, F., Förster, M., Kleinschmit, B., 2022a. Mapping evapotranspirative and radiative cooling services in an urban environment. Sustain. Cities Soc. 85, 104051. https://doi.org/10.1016/j.scs.2022.104051

Rocha, A.D., Vulova, S., van der Tol, C., Förster, M., Kleinschmit, B., 2022b. Modelling hourly evapotranspiration in urban environments with SCOPE using open remote sensing and meteorological data. Hydrol. Earth Syst. Sci. 26, 1111–1129. https://doi.org/10.5194/hess-26-1111-2022

Rolnick, D., Donti, P.L., Kaack, L.H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A.S., Maharaj, T., Sherwin, E.D., Mukkavilli, S.K., Kording, K.P., Gomes, C.P., Ng, A.Y., Hassabis, D., Platt, J.C., Creutzig, F., Chayes, J., Bengio, Y., 2023. Tackling Climate Change with Machine Learning. ACM Comput. Surv. https://doi.org/10.1145/3485128

Román, M.O., Wang, Z., Sun, Q., Kalb, V., Miller, S.D., Molthan, A., Schultz, L., Bell, J., Stokes, E.C., Pandey, B., Seto, K.C., Hall, D., Oda, T., Wolfe, R.E., Lin, G., Golpayegani, N., Devadiga, S., Davidson, C., Sarkar, S., Praderas, C., Schmaltz, J., Boller, R., Stevens, J., Ramos González, O.M., Padilla, E., Alonso, J., Detrés, Y., Armstrong, R., Miranda, I., Conte, Y., Marrero, N., MacManus, K., Esch, T., Masuoka, E.J., 2018. NASA's Black Marble nighttime lights product suite. Remote Sens. Environ. 210, 113–143. https://doi.org/10.1016/j.rse.2018.03.017

Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020a. Explain it to me-facing remote sensing challenges in the bio-and geosciences with explainable machine learning, in: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Copernicus GmbH, pp. 817–824. https://doi.org/10.5194/isprs-Annals-V-3-2020-817-2020

Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020b. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access 8, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

Rumora, L., Majić, I., Miler, M., Medak, D., 2021. Spatial video remote sensing for urban vegetation mapping using vegetation indices. Urban Ecosyst. 24, 21–33. https://doi.org/10.1007/s11252-020-01002-5

Saher, R., Stephen, H., Ahmad, S., 2021. Urban evapotranspiration of green spaces in arid regions through two established approaches: a review of key drivers, advancements, limitations, and potential opportunities. Urban Water J. 18, 115–127. https://doi.org/10.1080/1573062X.2020.1857796

Sampson, C.C., Fewtrell, T.J., Duncan, A., Shaad, K., Horritt, M.S., Bates, P.D., 2012. Use of terrestrial laser scanning data to drive decimetric resolution urban inundation models. Adv. Water Resour. 41, 1–17. https://doi.org/10.1016/j.advwatres.2012.02.010

Santamouris, M., 2020. Recent progress on urban overheating and heat island research. Integrated assessment of the energy, environmental, vulnerability and health impact. Synergies with the global climate change. Energy Build. https://doi.org/10.1016/j.enbuild.2019.109482

Santamouris, M., Kolokotsa, D., 2015. On the impact of urban overheating and extreme climatic conditions on housing, energy, comfort and environmental quality of vulnerable population

in Europe. Energy Build. 98, 125–133. https://doi.org/10.1016/j.enbuild.2014.08.050

Scherer, D., Ament, F., Emeis, S., Fehrenbach, U., Leitl, B., Scherber, K., Schneider, C., Vogt, U., 2019. Three-dimensional observation of atmospheric processes in cities. Meteorol. Zeitschrift 28, 121–138. https://doi.org/10.1127/metz/2019/0911

Scherer, D., Fehrenbach, U., Lakes, T., Lauf, S., Meier, F., Schuster, C., 2013. Quantification of heat-stress related mortality hazard, vulnerability and risk in Berlin, Germany. Erde 144, 238–259. https://doi.org/10.12854/erde-144-17

Schifano, P., Cappai, G., De Sario, M., Michelozzi, P., Marino, C., Bargagli, A.M., Perucci, C.A., 2009. Susceptibility to heat wave-related mortality: a follow-up study of a cohort of elderly in Rome. Environ. Health 8, 50. https://doi.org/10.1186/1476-069x-8-50

Schmid, H.P., Oke, T.R., 1990. A model to estimate the source area contributing to turbulent exchange in the surface layer over patchy terrain. Q. J. R. Meteorol. Soc. 116, 965–988. https://doi.org/10.1002/qj.49711649409

Schmidt, A., Wrzesinsky, T., Klemm, O., 2008. Gap filling and quality assessment of CO_2 and water vapour fluxes above an urban area with radial basis function neural networks. Boundary-Layer Meteorol. 126, 389–413. https://doi.org/10.1007/s10546-007-9249-7

Schmidt, W., 1927. Die Verteilung der Minimumtemperaturen in der Frostnacht des 12.5.1927 im Gemeindegebiet von Wien. Fortschritte der Landwirtschaft 2, 681–686.

Schrotter, G., Hürzeler, C., 2020. The Digital Twin of the City of Zurich for Urban Planning. PFG - J. Photogramm. Remote Sens. Geoinf. Sci. 88, 99–112. https://doi.org/10.1007/s41064-020-00092-2

Scott, R.L., Huxman, T.E., Cable, W.L., Emmerich, W.E., 2006. Partitioning of evapotranspiration and its relation to carbon dioxide exchange in a Chihuahuan Desert shrubland. Hydrol. Process. 20, 3227–3243. https://doi.org/10.1002/hyp.6329

Seebacher, D., Miller, M., Polk, T., Fuchs, J., Keim, D.A., 2019. Visual Analytics of Volunteered Geographic Information: Detection and Investigation of Urban Heat Islands. IEEE Comput. Graph. Appl. 39, 83–95. https://doi.org/10.1109/MCG.2019.2926242

Seidel, J., Ketzler, G., Bechtel, B., Thies, B., Philipp, A., Böhner, J., Egli, S., Eisele, M., Herma, F., Langkamp, T., Petersen, E., Sachsen, T., Schlabing, D., Schneider, C., 2016. Mobile measurement techniques for local and micro-scale studies in urban and topo-climatology. Erde 147, 15–39. https://doi.org/10.12854/erde-147-2

Senate Department for Urban Development and Housing, 2017. 01.02 Impervious Soil Coverage(Sealing of Soil Surface) (Edition 2017).

SenUVK, 2021. Anteil öffentlicher Grünflächen in Berlin. Berlin.

Shafieiyoun, E., Gheysari, M., Khiadani, M., Koupai, J.A., Shojaei, P., Moomkesh, M., 2020. Assessment of reference evapotranspiration across an arid urban environment having poor data monitoring system. Hydrol. Process. 34, 4000–4016. https://doi.org/10.1002/hyp.13851

Shandas, V., Voelkel, J., Williams, J., Hoffman, J., 2019. Integrating satellite and ground measurements for predicting locations of extreme urban heat. Climate 7, 1–13. https://doi.org/10.3390/cli7010005

Shaposhnikov, D., Revich, B., Bellander, T., Bedada, G.B., Bottai, M., Kharkova, T., Kvasha, E., Lezina, E., Lind, T., Semutnikova, E., Pershagen, G., 2014. Mortality related to air pollution with the Moscow heat wave and wildfire of 2010. Epidemiology 25, 359–364. https://doi.org/10.1097/EDE.0000000000000090

Shashua-Bar, L., Pearlmutter, D., Erell, E., 2011. The influence of trees and grass on outdoor thermal comfort in a hot-arid environment. Int. J. Climatol. 31, 1498–1506. https://doi.org/10.1002/joc.2177

Shojaei, P., Gheysari, M., Nouri, H., Myers, B., Esmaeili, H., 2018. Water requirements of urban landscape plants in an arid environment: The example of a botanic garden and a forest park. Ecol. Eng. 123, 43–53. https://doi.org/10.1016/j.ecoleng.2018.08.021

Simpson, D., Lindgren, F., Rue, H., 2012. Think continuous: Markovian Gaussian models in spatial statistics. Spat. Stat. 1, 16–29. https://doi.org/10.1016/j.spasta.2012.02.003

Smith, R.B., 2010. The heat budget of the earth's surface deduced from space [WWW Document]. URL https://yceo.yale.edu/sites/default/files/files/Surface_Heat_Budget_From_Space.pdf (accessed 9.15.19).

Sobrino, J.A., Jiménez-Muñoz, J.C., Paolini, L., 2004. Land surface temperature retrieval from LANDSAT TM 5. Remote Sens. Environ. 90, 434–440. https://doi.org/10.1016/j.rse.2004.02.003

Somarakis, G., Stagakis, S., Chrysoulakis, N., 2019. ThinkNature Nature-Based Solutions Handbook. European Union. https://doi.org/10.26225/JERV-W202

Spiess, A.N., Neumeyer, N., 2010. An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: A Monte Carlo approach. BMC Pharmacol. 10, 1–11. https://doi.org/10.1186/1471-2210-10-6

Spronken-Smith, R.A., Oke, T.R., 1999. Scale modelling of nocturnal cooling in urban parks. Boundary-Layer Meteorol. 93, 287–312. https://doi.org/10.1023/A:1002001408973

Statistical Office of Berlin-Brandenburg, 2019. Inhabitants of the state of Berlin on 31 December 2018. Potsdam.

Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. Bull. Am. Meteorol. Soc. 93, 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., Eklundh, L., 2007. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. Remote Sens. Environ. 110, 262–274. https://doi.org/10.1016/j.rse.2007.02.025

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85–99. https://doi.org/10.5194/hess-6-85-2002

Sun, Y.J., Wang, J.F., Zhang, R.H., Gillies, R.R., Xue, Y., Bo, Y.C., 2005. Air temperature retrieval from remote sensing data based on thermodynamics. Theor. Appl. Climatol. 80, 37–48. https://doi.org/10.1007/s00704-004-0079-y

Takaku, J., Tadono, T., Tsutsui, K., 2014. Generation of high resolution global DSM from ALOS PRISM. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 40, 243–248. https://doi.org/10.5194/isprsarchives-XL-4-243-2014

Teuling, A.J., 2018. A Forest Evapotranspiration Paradox Investigated Using Lysimeter Data; A Forest Evapotranspiration Paradox Investigated Using Lysimeter Data. Vadose Zo. J. | Adv. Crit. Zo. Sci. https://doi.org/10.2136/v

Teuling, A.J., Seneviratne, S.I., Stöckli, R., Reichstein, M., Moors, E., Ciais, P., Luyssaert, S., Van Den Hurk, B., Ammann, C., Bernhofer, C., Dellwik, E., Gianelle, D., Gielen, B., Grünwald, T., Klumpp, K., Montagnani, L., Moureaux, C., Sottocornola, M., Wohlfahrt, G., 2010. Contrasting response of European forest and grassland energy exchange to heatwaves. Nat. Geosci. 3, 722–727. https://doi.org/10.1038/ngeo950

Timm, A., Kluge, B., Wessolek, G., 2018. Hydrological balance of paved surfaces in moist mid-latitude climate – A review. Landsc. Urban Plan. 175, 80–91. https://doi.org/10.1016/j.landurbplan.2018.03.014

Tooke, T.R., van der Laan, M., Coops, N.C., 2014. Mapping demand for residential building thermal energy services using airborne LiDAR. Appl. Energy 127, 125–134. https://doi.org/10.1016/j.apenergy.2014.03.035

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Turrini, T., Knop, E., 2015. A landscape ecology approach identifies important drivers of urban biodiversity. Glob. Chang. Biol. 21, 1652–1667. https://doi.org/10.1111/gcb.12825

U.S. Geological Survey (USGS), 2019a. U.S. Geological Survey (USGS) EarthExplorer [WWW Document]. URL https://earthexplorer.usgs.gov/ (accessed 9.1.19).

U.S. Geological Survey (USGS), 2019b. LSDS Science Research and Development (LSRD)

[WWW Document]. URL https://espa.cr.usgs.gov/ (accessed 9.5.19).

United Nations, 2019. World Urbanization Prospects: The 2018 Revision. Department of Economic and Social Affairs, Population Division, New York, United States.

United Nations, 2018. The World's Cities in 2018. World's Cities 2018 - Data Bookl. (ST/ESA/ SER.A/417).

Vapnik, V.N., 1999. An Overview of Statistical Learning Theory, IEEE TRANSACTIONS ON NEURAL NETWORKS.

Vaz Monteiro, M., Doick, K.J., Handley, P., Peace, A., 2016. The impact of greenspace size on the extent of local nocturnal air temperature cooling in London. Urban For. Urban Green. 16, 160–169. https://doi.org/10.1016/j.ufug.2016.02.008

Venter, Z.S., Brousse, O., Esau, I., Meier, F., 2020. Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. Remote Sens. Environ. 242, 111791. https://doi.org/10.1016/j.rse.2020.111791

Vesala, T., Järvi, L., Launiainen, S., Sogachev, A., Rannik, Ü., Mammarella, I., Siivola, E., Keronen, P., Rinne, J., Riikonen, A., Nikinmaa, E., 2008. Surface-atmosphere interactions over complex urban terrain in Helsinki, Finland. Tellus, Ser. B Chem. Phys. Meteorol. 60 B, 188–199. https://doi.org/10.1111/j.1600-0889.2007.00312.x

Vickers, D., Mahrt, L., 1997. Quality control and flux sampling problems for tower and aircraft data. J. Atmos. Ocean. Technol. 14, 512–526. https://doi.org/10.1175/1520-0426(1997)014<0512:QCAFSP>2.0.CO;2

Vidal, T.H.G., Gamet, P., Olioso, A., Jacob, F., 2022. Optimizing TRISHNA TIR channels configuration for improved land surface temperature and emissivity measurements. Remote Sens. Environ. 272, 112939. https://doi.org/10.1016/j.rse.2022.112939

Voogt, J.A., Oke, T.R., 2003. Thermal remote sensing of urban climates. Remote Sens. Environ. 86, 370–384. https://doi.org/10.1016/S0034-4257(03)00079-8

Vulova, S., Kleinschmit, B., 2019. Thermal behavior and its seasonal and diurnal variability of urban green infrastructure in a mid-latitude city - Berlin, in: 2019 Joint Urban Remote Sensing Event (JURSE). Vannes, France, pp. 9–12. https://doi.org/10.1109/JURSE.2019.8809011

Vulova, S., Kuhlemann, L.M., Tetzlaff, D., Soulsby, C., Kleinschmit, B., 2019. Assessment of evapotranspiration from urban vegetation across space and time: A case study in Berlin, in: 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). Shanghai, China, pp. 1–4. https://doi.org/10.1109/Multi-Temp.2019.8866903

Vulova, S., Meier, F., Fenner, D., Nouri, H., Kleinschmit, B., 2020. Summer Nights in Berlin, Germany: Modeling Air Temperature Spatially With Remote Sensing, Crowdsourced Weather Data, and Machine Learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 5074–5087. https://doi.org/10.1109/JSTARS.2020.3019696

Vulova, S., Meier, F., Rocha, A.D., Quanz, J., Nouri, H., Kleinschmit, B., 2021. Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence. Sci. Total Environ. 786, 147293. https://doi.org/10.1016/j.scitotenv.2021.147293

Wan, H., McLaughlin, D., Shao, Y., van Eerden, B., Ranganathan, S., Deng, X., 2021. Remotely-sensed evapotranspiration for informed urban forest management. Landsc. Urban Plan. 210. https://doi.org/10.1016/j.landurbplan.2021.104069

Wang, C., Yang, J., Myint, S.W., Wang, Z.H., Tong, B., 2016. Empirical modeling and spatio-temporal patterns of urban evapotranspiration for the Phoenix metropolitan area, Arizona. GIScience Remote Sens. 53, 778–792. https://doi.org/10.1080/15481603.2016.1243399

Wang, Chenghao, Wang, Z.H., Wang, Chuyuan, Myint, S.W., 2019. Environmental cooling provided by urban trees under extreme heat and cold waves in U.S. cities. Remote Sens. Environ. 227, 28–43. https://doi.org/10.1016/j.rse.2019.03.024

Wang, F., Qin, Z., Song, C., Tu, L., Karnieli, A., Zhao, S., 2015. An Improved Mono-Window Algorithm for Land Surface Temperature Retrieval from Landsat 8 Thermal Infrared Sensor Data. Remote Sens. 7, 4268–4289. https://doi.org/10.3390/rs70404268

Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. Rev. Geophys. https://doi.org/10.1029/2011RG000373

Wang, Y., Ouyang, W., 2021. Investigating the heterogeneity of water cooling effect for cooler cities. Sustain. Cities Soc. 75. https://doi.org/10.1016/j.scs.2021.103281

Ward, H.C., Evans, J.G., Grimmond, C.S.B., 2013. Multi-season eddy covariance observations of energy, water and carbon fluxes over a suburban area in Swindon, UK. Atmos. Chem. Phys. 13, 4645–4666. https://doi.org/10.5194/acp-13-4645-2013

Ward, H.C., Kotthaus, S., Järvi, L., Grimmond, C.S.B., 2016. Surface Urban Energy and Water Balance Scheme (SUEWS): Development and evaluation at two UK sites. Urban Clim. 18, 1–32. https://doi.org/10.1016/j.uclim.2016.05.001

Webb, E.K., Pearman, G.I., Leuning, R., 1980. Correction of flux measurements for density effects due to heat and water vapour transfer. Q. J. R. Meteorol. Soc. 106, 85–100. https://doi.org/10.1002/qj.49710644707

Wei, Z., Yoshimura, K., Wang, L., Miralles, D.G., Jasechko, S., Lee, X., 2017. Revisiting the

contribution of transpiration to global terrestrial evapotranspiration. Geophys. Res. Lett. 44, 2792–2801. https://doi.org/10.1002/2016GL072235

Wetherley, E.B., McFadden, J.P., Roberts, D.A., 2018. Megacity-scale analysis of urban vegetation temperatures. Remote Sens. Environ. 213, 18–33. https://doi.org/10.1016/j.rse.2018.04.051

Wicki, A., Parlow, E., Feigenwinter, C., 2018. Evaluation and modeling of urban heat island intensity in Basel, Switzerland. Climate 6. https://doi.org/10.3390/cli6030055

Wilby, R.L., 2007. A review of climate change impacts on the built environment. Built Environ. 33, 31–45. https://doi.org/10.2148/benv.33.1.31

Williams, C.A., Reichstein, M., Buchmann, N., Baldocchi, D., Beer, C., Schwalm, C., Wohlfahrt, G., Hasler, N., Bernhofer, C., Foken, T., Papale, D., Schymanski, S., Schaefer, K., 2012. Climate and vegetation controls on the surface water balance: Synthesis of evapotranspiration measured across a global network of flux towers. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011586

Wong, N.H., Tan, C.L., Kolokotsa, D.D., Takebayashi, H., 2021. Greenery as a mitigation and adaptation strategy to urban heat. Nat. Rev. Earth Environ. 2, 166–181. https://doi.org/10.1038/s43017-020-00129-5

Wouters, H., Demuzere, M., Ridder, K. De, Van Lipzig, N.P.M., 2015. The impact of impervious water-storage parametrization on urban climate modelling. Urban Clim. 11, 24–50. https://doi.org/10.1016/j.uclim.2014.11.005

Xenakis, G., 2016. FREddyPro: Post-Processing EddyPro Full Output File (R package version 1.0) [WWW Document]. URL https://cran.r-project.org/package=FREddyPro

Yebra, M., Van Dijk, A., Leuning, R., Huete, A., Guerschman, J.P., 2013. Evaluation of optical remote sensing to estimate actual evapotranspiration and canopy conductance. Remote Sens. Environ. 129, 250–261. https://doi.org/10.1016/j.rse.2012.11.004

Yoo, C., Im, J., Park, S., Quackenbush, L.J., 2018. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. ISPRS J. Photogramm. Remote Sens. 137, 149–162. https://doi.org/10.1016/j.isprsjprs.2018.01.018

Zakšek, K., Oštir, K., 2012. Downscaling land surface temperature for urban heat island diurnal cycle analysis. Remote Sens. Environ. 117, 114–124. https://doi.org/10.1016/j.rse.2011.05.027

Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. Int. J. Remote Sens. 24, 583–594. https://doi.org/10.1080/01431160304987

Zhang, F., Qiang Sun, Y., Magnusson, L., Buizza, R., Lin, S.J., Chen, J.H., Emanuel, K., 2019. What is the predictability limit of midlatitude weather? J. Atmos. Sci. 76, 1077–1091. https://doi.org/10.1175/JAS-D-18-0269.1

Zhang, J., Wu, P., Xu, X., Han, M., Pan, B., 2022. PCS-LSTM: A hybrid deep learning model for multi-stations joint temperature prediction based on periodicity and closeness. Neurocomputing 501, 151–161. https://doi.org/10.1016/j.neucom.2022.06.015

Zhang, K., Li, Y., Schwartz, J.D., O'Neill, M.S., 2014. What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods. Environ. Res. 132, 350–359. https://doi.org/10.1016/j.envres.2014.04.004

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 Sens. 11, 1–36. https://doi.org/10.3390/rs11010048

Zhou, Q., 2014. A Review of Sustainable Urban Drainage Systems Considering the Climate Change and Urbanization Impacts. Water 6, 976–992. https://doi.org/10.3390/w6040976

Zhu, X., Zhang, Q., Xu, C.Y., Sun, P., Hu, P., 2019. Reconstruction of high spatial resolution surface air temperature data across China: A new geo-intelligent multisource data-based machine learning technique. Sci. Total Environ. 665, 300–313. https://doi.org/10.1016/j.scitotenv.2019.02.077

Declaration of independent work

I hereby declare that this thesis and the work presented in it is entirely my own except where otherwise indicated. I have only used the documented utilities and references.

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

_____

Stenka Vulova

Chapter 6: Appendices

Appendix A   Modeling

urban air temperature

Supplementary material to Chapter 2: Summer nights in Berlin, Germany: Modeling air temperature spatially with remote sensing, crowdsourced weather data, and machine learning

Table A.1: Locations of reference stations used for validation. Latitude and longitude are in WGS84. Operators: DWD - Deutscher Wetterdienst, TUB - Technische Universität Berlin. Altitude is according to Shuttle Radar Topography Mission (SRTM) data (Jarvis et al., 2008). Measurement height at all sites is 2 m above ground except at Bamberger (2.5 m) and Dessauer (3.5 m). [a] Local Climate Zone (LCZ) as mapped in Fenner et al. (2017)

Site (Operator)	Longitude (°)	Latitude (°)	Altitude (m above mean sea level)	LCZ[a]
Albrecht (TUB)	13.3486	52.4446	42	6
Alexanderplatz (DWD)	13.4057	52.5198	48	2
Bamberger (TUB)	13.3376	52.4965	47	2
Buch (DWD)	13.5022	52.6309	61	6
Dahlem (DWD)	13.3017	52.4537	52	6
Dahlemerfeld (TUB)	13.2253	52.4776	54	B
Dessauer (TUB)	13.3783	52.5047	38	2
Jagen (TUB)	13.2251	52.4732	67	A
Kaniswall (DWD)	13.7309	52.4040	33	B
Koepenick (TUB)	13.6157	52.4330	43	A
Marzahn (DWD)	13.5598	52.5447	60	4
Rothenburg (TUB)	13.3158	52.4572	52	6
Schlosscharl (TUB)	13.2945	52.5213	34	6
Spandauer (TUB)	13.1584	52.5364	30	5
Swinemuender (TUB)	13.3968	52.5431	45	5
Tegel (DWD)	13.3088	52.5644	31	D
Tiergarten (TUB)	13.3636	52.5145	36	A

Tuning

Table A.2: Optimal hyperparameters for avNNet revealed during parameter tuning. Model tuning was performed using 10-fold cross-validation. The optimal values leading to the lowest RMSE based on 10-fold cross-validation were used in the final models. Date refers to the predictor date left out of model training.

Predictors	Date	Decay	Size
All	2 June 2017	0.06	12
	23 July 2018	0.06	14
	24 June 2019	0.04	12
	26 July 2019	0.06	12
GIS	2 June 2017	0.06	14
	23 July 2018	0.02	12
	24 June 2019	0.02	14
	26 July 2019	0.02	14
Tcrowd	2 June 2017	0	2
	23 July 2018	0	10
	24 June 2019	0	4
	26 July 2019	0.08	6

Table A.3: Optimal hyperparameters for GBM revealed during parameter tuning. Model tuning was performed using 10-fold cross-validation. The optimal values leading to the lowest RMSE based on 10-fold cross-validation were used in the final models. Date refers to the predictor date left out of model training.

Predictors	Date	n.trees	interaction.depth	shrinkage	n.minobsinnode
All	2 June 2017	850	7	0.01	8
	23 July 2018	1000	11	0.01	10
	24 June 2019	600	11	0.05	8
	26 July 2019	850	11	0.01	10
GIS	2 June 2017	1000	11	0.05	8
	23 July 2018	950	11	0.01	6
	24 June 2019	700	11	0.05	10
	26 July 2019	1000	11	0.05	10
Tcrowd	2 June 2017	950	1	0.01	10
	23 July 2018	600	3	0.01	8
	24 June 2019	200	1	0.05	6
	26 July 2019	1000	1	0.01	6

Table A.4: Optimal hyperparameters for RF revealed during parameter tuning. Model tuning was performed using 10-fold cross-validation. The optimal values leading to the lowest RMSE based on 10-fold cross-validation were used in the final models. Date refers to the predictor date left out of model training. RF models run with one predictor (Tcrowd) were not tuned, with hyperparameter 'mtry' kept constant at a default value of 2.

Predictors	Date	mtry
All	2 June 2017	8
	23 July 2018	10
	24 June 2019	8
	26 July 2019	10
GIS	2 June 2017	4
	23 July 2018	10
	24 June 2019	5
	26 July 2019	10

Focal Buffer Analyses

Tcrowd was averaged from 22 UTC to 2 UTC on the same day as Landsat predictors. Spearman rank correlation coefficient (Spearman's rho) was used to quantify the relationship between focal buffers and Tcrowd (Figure A.1). The advantage of the Spearman coefficient is that it is a nonparametric measure of rank correlation that assesses how well two variables are monotonically related, even if their relationship is nonlinear. The p-values of the correlations were adjusted according to the Benjamini and Hochberg (1995) method to account for multiple tests. The families adjusted for were all possible radii (9 radii) and all GIS predictors (13 predictors), or 117 families per study date.

For each parameter and study date, the absolute values of Spearman's rho coefficients (representing the strength of correlation) were ranked from highest to lowest. The three most strongly and significantly ($p < 0.05$) correlated radii per study date and parameter were assigned points based on their rank of the absolute values of Spearman's rho coefficients (e.g., highest correlation: 3 points; second-highest: 2 points, etc.). Then, the points assigned per parameter and radius were summed for all four study dates. The radius with the most points for a certain parameter was identified as the optimal radius (see Table 2.2).
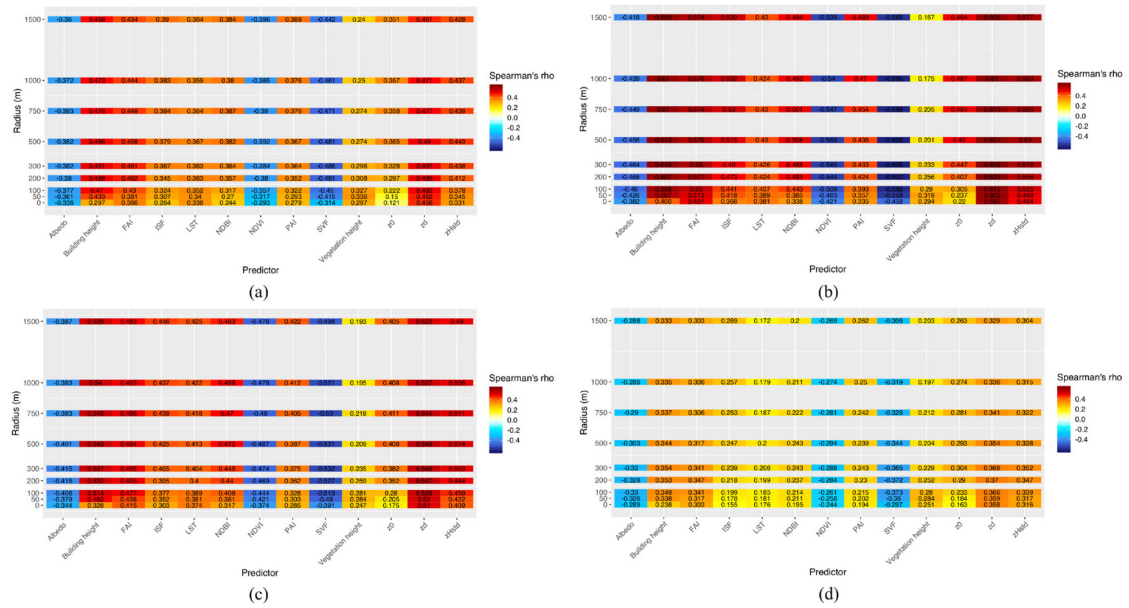
Figure A.1: Heat maps showing Spearman's rho correlation coefficients between Tcrowd and GIS predictors with different focal buffer radii for (a) 2 June 2017, (b) 23 July 2018, (c) 24 June 2019, and (d) 26 July 2019. P-values of the correlations were adjusted according to the Benjamini and Hochberg (1995) method to account for multiple tests. All Spearman correlations between GIS predictors and Tcrowd were statistically significant (all adjusted p-values < 0.05).

Table A.5: Optimal radii in meters for focal buffers for each study date based on Spearman's rho correlation analyses.

Predictor	2 June 2017	23 July 2018	24 June 2019	26 July 2019
Albedo	750	200	200	100
BH	300	300	750	300
FAI	200	300	750	200
ISF	1500	1500	1500	1500
LST	500	500	1500	300
NDBI	750	500	500	500
NDVI	1500	500	500	500
PAI	1500	1500	1500	1500
SVF	300	300	300	100
VH	50	50	50	50
z0	500	1000	750	300
zd	300	300	500	200
zHstd	500	500	500	300

Testing

Table A.6: "Leave-one-date-out" testing performance metrics for the test date 3 June 2017. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).

Predictors	Model	RMSE (K)		MAE (K)		pbias (%)		R^2 (-)	
		Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$
All	avNNet	2.05	3.29	1.86	3.25	247.80	18.60	0.657	0.751
	GBM	1.94	3.39	1.79	3.33	238.90	19.10	0.634	0.659
	RF	1.95	3.41	1.80	3.37	239.10	19.30	0.641	0.734
GIS	avNNet	1.12	1.83	0.94	1.73	108.40	9.90	0.394	0.686
	GBM	1.21	1.98	1.00	1.77	112.00	10.00	0.420	0.451
	RF	1.09	1.84	0.91	1.73	107.50	9.90	0.485	0.646
Tcrowd	avNNet	2.11	3.75	1.94	3.67	260.10	21.00	0.605	0.418
	GBM	2.02	3.75	1.87	3.66	250.20	21.00	0.578	0.392
	RF	2.11	3.76	1.93	3.65	258.60	20.90	0.588	0.259

Table A.7: "Leave-one-date-out" testing performance metrics for the test date 24 July 2018. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).

Predictors	Model	RMSE (K)		MAE (K)		pbias (%)		R^2 (-)	
		Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$
All	avNNet	0.95	2.42	0.75	1.98	-42.50	8.80	0.803	0.568
	GBM	0.95	2.22	0.75	1.91	-45.00	8.80	0.831	0.790
	RF	0.96	2.25	0.75	1.92	-45.40	9.00	0.826	0.834
GIS	avNNet	1.28	2.28	1.07	1.82	-33.70	7.70	0.491	0.654
	GBM	1.27	2.16	1.05	1.74	-35.40	7.30	0.554	0.655
	RF	1.16	2.23	0.95	1.75	-33.40	7.60	0.655	0.653
Tcrowd	avNNet	1.01	2.45	0.81	2.11	-49.10	9.30	0.821	0.677
	GBM	1.01	2.45	0.81	2.10	-49.30	9.50	0.817	0.695
	RF	1.03	2.27	0.82	1.84	-48.20	8.20	0.778	0.563

Table A.8: "Leave-one-date-out" testing performance metrics for the test date 25 June 2019. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).

Predictors	Model	RMSE (K)		MAE (K)		pbias (%)		R^2 (-)	
		Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$
All	avNNet	1.33	2.96	1.11	2.33	-16.60	8.10	0.480	0.019
	GBM	1.26	3.16	1.05	2.56	-6.90	10.00	0.572	0.045
	RF	1.22	3.09	1.03	2.47	0	10.50	0.637	0.386
GIS	avNNet	1.47	2.63	1.25	2.08	-7.00	8.40	0.411	0.543
	GBM	1.40	2.76	1.18	2.17	-5.60	8.70	0.487	0.401
	RF	1.34	2.75	1.14	2.22	-0.60	9.00	0.606	0.616
Tcrowd	avNNet	1.16	3.34	0.98	2.75	9.60	11.60	0.744	0.266
	GBM	1.16	3.32	0.98	2.73	9.60	11.60	0.749	0.325
	RF	1.16	2.94	0.96	2.37	18.10	10.20	0.693	0.521

Table A.9: "Leave-one-date-out" testing performance metrics for the test date 27 July 2019. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).

Predictors	Model	RMSE (K)		MAE (K)		pbias (%)		R^2 (-)	
		Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$	Testing$_{crowd}$	Testing$_{ref}$
All	avNNet	0.73	0.41	0.60	0.35	-45.30	0.80	0.730	0.720
	GBM	0.78	0.57	0.64	0.48	-51.90	0.70	0.739	0.490
	RF	0.81	0.47	0.66	0.40	-54.30	0.60	0.734	0.570
GIS	avNNet	1.10	1.07	0.88	0.94	-5.40	2.80	0.192	0.427
	GBM	1.13	1.27	0.92	1.14	-6.80	3.10	0.279	0.351
	RF	1.05	1.05	0.86	0.95	-10.00	2.90	0.301	0.409
Tcrowd	avNNet	0.80	0.49	0.75	0.40	-64.00	0.90	0.904	0.597
	GBM	0.81	0.49	0.75	0.40	-64.60	0.90	0.901	0.599
	RF	0.90	0.43	0.81	0.39	-68.00	0.30	0.796	0.660

LULC Analysis

Testing within each buffer size, there were significant differences between the majority of LULC classes (Dunn's test, adjusted p-values < 0.05). All differences between LULC classes were significant at the 0-5 km buffer. At the >5-10 km buffer, classes 12100 (mainly industrial areas) and 50000 (water) were not significantly different. At the >10-20 km buffer, classes 12220 and 14100 were not significantly different. At >20-30 km, class 11100 was not significantly different from three other classes (11220, 12100, and 12220) and class 11210 was not significantly different from class 14200.

Performance without Vegetation Height

There was no significant difference between models with VH as a predictor and without VH based on all performance metrics for both testing$_{crowd}$ and testing$_{ref}$ (Figure A.2; Figure A.3) (Wilcoxon rank sum tests, p > 0.05).
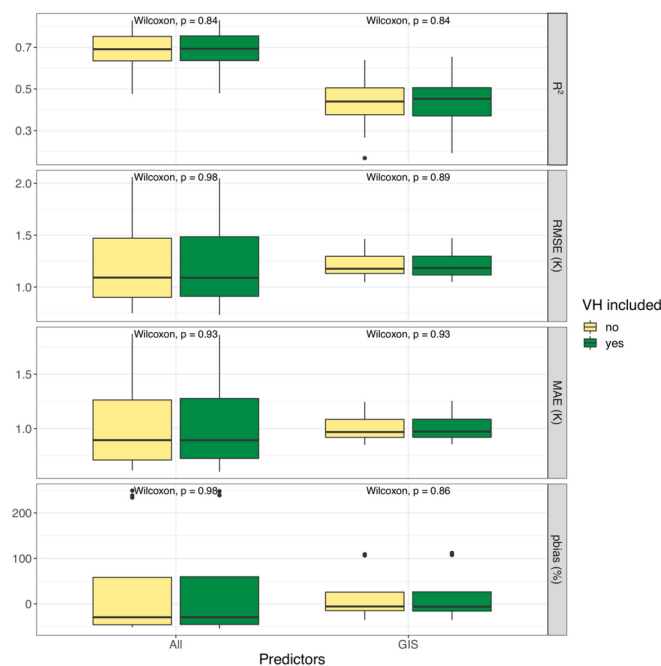
Figure A.2: Comparison of the Tair prediction performances of the three ML algorithms for testing$_{crowd}$ with all or only GIS predictors and with/ without vegetation height (VH) as a predictor. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).
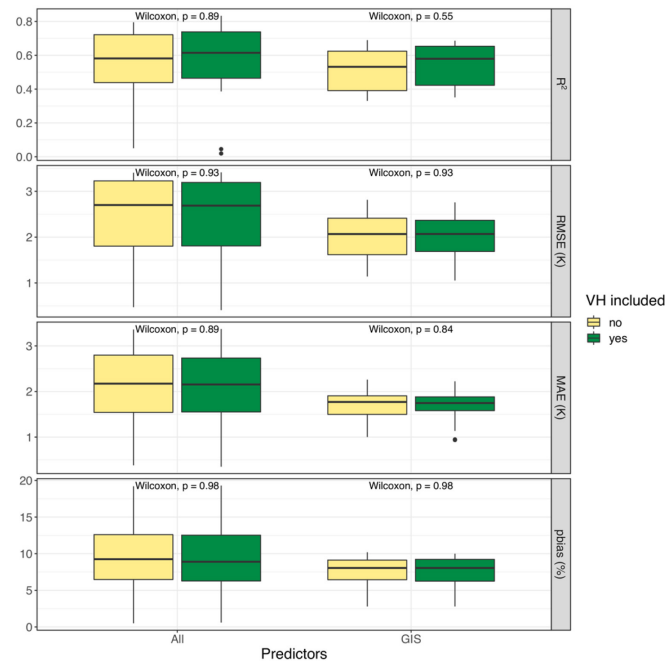
Figure A.3: Comparison of the Tair prediction performances of the three ML algorithms for testing$_{ref}$ with all or only GIS predictors and with/without vegetation height (VH) as a predictor. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).

Processing Time

Table A.10: Processing time in minutes for model tuning and training as well as just model training with the optimal hyperparameters. The date refers to the test date left out of training.

Predictors	Model	2 June 2017		23 July 2018		24 June 2019		26 July 2019	
		Tune and train	Train with optimal	Tune and train	Train with optimal	Tune and train	Train with optimal	Tune and train	Train with optimal
All	avNNet	17.18	0.17	14.89	0.19	17.81	0.21	14.21	0.18
	GBM	4.15	0.07	3.83	0.11	3.92	0.07	3.88	0.07
	RF	6.04	0.39	5.47	0.45	5.62	0.34	5.14	0.40
GIS	avNNet	10.76	0.16	11.17	0.12	10.61	0.16	9.77	0.14
	GBM	4.00	0.10	4.07	0.10	4.00	0.08	3.71	0.10
	RF	5.48	0.29	5.94	0.46	5.41	0.29	5.01	0.44
Tcrowd	avNNet	4.60	0.04	5.05	0.09	4.88	0.05	4.80	0.05
	GBM	0.69	0.01	0.73	0.01	0.68	0.01	0.68	0.01
	RF	-	0.16	-	0.17	-	0.16	-	0.16

Factoring Influencing Testing Accuracy

In order to analyze how testing_ref performance differs based on land cover, the Urban Atlas 2012 LULC classes of the reference weather stations (Table A.1) were grouped into four main land cover classes based on similarity (Table A.12). The new class "Urban fabric (S.L. 50% - 100%)" was a combination of LULC classes "Continuous urban fabric (S.L.: >80%)", "Discontinuous dense urban fabric (S.L.: 50% - 80%)", and "Airports." The new class "Urban fabric (S.L. 10% - 50%)" was a combination of LULC classes "Discontinuous medium density urban fabric (S.L.: 30% - 50%)", "Discontinuous low density urban fabric (S.L. : 10% - 30%)", and "Sports and leisure facilities." The new class "Urban meadows" was a combination of LULC classes "Green urban areas", "Pastures", and "Herbaceous vegetation associations." The new class "Forests" was based only on the LULC class "Forests."

Table A.11: Testing_ref performance metrics averaged for buffers representing distances (in km) from the geometric city center. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).

Buffer (km)	RMSE (K)	MAE (K)	pbias (%)	R^2 (-)	No. of stations	No. of data points
0-5	1.58	1.18	3.50	0.649	5	171
5-10	2.70	2.22	11.20	0.333	5	162
10-20	2.83	2.40	12.20	0.358	6	180
20-30	3.96	3.87	23.70	0.806	1	18

Table A.12: Testing_ref performance metrics averaged for land cover classes. The performance metrics are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and percent bias (pbias).

Land cover class	RMSE (K)	MAE (K)	pbias (%)	R^2 (-)	No. of stations	No. of data points
Urban fabric (S.L. 50% - 100%)	1.76	1.39	5.9	0.759	6	189
Urban fabric (S.L. 10% - 50%)	2.63	2.15	10.8	0.421	3	90
Urban meadows	2.93	2.40	10.8	0.331	6	198
Forests	2.79	2.48	12.7	0.362	2	54

Modeled Air Temperature and Urban Land Cover

The analysis of the relationship between T_{mod} and GIS predictors reinforced the importance of urban morphology to Tair variation (see Figure A.4). For 3 June 2017, the most strongly correlated GIS predictors were zd, LST, and zHstd. The most strongly correlated GIS predictors for 24 July 2018, 25 June 2019, and 27 July 2019 were zd, zHstd, and SVF.
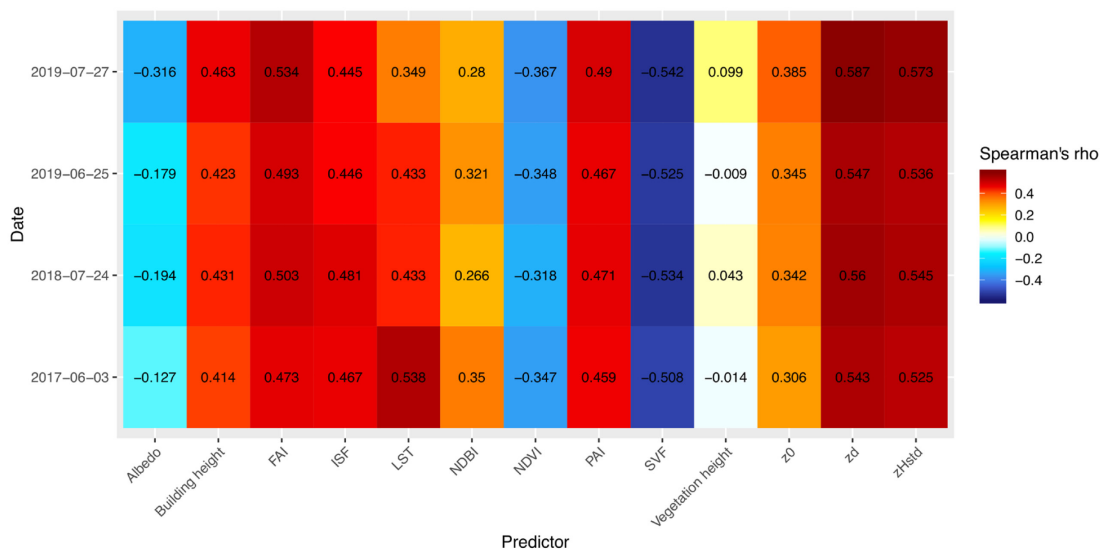

Figure A.4: Heat map showing Spearman's rho correlation coefficients between modelled air temperature (T_{mod}) and GIS predictors (from the previous day, in the case of Landsat predictors) without focal buffers for the four study dates. All Spearman correlations between GIS predictors and T_{mod} were statistically significant (all adjusted p-values < 0.05).

Appendix B   Modeling

urban ET

Supplementary material to Chapter 3: Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence

Reference evapotranspiration (ETo) parametrization

Reference ET (ETo) is defined as the ET rate from a homogeneous surface of dense and actively growing vegetation with a specific height and surface resistance, which is not constrained by water availability (Allen et al., 2005). A standardized calculation procedure for ETo based on the Penman-Monteith equation has been developed by the American Society of Civil Engineers (ASCE), which simplifies the vegetated surface for ETo as two types of hypothetical crops with predefined characteristics for reproducibility and convenience (Allen et al., 2005, 1998, 1994). The two hypothetical crop types are a short crop with a height of 0.12 m (similar to a clipped, cool-season grass) and a tall crop with a height of 0.50 m (similar to full-cover alfalfa) (Allen et al., 2005). In our study, preliminary exploratory data analysis revealed a higher correlation between measured ET and ETo for short crops rather than ETo for tall crops, with a Pearson's r of 0.862 and 0.841 at ROTH and 0.697 and 0.689 at TUCC, respectively. Reference evapotranspiration (ETo) was calculated with the hourly ASCE "Standardized Reference Evapotranspiration Equation" for short crops (Allen et al., 2005) using the "hourlyET" function of the R package "water" (Olmedo et al., 2016). Air temperature, wind speed, relative humidity, and shortwave downward radiation from the DWD were used as input. Hourly ETo was computed as:

$$ETo = \frac{0.408\,\Delta(R_n - G) + \gamma \dfrac{C_n}{T + 273}\,u_2(e_s - e_a)}{\Delta + \gamma(1 + C_d\,u_2)} \tag{1}$$

where R_n is the calculated net radiation at the crop surface (MJ m^{-2} h^{-1}); G is the soil flux density at the soil surface (MJ m^{-2} h^{-1}); T is mean hourly air temperature (°C); u_2 is mean hourly wind speed (m s^{-1}); e_s is saturation vapor pressure (kPa); e_a is actual vapor pressure (kPa); Δ is the slope

of the saturation vapor pressure-temperature curve (kPa °C^{-1}); γ is the psychometric constant (kPa °C^{-1}); C_n is a numerator constant that changes with the crop type (K mm s^3 Mg^{-1} h^{-1}); C_d is a denominator constant that changes with the crop type (s m^{-1}) (Allen et al., 2005).

Saturated vapor pressure in kPa (e_s) is computed as:

$$e_s = e^o(T) = 0.6108 \, exp\left(\frac{17.27 \, T}{T + 237.3}\right) \tag{1}$$

where e^o(T) is the saturation vapor pressure function and T is mean hourly air temperature (°C).

Actual vapor pressure in kPa (e_a) is calculated as:

$$e_a = \frac{RH}{100} \, e^o(T) \tag{2}$$

where RH is mean hourly relative humidity (%).

Net radiation in MJ m^{-2} h^{-1} (R_n) is calculated as:

$$R_n = R_{ns} - R_{nl} \tag{3}$$

where R$_{ns}$ is net shortwave radiation (MJ m^{-2} h^{-1}) and R$_{nl}$ is net outgoing longwave radiation (MJ m^{-2} h^{-1}).

R$_{ns}$ is calculated as:

$$R_{ns} = R_s - \alpha \, R_s = (1 - \alpha)R_s \tag{4}$$

where α is albedo or the canopy reflection coefficient and R$_s$ is the incoming solar radiation (MJ m^{-2} h^{-1}). A constant value of 0.23 is used for α according to Allen et al. (2005).

R$_{nl}$ is calculated as:

$$R_{nl} = \sigma \, f_{cd}\left(0.34 - 0.14 \, \sqrt{e_a}\right) T_{K \, hr}^4 \tag{5}$$

where σ is the Stefan-Boltzmann constant (2.042×10^{-10} MJ K^{-4} m^{-2} h^{-1}); f_{cd} is the cloudiness function (dimensionless); $T_{K\,hr}$ is the mean absolute temperature during the hourly period (K).

Soil heat flux density (G) is calculated as:

$$G_{hr\ daytime} = 0.1\ R_n \tag{6}$$

$$G_{hr\ nighttime} = 0.5\ R_n \tag{7}$$

where G and R_n are both in units of MJ m^{-2} h^{-1}. Nighttime is defined as when R_n is negative.

For further details on the equations and parametrization of ETo, please refer to Allen et al. (2005).

Convolutional neural networks (CNN) hyperparameter tuning

To avert convergence problems, feature-wise normalization was applied for each column in the input data matrix, subtracting the mean of the feature and dividing by the standard deviation (Chollet and Allaire, 2018). The mean and standard deviation used for centering and scaling were always computed using the training dataset.

Hyperparameter tuning for 1D convolutional neural networks (CNNs) was performed using the R package "tfruns" (Allaire, 2018), with tested values based upon the recommendations of Chollet and Allaire (2018). The following values were tested for the number of filters, kernel size, batch size, and pool size, respectively: (32, 48, 64), (3, 5, 7, 9), (16, 32, 128), and (2, 3). Each hyperparameter was kept constant across different layers (e.g., all convolutional layers have 32 filters in one scenario). The validation loss (mean squared error (MSE)) was utilized to select the optimal hyperparameters using a random 20% subset of the training data. If MSE did not differ between different hyperparameter combinations, mean absolute error (MAE) was used to select between the hyperparameter combinations with the lowest MSE. Optimal hyperparameters determined by this procedure are shown in Table B.1.

Table B.1: Optimal hyperparameters for 1D convolutional neural networks (CNNs) revealed during parameter tuning.

Tower (train)	Predictors	Filters		Kernel size		Batch size		Pool size	
		Train 2018/2020	Train 2019	Train 2018/2020	Train 2019	Train 2018/2020	Train 2019	Train 2018/2020	Train 2019
ROTH	ETo	64	32	7	5	128	128	3	3
	ETo and GIS	64	48	5	9	128	16	2	2
	Met	48	48	7	9	16	16	2	3
	Met and GIS	64	64	7	7	32	32	3	3
TUCC	ETo	64	48	5	3	16	32	2	3
	ETo and GIS	48	48	9	7	16	16	3	2
	Met	64	48	7	9	32	32	2	3
	Met and GIS	64	48	9	9	32	16	3	3

Random forest (RF) hyperparameter tuning

As with CNNs, the predictors were centered and scaled prior to modeling with RF. The number of predictors randomly selected at each split (mtry) was tuned for each value between two and the number of predictors, as recommended by Kuhn and Johnson (2013). The root mean square error (RMSE) was used to determine the optimal hyperparameters during ten-fold cross-validation on the training data. Optimal mtry values determined by tuning are given in Table B.2. Random forest (RF) models with a single predictor (ETo) were not tuned, with mtry kept constant at a default value of 2. The number of trees (ntree) was set to the default "caret" R package value of 500 (Kuhn, 2008), as tuning ntree is demonstrated to be unnecessary and the most substantial performance gain is reached with the first 100 trees (Probst and Boulesteix, 2018).

Table B.2: Optimal hyperparameters for RF revealed during parameter tuning.

Tower (train)	Predictors	mtry	
		Train 2018/ 2020	Train 2019
ROTH	ETo and GIS	2	2
	Met	13	4
	Met and GIS	11	5
TUCC	ETo and GIS	2	2
	Met	3	2
	Met and GIS	4	2

Harmonized Landsat and Sentinel-2 (HLS) data preprocessing

The Harmonized Landsat and Sentinel-2 (HLS) product provides a harmonized "Virtual Constellation" (VC) of the Landsat 8 Operational Land Imager (OLI) and Sentinel-2 Multi-Spectral Instrument (MSI) sensors by standardizing the products to a common pixel resolution, map projection, atmospheric correction, cloud mask, and nadir view geometry (Claverie et al., 2018).

Cloud, cirrus, and adjacent cloud pixels were masked out using the Quality Assessment (QA) layer (Claverie et al., 2018; Masek et al., 2018). Next, the Normalized Difference Vegetation Index (NDVI), which represents the reflection properties of vegetated surfaces in the visible red (RED) and near-infrared (NIR) range (Tucker, 1979), was computed:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{8}$$

where NIR is band 5 and band 8A for Landsat 8 OLI and Sentinel-2 MSI, respectively, and RED is band 4 for both sensors.

Lastly, each pixel was linearly interpolated to daily values under the assumption that the day-to-day fluctuations of NDVI were small. The resulting daily rasters were subsequently used to extract weighted averages of NDVI every half-hour by flux footprint modeling.

Flux footprint modeling

Footprint size, shape, and extent are controlled by the sensor height, wind direction, wind speed, atmospheric stability, surface roughness, and surface material composition (Kljun et al., 2002). To calculate footprints, the functions required wind speed and direction, z' (the effective measurement height), the standard deviation of the cross-stream wind component (sigmaV), friction velocity (u*) measured by the IRGASON, and the Monin-Obukhov stability parameter (ζ) as an indicator for atmospheric stability (Kormann and Meixner, 2001; Xenakis, 2016). The fetch size was set to 1500 meters and the resolution for the output grid was set to ten meters. All input meteorological data for footprint modeling were derived from the flux towers. SigmaV was calculated as the square root of northward wind.

The parameter z' was calculated as:

$$z' = zm - zd \tag{9}$$

where zm is the measurement height and zd is the zero-plane displacement length (Kotthaus and Grimmond, 2014).

The stability parameter ζ was calculated according to the equation from Oke et al. (2017):

$$\zeta = \frac{z'}{L_o} \qquad (10)$$

where L_o is the Obukhov length. The computation of L_o was executed within EddyPro.

The aerodynamic parameter zd (representing the height above ground where the wind speed is zero as a result of obstacles to the flow) used to calculate z' was estimated in an iterative process for each 30-minute interval (Grimmond and Oke, 1999). Morphometric parameters for each site were computed using the Urban Multi-scale Environmental Predictor (UMEP) QGIS plug-in (Lindberg et al., 2018) with a 1000 m radius and a 1° wind direction search interval, producing databases of morphometric parameters for 360 wind directions. Three different rasters were used as input for UMEP: vegetation height (VH), building height (BH), and a raster combining VH and BH (representing the maximum height of either VH or BH at any given pixel). Zd was calculated following the method of Kanda et al. (2013). Using zd only calculated from BH with the method of Kanda et al. (2013) would underestimate zd, as trees in the surrounding area increase the aerodynamic roughness. Therefore, the method from Kent et al. (2017) was used to correct zd to incorporate vegetation as a roughness element. A different aerodynamic porosity (P_{3D}) value, which is used to estimate the roughness of vegetation, was used for each season (winter, summer, and intermediate) due to changes in tree foliage (Kent et al., 2017). For further details refer to Kent et al. (2017) and Quanz (2018). The seasonally corrected one-degree zd values were then averaged over a running mean including the zd values from wind direction i±5° moving forward by one degree. According to the previously described method for zd calculation, zd was calculated for a 50x50 m grid within UMEP, generating zd grids for summer, winter, and the intermediate seasons.

These 'first-guess' estimates of zd values were assigned to the half-hourly EC data based upon wind direction and season and used to calculate initial 'first-guess' footprints. The output of footprint modeling are grids with a ten-meter resolution for the 1500 m surrounding area, where each grid cell holds the footprint probability. Footprint grid cells contributing to 90% of the signal were used to extract a weighted average of the zd grids depending on the season (Quanz, 2018).

The zd extracted from grids using 'first-guess' footprints differs from running mean zd and leads to a more precise description of the footprint area according to Kotthaus and Grimmond (2014). This grid-derived zd was used to run Kormann and Meixner (2001) models again,

producing 'final guess' footprint probability grids. The 'final guess' footprint grids were used to extract weighted averages of the surface cover (from a raster stack of six remote sensing and GIS predictors).

Evapotranspiration (ET) performance metrics

Table B.3: Testing performance metrics for training in 2018/ 2020 and testing in 2019. The performance metrics are root mean square error (RMSE), mean absolute error (MAE), percent bias (pbias), coefficient of determination (R^2), and normalized root mean square error (NRMSE). The best performance metrics for each tower training and testing combination (e.g., training in ROTH and testing in TUCC) are shown in bold.

Tower (train)	Tower (test)	Predictors	RMSE (mm/ hour)		MAE (mm/ hour)		pbias (%)		R^2 (-)		NRMSE (%)	
			CNN	RF	CNN	RF	CNN	RF	CNN	RF	CNN	RF
ROTH	ROTH	ETo	0.0306	0.0353	0.0190	0.0221	-12.00	-5.90	0.754	0.663	10.70	12.30
		ETo and GIS	0.0264	0.0267	0.0171	0.0170	-2.10	-3.70	0.809	0.808	9.20	9.30
		Met	0.0281	0.0255	0.0186	0.0162	5.20	-1.80	0.784	0.823	9.80	8.90
		Met and GIS	0.0258	**0.0241**	0.0166	**0.0153**	**0.00**	-1.40	0.816	**0.842**	9.00	**8.40**
TUCC	TUCC	ETo	0.0181	0.0204	0.0115	0.0138	-14.50	-2.20	0.526	0.380	11.50	12.90
		ETo and GIS	0.0173	0.0173	**0.0111**	0.0114	-9.30	**-1.10**	0.551	0.535	11.00	11.00
		Met	0.0181	0.0170	0.0125	0.0117	7.10	4.50	0.510	0.555	11.40	10.80
		Met and GIS	0.0175	**0.0167**	0.0114	0.0113	-4.80	2.20	0.531	**0.570**	11.10	**10.60**
ROTH	TUCC	ETo	0.0405	0.0446	0.0241	0.0265	62.70	75.40	0.524	0.490	25.60	28.20
		ETo and GIS	0.0758	**0.0217**	0.0464	**0.0147**	-168.40	**-12.30**	0.051	0.385	48.00	**13.80**
		Met	0.0471	0.0444	0.0295	0.0271	102.00	86.10	0.543	**0.558**	29.80	28.10
		Met and GIS	0.0325	0.0300	0.0230	0.0194	-50.20	50.40	0.250	0.548	20.50	19.00
TUCC	ROTH	ETo	0.0553	0.0545	0.0320	0.0320	-54.90	-49.10	0.749	0.510	19.30	19.10
		ETo and GIS	0.0614	0.0539	0.0365	0.0320	-63.20	-53.50	0.583	0.726	21.50	18.80
		Met	0.0517	**0.0506**	0.0317	**0.0296**	**-42.90**	-44.80	0.580	0.760	18.10	**17.70**
		Met and GIS	0.0592	0.0510	0.0355	0.0297	-61.50	-47.90	0.664	**0.782**	20.70	17.80

Table B.4: Testing performance metrics for training in 2019 and testing in 2018/ 2020. The performance metrics are root mean square error (RMSE), mean absolute error (MAE), percent bias (pbias), coefficient of determination (R^2), and normalized root mean square

error (NRMSE). The best performance metrics for each tower training and testing combination (e.g., training in ROTH and testing in TUCC) are shown in bold.

| Tower (train)	Tower (test)	Predictors	RMSE (mm/hour)		MAE (mm/hour)		pbias (%)		R^2 (-)		NRMSE (%)	
			CNN	RF	CNN	RF	CNN	RF	CNN	RF	CNN	RF
ROTH	ROTH	ETo	0.0303	0.0358	0.0201	0.0231	**-0.40**	7.80	0.735	0.656	10.40	12.30
		ETo and GIS	0.0269	0.0259	0.0184	0.0169	4.20	3.30	0.794	0.803	9.30	8.90
		Met	0.0266	0.0256	0.0169	0.0166	4.10	6.50	0.802	0.812	9.20	8.80
		Met and GIS	0.0241	**0.0237**	**0.0154**	**0.0154**	-6.30	5.00	0.832	**0.837**	8.30	**8.20**
TUCC	TUCC	ETo	0.0191	0.0211	0.0128	0.0143	6.50	3.80	0.447	0.340	11.90	13.20
		ETo and GIS	0.0175	0.0176	0.0117	0.0119	**0.00**	3.90	0.499	0.494	10.90	11.00
		Met	0.0175	0.0175	**0.0113**	0.0121	-8.30	7.30	0.506	0.502	10.90	10.90
		Met and GIS	0.0181	**0.0172**	0.0123	0.0117	5.50	6.70	0.474	**0.518**	11.20	**10.70**
ROTH	TUCC	ETo	0.0394	0.0504	0.0254	0.0305	79.10	92.50	0.447	0.415	24.50	31.40
		ETo and GIS	0.0511	**0.0235**	0.0332	**0.0158**	-105.80	**-2.60**	0.082	0.353	31.80	**14.60**
		Met	0.0475	0.0447	0.0286	0.0276	79.00	88.20	0.471	**0.484**	29.60	27.80
		Met and GIS	0.0947	0.0335	0.0556	0.0210	-187.80	53.80	0.039	0.481	58.90	20.80
TUCC	ROTH	ETo	**0.0465**	0.0503	**0.0274**	0.0301	**-42.30**	-44.20	0.739	0.539	**16.00**	17.30
		ETo and GIS	0.0552	0.0538	0.0333	0.0323	-60.20	-56.00	0.635	0.680	19.00	18.50
		Met	0.0515	0.0475	0.0308	0.0281	-51.30	-43.20	0.749	**0.790**	17.80	16.40
		Met and GIS	0.0573	0.0500	0.0348	0.0298	-60.80	-50.40	0.528	0.778	19.70	17.20
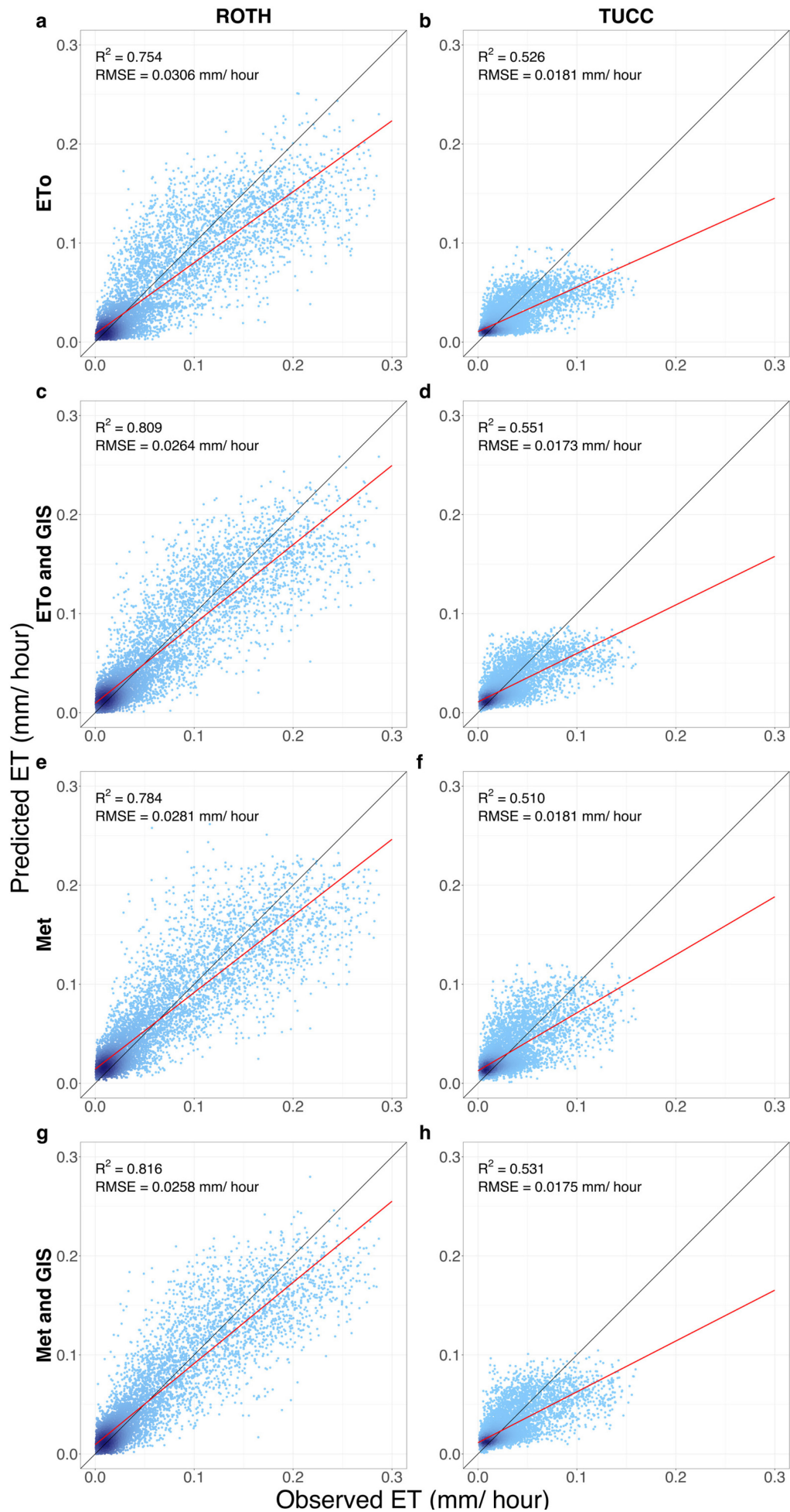
Figure B.1: Scatterplots showing the observed and predicted ET for 1D convolutional neural networks (CNN) models where training was performed in 2018/ 2020 and testing was performed for the year 2019. Training and testing were conducted on the same tower. The following scenarios are depicted: (a) "ETo" predictor scenario in ROTH, (b) "ETo" predictor scenario in TUCC, (c) "ETo and GIS" (ETo and GIS predictors) predictor scenario in ROTH, (d) "ETo and GIS" predictor scenario in TUCC, (e) "Met" (meteorological predictors) predictor scenario in ROTH, (f) "Met" predictor scenario in TUCC, (g) "Met and GIS" (meteorological and GIS predictors) predictor scenario in ROTH, and (h) "Met and GIS" predictor scenario in TUCC. The intensity of the blue color represents the data point density. The red line indicates the best fit (linear model), and the solid black line indicates the 1:1 line.
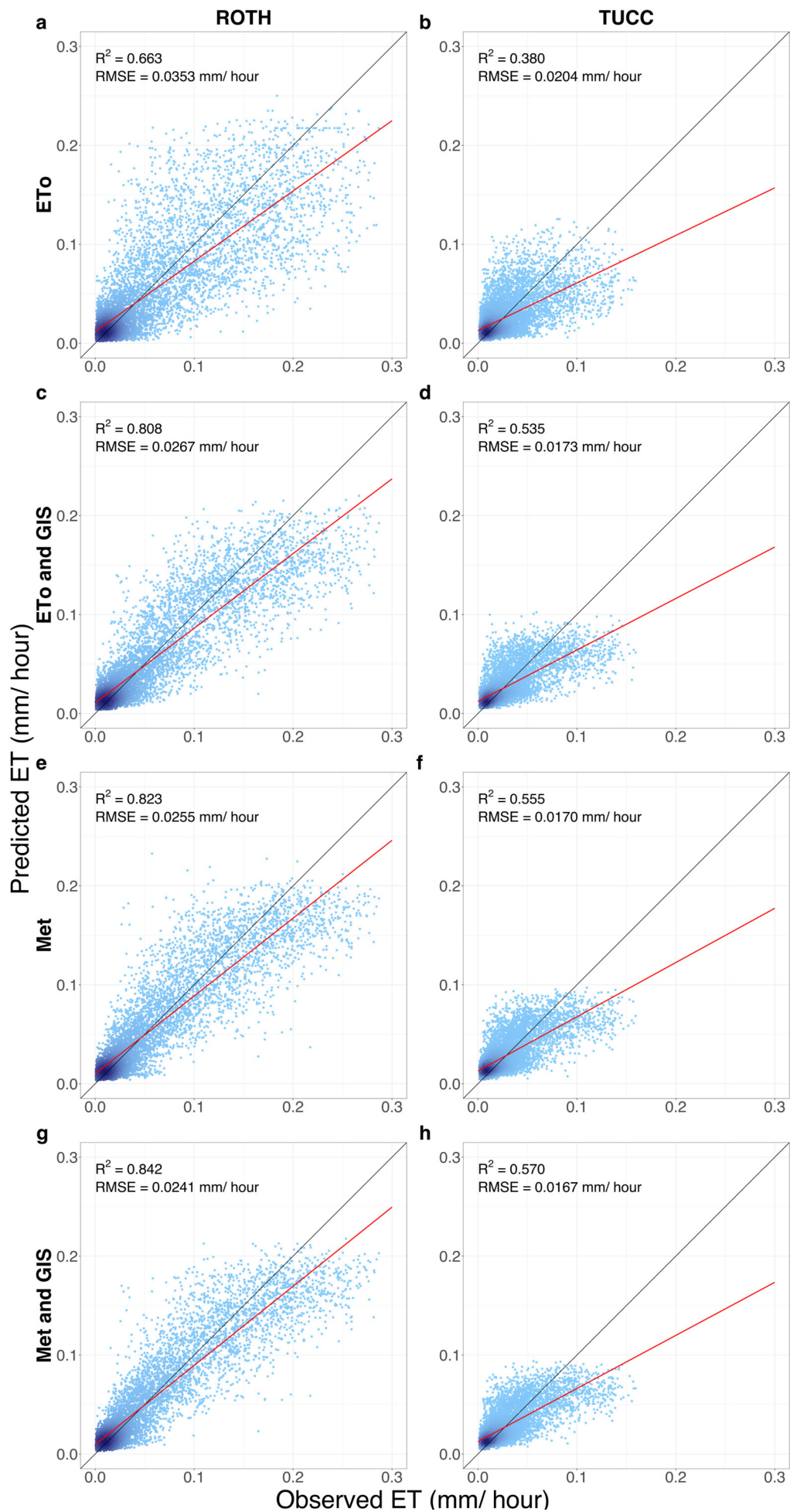
Figure B.2: Scatterplots showing the observed and predicted ET for random forest (RF) models where training was performed in 2018/ 2020 and testing was performed for the year 2019. Training and testing were conducted on the same tower. The following scenarios are depicted: (a) "ETo" predictor scenario in ROTH, (b) "ETo" predictor scenario in TUCC, (c) "ETo and GIS" (ETo and GIS predictors) predictor scenario in ROTH, (d) "ETo and GIS" predictor scenario in TUCC, (e) "Met" (meteorological predictors) predictor scenario in ROTH, (f) "Met" predictor scenario in TUCC, (g) "Met and GIS" (meteorological and GIS predictors) predictor scenario in ROTH, and (h) "Met and GIS" predictor scenario in TUCC. The intensity of the blue color represents the data point density. The red line indicates the best fit (linear model), and the solid black line indicates the 1:1 line.
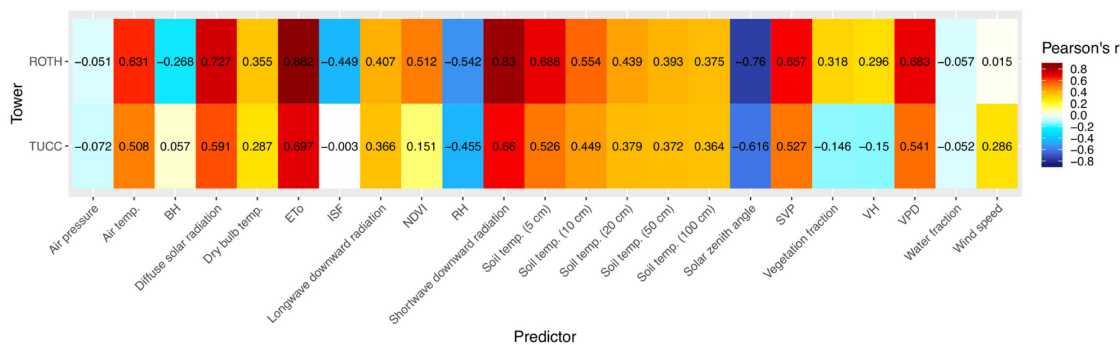

Figure B.3: Heat map showing Pearson's correlation coefficients between evapotranspiration (ET) and predictors. The entire datasets used for modeling (from 1 June 2018 to 1 June 2020 with quality control filtering applied) were used to compute the correlation coefficients. Two of the correlations with ET were statistically insignificant (p-values > 0.05): wind speed at the ROTH tower and impervious surface fraction (ISF) at the TUCC tower.

Precipitation and gap-filled ET

There were significant differences between gap-filled ET during precipitation versus during no precipitation in all cases (Figure B.4) (Wilcoxon rank sum tests, p < 0.05). Gap-filled ET was lower on average during precipitation events in all cases (Table 7; Figure B.4).
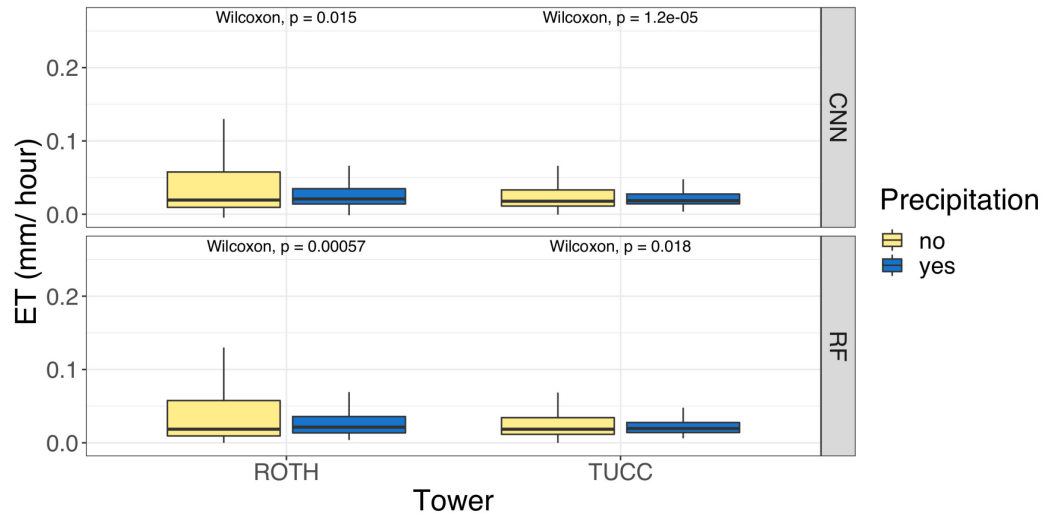
Figure B.4: Comparison of gap-filled evapotranspiration (ET) for the year 2019 when precipitation has fallen or not. Wilcoxon rank sum tests were applied to each gap-filled dataset (with a single tower and model) to determine if there were significant differences in gap-filled ET in the precipitation versus the no-precipitation group.

Table B.5: Mean and standard deviation (SD) of gap-filled evapotranspiration (ET) for the year 2019 when precipitation has fallen (Y) or not (N).

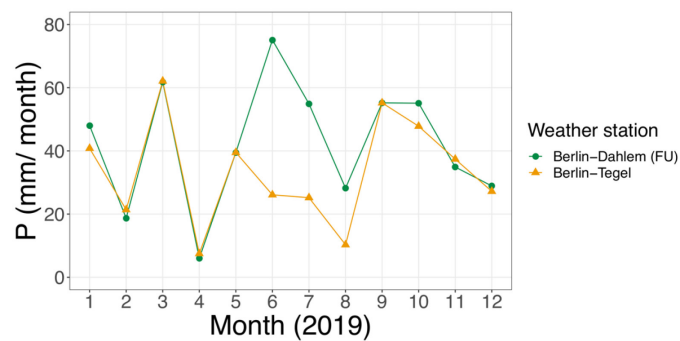| Tower	Model	Precipitation (Y/N)	Mean ET	SD ET
ROTH	CNN	N	0.0447	0.0544
		Y	0.0308	0.0291
	RF	N	0.0443	0.0537
		Y	0.0316	0.0291
TUCC	CNN	N	0.0257	0.0212
		Y	0.0232	0.0133
	RF	N	0.0261	0.0210
		Y	0.0230	0.0126

Monthly precipitation

Figure B.5: Monthly precipitation (P) sums in 2019 from two German Meteorological Service (DWD) weather stations in Berlin.

Temporal modeling accuracy

Modeling accuracy varies considerably seasonally and between the two towers, whereas it is similar between the two AI algorithms. The monthly averages of the performance metrics RMSE, R^2, NRMSE, and pbias with the "Met and GIS" predictor scenario based on the testing set (when training and testing on the same tower) are shown in Figure B.6. In order to assess the variation in accuracy between the two towers, only timestamps which were available for both towers in the testing datasets were used, leaving 5731 timestamps for the 2019 testing dataset and 5948 timestamps for the 2018/ 2020 testing dataset. For both towers, RMSE is highest in the summer months and lowest in the winter months. As an example, modeling with RF at ROTH, the highest RMSE is in August (0.0309 mm/hour), whereas the lowest RMSE is in December (0.0086 mm/ hour). R^2 is highest in summertime and lowest in wintertime for both towers. With RF at ROTH, June has the highest R^2 (0.870), whereas December has the lowest R^2 (0.505). This finding is in line with Ward et al. (2016), who also reported a higher R^2 (~0.8 for the suburban site) in summertime than in wintertime. Both higher RMSE and higher R^2 values in the warmer months can largely be explained by the higher variability of ET in summertime.

When normalizing RMSE using the range (NRMSE), the less vegetated site (TUCC) presents proportionally higher error than the more vegetated site (ROTH). With NRMSE, the seasonal component of error is not as pronounced as with RMSE and R^2. At ROTH, the highest NRMSE occurs in August, November, and September. At TUCC, NRMSE still implies a higher modeling error in summertime: the highest NRMSE occurs in July, May, and June. With RF, NRMSE is lowest in the transitional and winter months (March, February, and May at ROTH; January, March, and February at TUCC). Measured ET at TUCC is likely more affected by water emissions from traffic and other anthropogenic sources and interception loss from impervious surfaces due to its location on a roof in the city center, which reduces modeling accuracy (Karsisto et al., 2016; Miralles et al., 2020). Although summertime EC measurements are usually subject to fewer measurement issues (Ward et al., 2016), our study indicates higher error in summertime based upon NRMSE and RMSE. However, pbias is closer to zero in summer months than in transitional and winter months.

RF models overestimate ET (positive pbias) in the cooler months (November, January-April), with overestimation particularly pronounced in the transitional months (April for both towers and November especially for TUCC). In summertime, pbias fluctuates around zero.

Models tend to overestimate ET more at TUCC (higher pbias). CNN models are more prone to underestimation than RF models, which can be partially attributed to prediction of negative ET values.
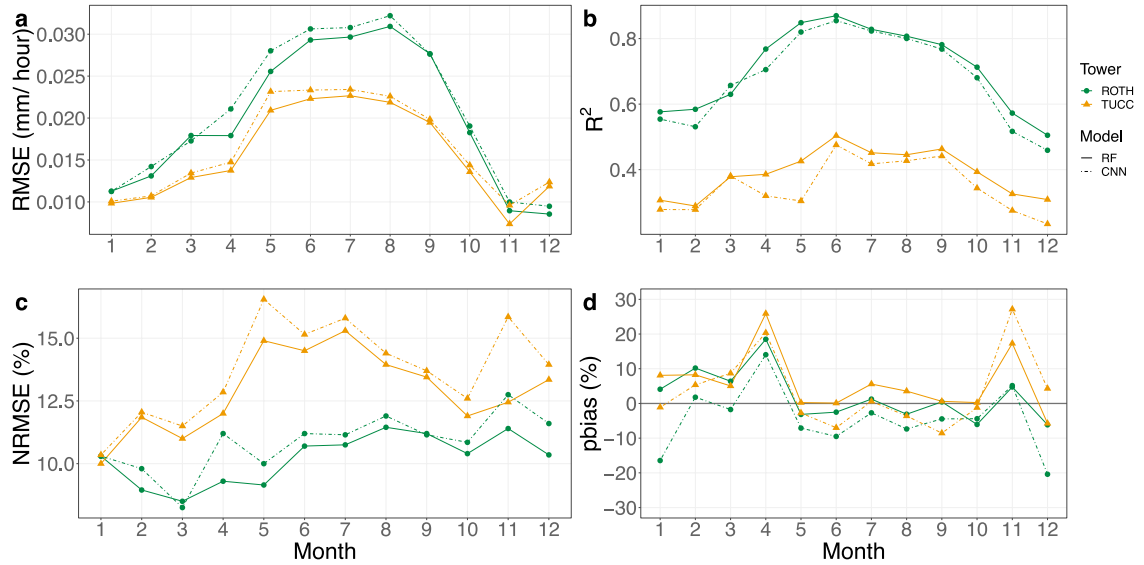

Figure B.6: Monthly averages of (a) root mean square error (RMSE), (b) coefficient of determination (R^2), (c) normalized root mean square error (NRMSE), and (d) percent bias (pbias) based on the testing set (averaged across the two training and testing splits, with training in 2018/ 2020 and testing in 2019 and vice versa). The predictor scenario depicted is "Met and GIS" (meteorological and GIS predictors with 22 predictors in total). Training and testing were conducted on the same tower. Only timestamps which were available for both towers in the testing datasets were used.

Model accuracy varies not only on an annual, but also on a diurnal, basis (Figures B.7-B.10). In the transitional seasons (spring and autumn) at ROTH, NRMSE is higher during the night, whereas NRMSE does not show clear diurnal variation in summer and winter. Higher modeling error at night may be due to inaccuracies in the nighttime storage heat flux (Ward et al., 2016, 2013). While pbias does not display a diurnal pattern in summertime, it varies considerably diurnally for all other seasons. RF models generally overestimate ET more at nighttime (positive pbias) in spring, autumn, and winter. In contrast, at ROTH, CNN models underestimate ET at nighttime in spring and winter (negative pbias). CNN models assign a higher importance to meteorological predictors, which imply a lower ET due to reduced solar radiation and temperature, whereas RF models assign a high importance to GIS predictors (Figure 3.4), which may explain the discrepancy in nighttime pbias between the AI algorithms.
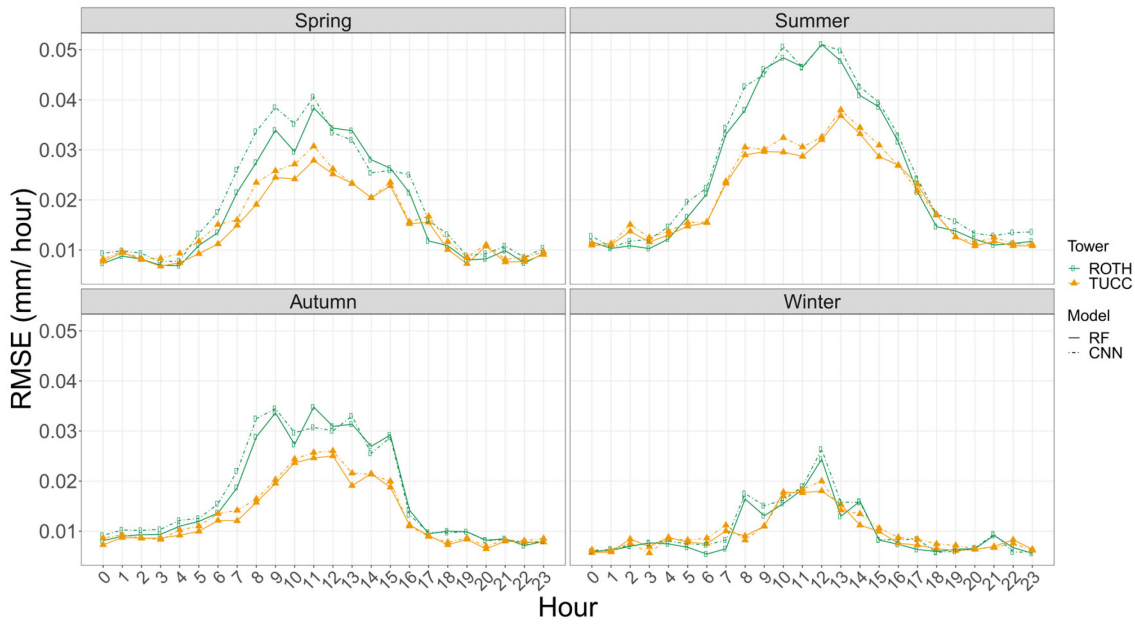
Figure B.7: Hourly averages of root mean square error (RMSE) for each season based on the testing set (averaged across the two training and testing splits, with training in 2018/ 2020 and testing in 2019 and vice versa). The predictor scenario depicted is "Met and GIS" (meteorological and GIS predictors with 22 predictors in total). Training and testing were conducted on the same tower. Only timestamps which were available for both towers in the testing datasets were used.
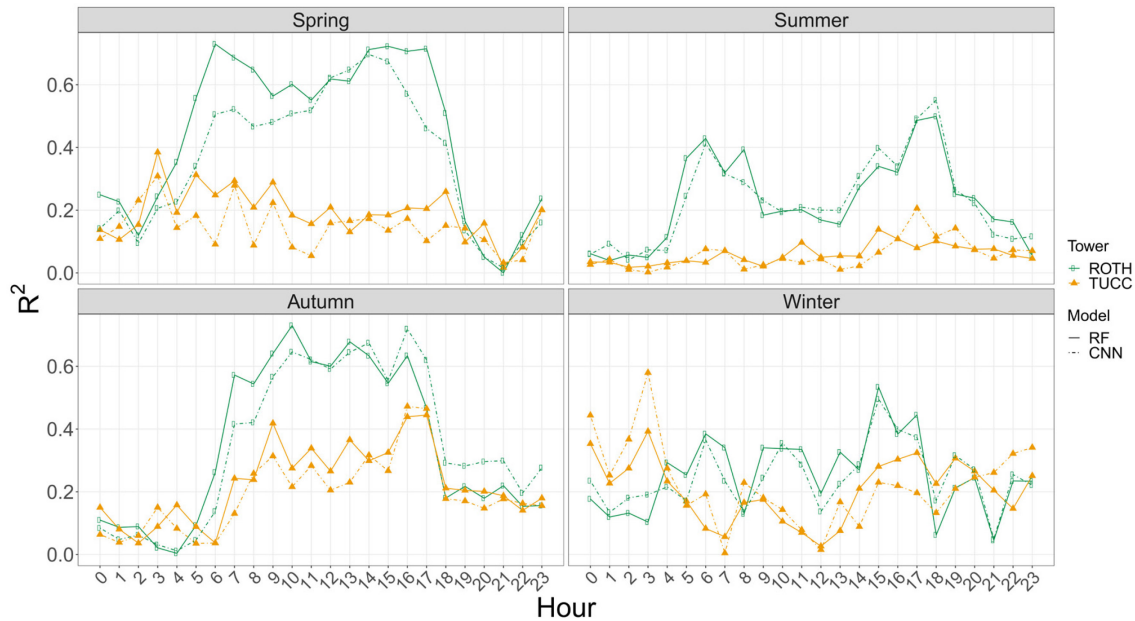

Figure B.8: Hourly averages of the coefficient of determination (R^2) for each season based on the testing set (averaged across the two training and testing splits, with training in 2018/ 2020 and testing in 2019 and vice versa). The predictor scenario depicted is "Met and GIS"

(meteorological and GIS predictors with 22 predictors in total). Training and testing were conducted on the same tower. Only timestamps which were available for both towers in the testing datasets were used.
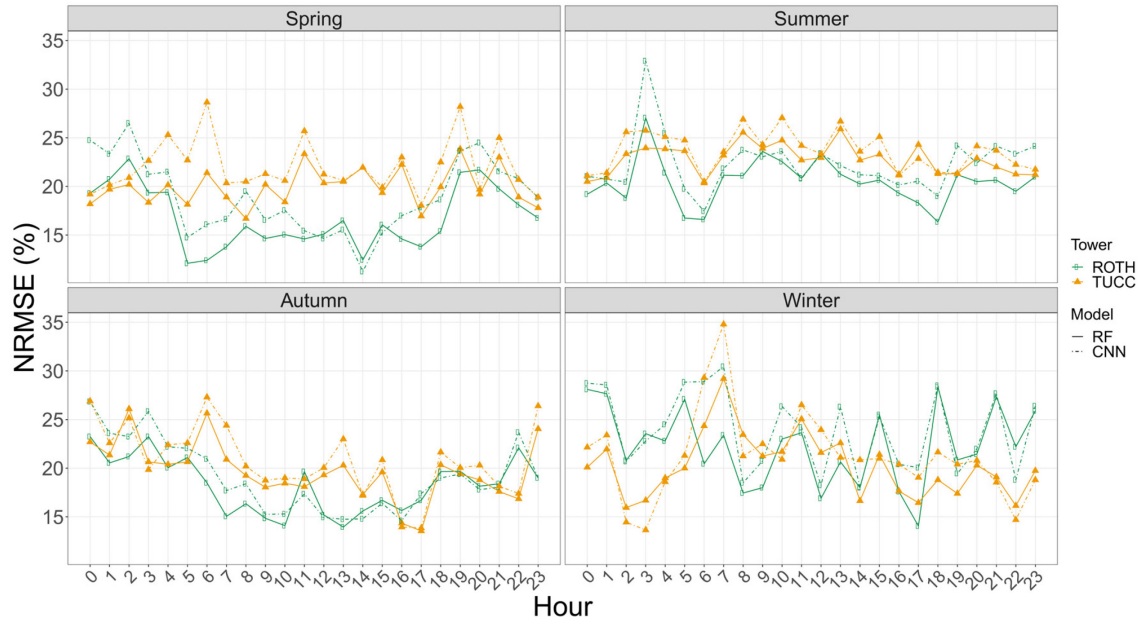

Figure B.9: Hourly averages of normalized root mean square error (NRMSE) for each season based on the testing set (averaged across the two training and testing splits, with training in 2018/ 2020 and testing in 2019 and vice versa). The predictor scenario depicted is "Met and GIS" (meteorological and GIS predictors with 22 predictors in total). Training and testing were conducted on the same tower. Only timestamps which were available for both towers in the testing datasets were used.
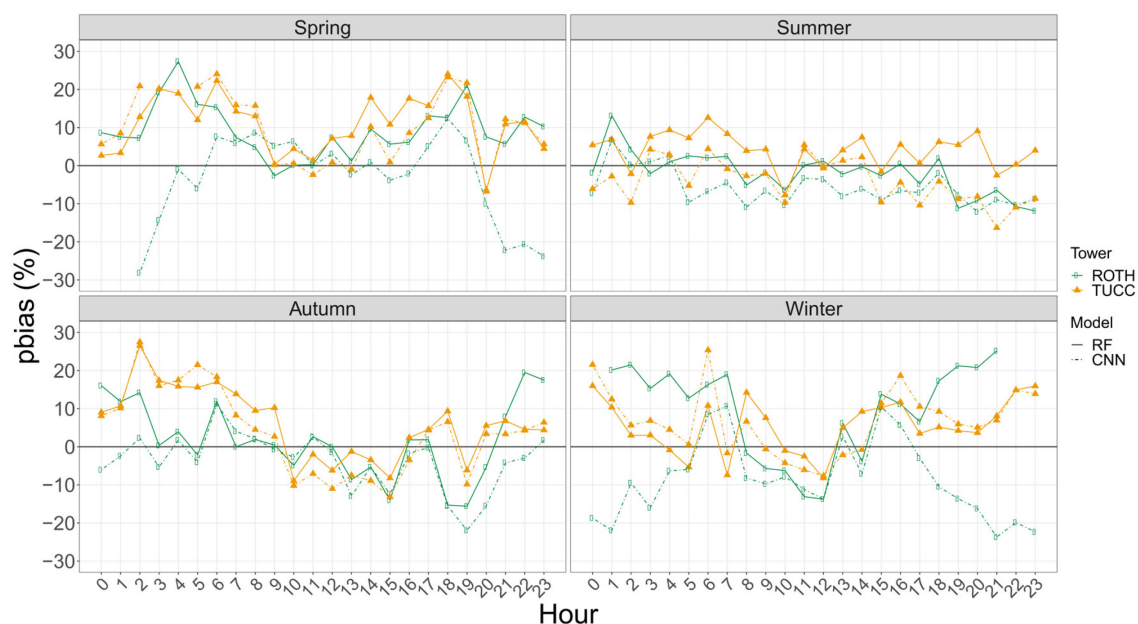
Figure B.10: Hourly averages of percent bias (pbias) for each season based on the testing set (averaged across the two training and testing splits, with training in 2018/ 2020 and testing in 2019 and vice versa). The predictor scenario depicted is "Met and GIS" (meteorological and GIS predictors with 22 predictors in total). Training and testing were conducted on the same tower. Only timestamps which were available for both towers in the testing datasets were used.

Appendix C   Mapping urban

ET

Supplementary material to Chapter 4: Mapping hourly urban evapotranspiration using Sentinel-2, open geodata and machine learning
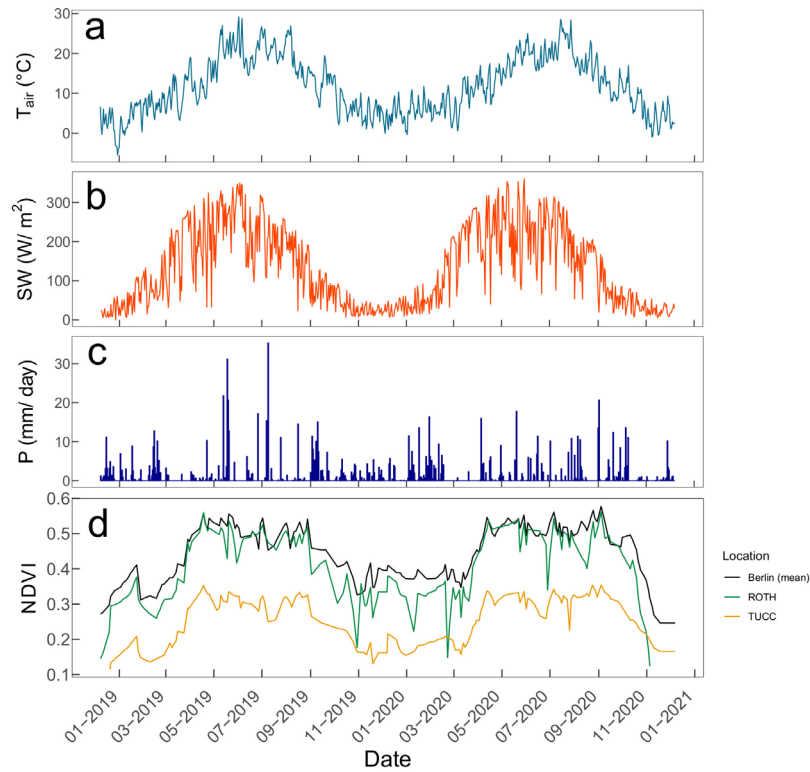

Figure C.1: Daily meteorological and vegetation greenness conditions during the study period (2019 - 2020): (a) averaged air temperature (Tair), (b) averaged shortwave downward radiation (SW), (c) precipitation (P) and (d) Normalized Difference Vegetation Index (NDVI). The NDVI time series was extracted from a Sentinel-2 L2A product (Gorelick et al., 2017) for Berlin and the respective flux tower locations at Rothenburgstrasse (ROTH) and TU Berlin Campus Charlottenburg (TUCC) using a 500-m buffer.
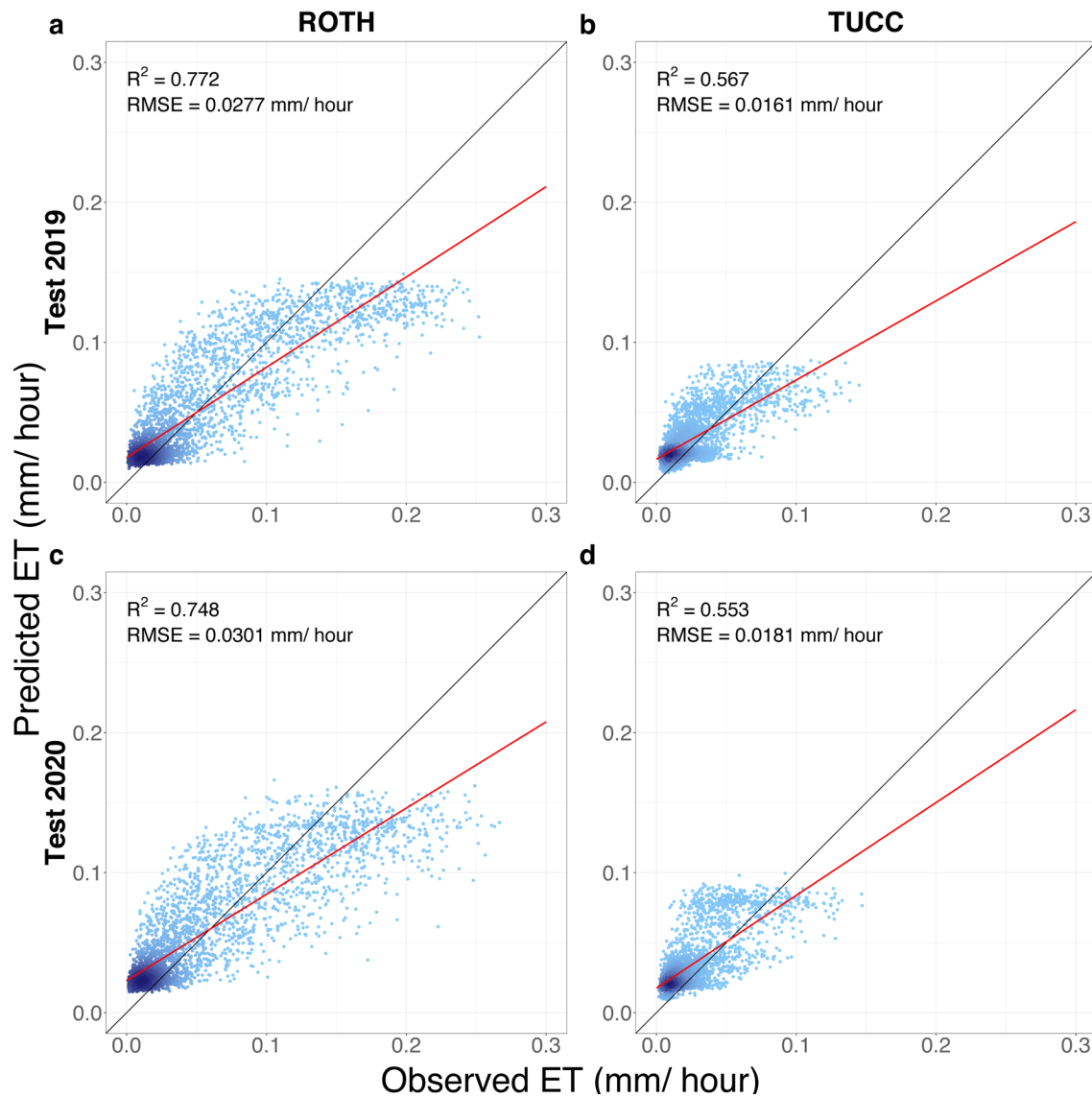
Figure C.2: Scatterplots showing the observed and predicted ET when testing at two flux towers (ROTH and TUCC) and two years (2019 and 2020). The intensity of the blue color represents the data point density. The red line indicates the best fit (linear model) and the solid black line indicates the 1:1 line.
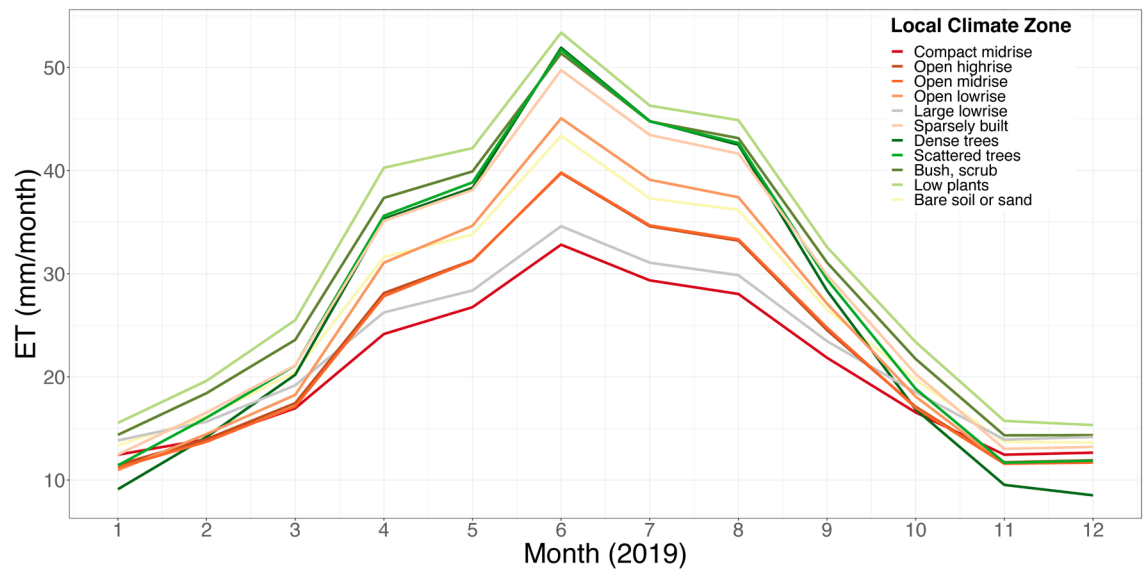
Figure C.3: Mean monthly ET (mm/ month) for Local Climate Zones over Berlin, Germany in 2019.

Appendix D   List of publications

Publications of cumulative doctoral thesis (in chronological order):

1. *Chapter 2 (as postprint):* **Vulova, S.**, Meier, F., Fenner, D., Nouri, H., Kleinschmit, B., 2020. Summer Nights in Berlin, Germany: Modeling Air Temperature Spatially With Remote Sensing, Crowdsourced Weather Data, and Machine Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 5074–5087. https://doi.org/10.1109/JSTARS.2020.3019696

2. *Chapter 3 (as postprint):* **Vulova, S.**, Meier, F., Rocha, A.D., Quanz, J., Nouri, H., Kleinschmit, B., 2021. Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence. Science of the Total Environment 786, 147293. https://doi.org/10.1016/j.scitotenv.2021.147293

3. *Chapter 4 (as preprint):* **Vulova, S.**, Rocha, A.D., Meier, F., Nouri, H., Schulz, C., Soulsby, C., Tetzlaff, D., Kleinschmit, B., 2022. Mapping hourly urban evapotranspiration using Sentinel-2, open geodata and machine learning. Under review: Remote Sensing of Environment.

Other publications:

Co-authored publications (in chronological order):

1. Rocha, A.D., **Vulova, S.**, van der Tol, C., Förster, M., Kleinschmit, B., 2022a. Modelling hourly evapotranspiration in urban environments with SCOPE using open remote sensing and meteorological data. Hydrology and Earth System Sciences 26, 1111–1129. https://doi.org/10.5194/hess-26-1111-2022

2. Rocha, A.D., **Vulova, S.**, Meier, F., Förster, M., Kleinschmit, B., 2022b. Mapping evapotranspirative and radiative cooling services in an urban environment. Sustainable Cities and Society 85, 104051. https://doi.org/10.1016/j.scs.2022.104051

Conference papers (in chronological order):

1. **Vulova, S.**, Kleinschmit, B., 2019. Thermal behavior and its seasonal and diurnal variability of urban green infrastructure in a mid-latitude city - Berlin, in: 2019 Joint Urban Remote Sensing Event (JURSE). Vannes, France, pp. 9–12. https://doi.org/10.1109/JURSE.2019.8809011

2. **Vulova, S.**, Kuhlemann, L.M., Tetzlaff, D., Soulsby, C., Kleinschmit, B., 2019. Assessment of evapotranspiration from urban vegetation across space and time: A case study in Berlin, in: 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). Shanghai, China, pp. 1–4. https://doi.org/10.1109/Multi-Temp.2019.8866903

3. **Vulova, S.**, Rocha, A.D., Meier, F., Nouri, H., Schulz, C., Kleinschmit, B., 2022. Modeling urban evapotranspiration with Sentinel-2, open geodata, and machine learning in summertime. International Geoscience and Remote Sensing Symposium (IGARSS) 2022. Kuala Lumpur, Malaysia. Accepted for publication.

Conference presentations (in chronological order):

1. *Thermal behavior and its seasonal and diurnal variability of urban green infrastructure in a mid-latitude city – Berlin* (2019). European Geosciences Union (EGU) General Assembly. Vienna, Austria. **Oral presentation.**

2. *Thermal behavior and its seasonal and diurnal variability of urban green infrastructure in a mid-latitude city – Berlin* (2019). Joint Urban Remote Sensing Event (JURSE). Vannes, France. **Oral presentation** and **poster**.

3. *Assessment of evapotranspiration from urban vegetation across space and time: a case study in Berlin* (2019). 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). Shanghai, China. **Oral presentation.**

4. *Modelling urban evapotranspiration with remote sensing and machine learning* (2020). First International Conference on Urban Water Interfaces. Online. **Oral presentation.**

5. *A data-driven approach to quantifying urban evapotranspiration using remote sensing, footprint modeling, and deep learning* (2021). European Geosciences Union (EGU) General Assembly. Online. **Oral presentation.**

6. *The city is breathing: Mapping urban evapotranspiration with remote sensing, artificial intelligence, and footprint modeling* (2022). Living Planet Symposium. Bonn, Germany. **Oral presentation.**

7. *Modeling urban evapotranspiration with Sentinel-2, open geodata, and machine learning in summertime* (2022). International Geoscience and Remote Sensing Symposium (IGARSS). Kuala Lumpur, Malaysia. **Oral presentation.**