UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

Mapping the time-varying spatial heterogeneity of temperature processes over the urban landscape of Augsburg, Germany

Matthias Wild, Svenia Behm, Christoph Beck, Josef Cyrys, Alexandra Schneider, Kathrin Wolf, Harry Haupt

Angaben zur Veröffentlichung / Publication details:

Wild, Matthias, Svenia Behm, Christoph Beck, Josef Cyrys, Alexandra Schneider, Kathrin Wolf, and Harry Haupt. 2022. "Mapping the time-varying spatial heterogeneity of temperature processes over the urban landscape of Augsburg, Germany." *Urban Climate* 43: 101160. https://doi.org/10.1016/j.uclim.2022.101160.

Nutzungsbedingungen / Terms of use:

CC BY-NC-ND 4.0

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
CC-BY-NC-ND 4.0: Creative Commons: Namensnennung - Nicht kommerziell - Keine Bearbeitung
Weitere Informationen finden Sie unter: / For more information see:
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de

Mapping the time-varying spatial heterogeneity of temperature processes over the urban landscape of Augsburg, Germany

Matthias Wild [a], Svenia Behm [a], Christoph Beck [b], Josef Cyrys [c], Alexandra Schneider [c], Kathrin Wolf [c], Harry Haupt [a,*]

[a] University of Passau, School of Business, Economics and Information Systems, Chair of Stantistics and Data Analytics, Innstraße 27, D-94030 Passau, Germany
[b] University of Augsburg, Institute of Geography, Alter Postweg 118, D-86159 Augsburg, Germany
[c] Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

ARTICLE INFO

Keywords:
Urban air temperature
Urban heat island
Spatial trend modelling
Seasonal and diurnal cycles
Nonparametric regression
Kriging

ABSTRACT

Generating high-resolution spatial interpolations of temperature processes is a vital task for studying urban climate anomalies and their various consequences. Such processes often constitute a complex and demanding data environment: Anthropogenic and natural conditions of the urban landscape result in anisotropic spatial dependencies and trend patterns that often vary in diurnal and seasonal cycles. Two-step geostatistical methods such as residual kriging take spatial heterogeneity into account but ignore the temporal dimension, which can result in a significant loss of potentially useful information. In this study, we propose nonparametric spatial detrending to obtain a process that fulfills the assumptions of ordinary kriging. In our application to urban air temperature series from monitoring sites distributed over the urban and suburban area of Augsburg, Germany, we provide an in-depth analysis of time-varying spatial heterogeneity. By using sub-sampling to account for diurnal, seasonal, and spatial trends, we produce interpolation maps with a resolution of 100 m × 100 m. The validation in a narrower sense is based on cross-validation and shows favorable behavior of the proposed method even when sub-samples are neglected. The broader sense validation is based on hold-out monitoring sites and provides further empirical support for our proposal.

1. Introduction

As more and more people live in urban regions or metropolitan areas, the assessment of urban air temperature anomalies becomes increasingly important. In particular, urban-rural differences, reflected in the phenomenon of so-called urban heat islands (Rizwan et al., 2008; Stewart, 2011; Mirzaei, 2015), have a strong impact on the biophysical environment in cities which may be associated with negative effects on the quality of life and health of urban residents, e.g. urban heat stress and thermal discomfort (McCarthy et al., 2010; Fischer et al., 2012; Van Hove et al., 2015; Martilli et al., 2020). Urban heat adaptation and mitigation strategies are widely

* Corresponding author.
E-mail addresses: matthias.wild@uni-passau.de (M. Wild), svenia.behm@uni-passau.de (S. Behm), christoph.beck@geo.uni-augsburg.de (C. Beck), cyrys@helmholtz-muenchen.de (J. Cyrys), alexandra.schneider@helmholtz-muenchen.de (A. Schneider), kathrin.wolf@helmholtz-muenchen.de (K. Wolf), harry.haupt@uni-passau.de (H. Haupt).

discussed and include cool roofs (Oleson et al., 2010; Gilabert et al., 2021), urban greening (Chun and Guldmann, 2014; Zhang et al., 2017; Saaroni et al., 2018; Ortiz Porangaba et al., 2021), urban transportation planning concepts (Nieuwenhuijsen, 2020; Mueller et al., 2020), and urban redevelopment (Yi et al., 2022). Related work addressing such strategies emphasises the environmental injustice that disadvantages low-income and underprivileged people when, for example, urban vegetation is spatially unevenly distributed (Pham et al., 2012; Zhang et al., 2017). Comprehensive analyses of the spatio-temporal distribution of air temperature – in particularly with regard to canopy layer urban heat islands (Stewart, 2011) – and accurate predictions on high-resolution spatial maps can help urban planners and decision makers develop equitable, sustainable, and effective adaptation strategies (Van Hove et al., 2015; Esau et al., 2019). In addition, air temperature predictions serve as tools and are used to improve prediction of other variables in urban ecosystems, climatology, economics, and especially epidemiology (e.g., Basu and Ostro, 2008; Kloog et al., 2012; Aubrecht and Özceylan, 2013; Beck et al., 2018; He et al., 2019).

However, modelling and prediction of urban temperature in space and time is not a straightforward task. The spatial dimension of urban temperature reveals anisotropic dependence structures caused by influences such as building development and transport infrastructure (e.g., Nieuwenhuijsen, 2020; Mueller et al., 2020), rendering direct application of simple interpolation techniques such as inverse distance weighting and ordinary kriging inappropriate. If neighboring locations do not have comparable environments, it is misleading to rely only on the distance between them to assess a relationship in contemporary air temperature measurements. As for the temporal dimension, urban temperature is usually characterised by diurnal and seasonal patterns (e.g., Fenner et al., 2014; Lazzarini et al., 2015; Van Hove et al., 2015; Zhang et al., 2017; Amorim et al., 2021; Núñez-Peiró et al., 2021), with urban heat island values being highest in summer and at night during calm and cloudless weather.

Even if the goal of modelling is to produce accurate interpolation maps with high spatial resolution, considering only the spatial dimension while neglecting the temporal dimension and its diurnal and seasonal patterns can result in a significant loss of potentially useful information and thus a loss of predictive power. On the other hand, the identification and modelling of a pure three-dimensional spatio-temporal dependence structure is very complex and usually relies on a large number of assumptions (De Iaco et al., 2002). As a compromise between these positions, the residual interpolation optimised (RIO) framework (Hooyberghs et al., 2006; Janssen et al., 2008) pursues an alternative idea. It basically proposes to filter spatial non-stationarity (in a first step) assuming stationarity in the time dimension. Thereby it enables (in a second step) residual ordinary kriging for spatial interpolation.

In this paper, we apply the RIO concept to urban canopy layer temperature processes. We use multiple nonparametric regression for spatial detrending (Behm et al., 2018) in the first step to avoid the effort of searching for trend specifications and the problems of possible misspecification. As a prerequisite, we propose a careful examination of assumptions about the temporal stability of spatial trends and variograms through an in-depth analysis of temperature trajectories. Our empirical results suggest that this works well and makes it possible to apply a simple procedure such as ordinary kriging in a second step. As a further improvement, we propose to divide the data into time windows by season and time of day before analysis. This will allow us to more accurately capture the distinctly different patterns of spatial heterogeneity in the diurnal and seasonal cycles of urban temperature series for the two steps of our procedure.

The remainder of this study is organised as follows: Section 2 formulates the methodological framework. Section 3 introduces the Augsburg temperature data, the employed predictors, and exploratory tools to detect temporal and spatial patterns of the temperature series. Section 4 evaluates the proposed methodology on basis of the dataset while Section 5 discusses and Section 6 concludes.

2. Methods

For study area D_s and time period D_t consider a geostatistical process (GSP)

$$\{Z(\boldsymbol{s};t) : \boldsymbol{s} \in D_s, t \in D_t\}, \tag{1}$$

that was observed at N locations $\boldsymbol{s}_n = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \in D_s$ and T time stamps, $t \in \{1, \ldots, T\} \in D_t$, producing observations $z_{n,t}$. If, at time $t \in D_t$ and over all locations $\boldsymbol{s}, \boldsymbol{s}' \in D_s$,

$$\begin{aligned} &\text{(a)} \quad \mathbb{E}[Z(\boldsymbol{s};t)] = \mu, \\ &\text{(b)} \quad \mathrm{Var}[Z(\boldsymbol{s};t)] = \sigma^2, \\ &\text{(c)} \quad \mathrm{Cov}[Z(\boldsymbol{s};t), Z(\boldsymbol{s}';t)] = \sigma^2\rho(h), \end{aligned}$$

hold, where $\rho(h) = Corr[Z_t(\boldsymbol{s}), Z_t(\boldsymbol{s}')]$ depends only on distance $h = ||\boldsymbol{s} - \boldsymbol{s}'||$, the process $Z(\boldsymbol{s};t)$ is second-order stationary and isotropic.

In theory, if the assumptions (a)–(c) of the GSP in (1) are approximately true, a well-established theory can be used for statistical inference (e.g. Cressie, 1993; Chilés and Delfiner, 2012). In empirical practice, when the interest is in spatial interpolations of spatio-temporal processes in a given time period, one faces the problem that the underlying processes (by assumption time-invariant) exhibit spatial heterogeneity in mean $\mu(\boldsymbol{s})$ and standard deviation $\sigma(\boldsymbol{s})$, thus violating assumptions (a)–(c). For a more comprehensive overview of GSP assumptions, related methods and applied examples, we refer to Zimmerman et al. (2010), Kauermann et al. (2012) or Montero et al. (2015).

As a potential remedy, the residual interpolation optimised (RIO) technique suggests the use of a two-step procedure for spatial interpolation, where the initial spatial detrending step aims to filter the process such that assumptions (a)–(c) hold approximately (e.g., Janssen et al., 2008; Behm et al., 2018). In the following, we describe the two steps of our implementation, which we will refer to as NP-OK: A nonparametric approach (NP) is used for the spatial detrending step, while the interpolation is based on ordinary kriging (OK):

STEP 1 (spatial detrending). The spatial trends of mean and standard deviation of $Z(\boldsymbol{s};t)$ are assumed to be time-invariant functions of predictors $\boldsymbol{X}$, i.e., $\mu(\boldsymbol{s};t) \approx \mu(\boldsymbol{s}) \approx m_\mu(\boldsymbol{X_s})$ and $\sigma(\boldsymbol{s};t) \approx \sigma(\boldsymbol{s}) \approx m_\sigma(\boldsymbol{X_s})$. Using empirical estimates for mean and standard deviation on basis of time series $\boldsymbol{z}_n$ allows fitting of $\widehat{\mu}(s) = \widehat{m}_\mu(X_s)$ and $\widehat{\sigma}(s) = \widehat{m}_\sigma(X_s)$. Employing nonparametric trend modelling to estimate $m_\mu(\cdot)$ and $m_\sigma(\cdot)$ overcomes inconveniences associated with parametric trend modelling, i.e. ad hoc assumptions on the degree of the polynomials, potential misspecification bias, or the extensive computational costs of specification search (Behm et al., 2018). In the present study, the mean (and, analogously, the standard deviation) is approximated by using a local-linear kernel smoothing estimator of $E(\overline{Z}|\boldsymbol{X}) = m_\mu(\boldsymbol{X})$ in the trend regression model $\overline{Z} = m_\mu(\boldsymbol{X}) + U$ with $E(U|\boldsymbol{X}) = 0$. We denote the weighted least squares estimator as $\widehat{m}_{\mu,\mathrm{NP}}$, where $(\widehat{m}, \widehat{\gamma})$ minimises $\sum_{n=1}^{N} [\overline{Z}_n - m - \gamma(\boldsymbol{X})]^2 K(\boldsymbol{X},\boldsymbol{h})$, with product kernel K and vector of bandwidths $\boldsymbol{h}$ estimated by expected Kullback-Leibler cross-validation (Hurvich et al., 1998).

Then, given arbitrarily predefined reference levels μ^{ref} and σ^{ref}, detrending of the observed data $\boldsymbol{z}_n$ is carried out through

$$z_n^* = z_n + (\mu^{\mathrm{ref}} - \widehat{\mu}(\boldsymbol{s}_n)) \tag{2a}$$

$$z_n^{**} = (z_n^* - \overline{z}_n^*)\frac{\sigma^{\mathrm{ref}}}{\widehat{\sigma}(\boldsymbol{s}_n)} + \overline{z}_n^*, \tag{2b}$$

with average over semi-detrended data $\overline{z}_n^*$, where the vector $\boldsymbol{z}_n = (z_{n,1}, \ldots, z_{n,T})$ defines the time series observed at monitoring site $\boldsymbol{s}_n$. If the assumption of time-invariant spatial trending holds, z_n^{**} can be interpreted as a time series vector of realisations observed from residual random field $Z^{**}(\boldsymbol{s}_n;t)$ at location $\boldsymbol{s}_n$, for which second-order stationarity conditions analogous to assumptions (a)–(c) hold for all $t \in D_t$.

STEP 2 (residual interpolation). Under such circumstances the ordinary kriging interpolator enjoys well known optimality properties (Cressie, 1993). As a prerequisite for kriging, we need to estimate and quantify the spatial dependence structures underlying the detrended data in the form of a variogram. In doing so, we assume that all historical detrended measurements $z_{n,t}^{**}$ originate from the same spatial random process and employ a pooled variogram estimation, i.e., we estimate the empirical variogram not only on the basis of data from one cross-section, but consolidate all observations available over time (Gräler et al., 2012). We consider several theoretical variogram models for fitting to the empirical estimates, the Linear, Spherical, Matern, Power, and Exponential models (see e.g. Armstrong, 1998; Chilés and Delfiner, 2012), each with a nugget effect. Finally, we choose the model that provides best weighted least squares fit to the empirical variogram estimate, see Calder and Cressie (2009) for more technical background. Using this theoretical variogram in ordinary kriging, the detrended observed data are interpolated to obtain a prediction $\widehat{Z}^{**}(s_0;t)$ for unobserved location of interest $\boldsymbol{s}_0$ at time t. Retrending $\widehat{Z}^{**}(s_0;t)$ provides prediction $\widehat{Z}(s_0;t)$ for the original unobserved process and is carried out via

$$\widehat{Z}^{*}(s_0;t) = (\widehat{Z}^{**}(s_0;t) - \overline{\widehat{Z}}^{**}(\boldsymbol{s}_0))\frac{\widehat{\sigma}(\boldsymbol{s}_0)}{\sigma^{\mathrm{ref}}} + \overline{\widehat{Z}}^{**}(\boldsymbol{s}_0) \tag{3a}$$

$$\widehat{Z}(s_0;t) = \widehat{Z}^{*}(s_0;t) - (\mu^{\mathrm{ref}} - \widehat{\mu}(\boldsymbol{s}_0)), \tag{3b}$$

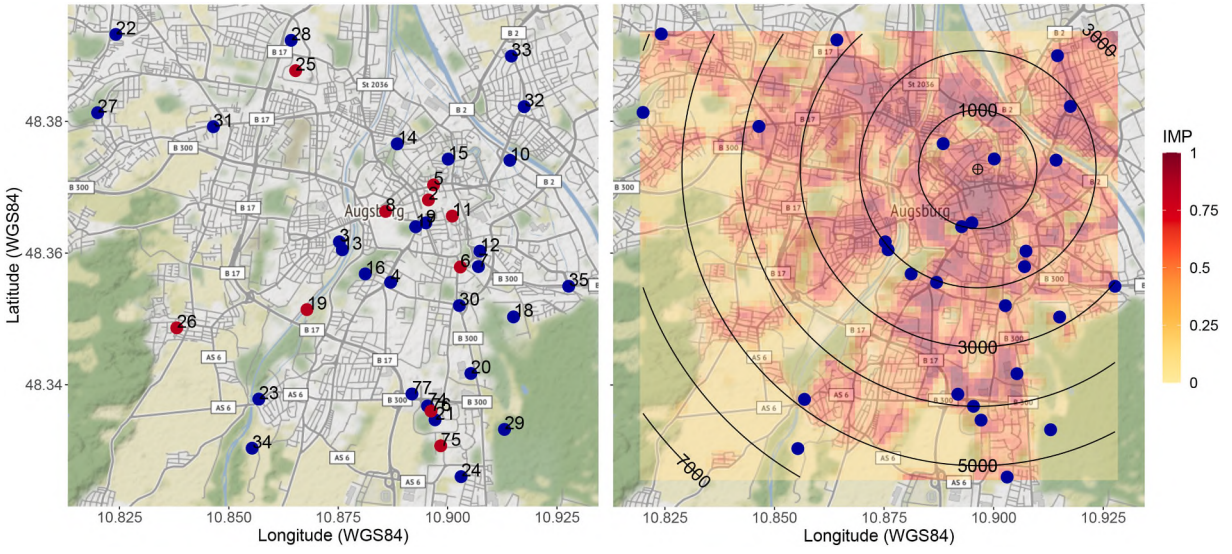

Fig. 1. Locations of monitoring sites in the urban and suburban area of Augsburg, Germany (left) and spatial distribution of predictors IMP and DTC (right). Sites considered in further analysis of diurnal, seasonal, and spatial patterns of heterogeneity are labelled with blue dots, hold-out sites with red dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with time average over interpolated values $\overline{\widehat{Z}}^{**}(s_0)$ and estimates for spatial mean and standard deviation at unobserved location s_0, i.e. $\widehat{\mu}(s_0) = \widehat{m}_\mu(X_{s_0})$ and $\widehat{\sigma}(s_0) = \widehat{m}_\sigma(X_{s_0})$, given by the formerly fitted trend functions.

3. Data

The raw data set considered for the analysis contains temperature measurements with a resolution of 4 minutes over the period from 2013 to 2019 at 82 sites in the city of Augsburg and the two adjacent counties in Bavaria, Germany. The entire measurement network covers an area of approximately 45×45 km^2. The corresponding quality assurance is discussed in detail by Beck et al. (2018). Aggregating the raw data into hourly averages (when at least 75% of the 4-minute measurements are available) and restricting the data to 38 of 82 monitoring stations (Augsburg city and suburbs, covering approximately an area of 9×10 km^2) for the years 2016 and 2017 (January to December in each case), we achieve reasonable data coverage (≈0.8% missing). The remaining years and stations are discarded due to long patterns of consecutive missings. This determination leads to a total of 661,359 hourly observations for further analysis, covering the full range of weather conditions, times of year, and times of day. For reasons of uniformity and to avoid confusion with German summer time, all time data in the following study are given throughout in German winter time, i.e., UTC+1.

To complement the results collected during the study, we made a further split by randomly assigning $H = 10$ of the 38 measurement series to a hold-out set. These series are not included in the following analysis of diurnal, seasonal, and spatial patterns of heterogeneity, but are withheld until the end of the paper, where we use these locations for a robustness check and a final validation of the predictive performance of the models considered. Fig. 1 (left) shows the spatial distribution of the selected 38 monitoring sites in the urban and suburban area of Augsburg, with red dots indicating the randomly assigned hold-out status.

To identify additional predictors of trends in spatial mean and standard deviation, we follow Coseo and Larsen (2014), Chun and Guldmann (2014) and Przybylak et al. (2015), who identified the degree of soil sealing (imperviousness) as a major factor in urban heat island effects. We define the predictor IMP as the percentage of sealed soil surrounding a site, averaged and within a 100-meter radius buffer area using 2015 data ([dataset] EEA, 2016) provided by the Copernicus Land Monitoring Service. Including all types of man-made development, i.e., residential areas, transportation areas (roads, airports, ports, train stations, parking lots), industrial and commercial areas, factories, power generation and distribution facilities, the predictor IMP is associated with many potential sources of waste heat and serves as indicator of materials with characteristic properties that increase heat storage capacity, thus provides a reasonable proxy for the implicit concept of small-scale urbanisation.

To complement IMP with large-scale information on urbanisation, we consider the additional predictor distance to the city center DTC (in meters), which we define heuristically as the geographic distance to the WGS84 coordinates (48.372721, 10.896377) reflecting the approximate center of the city core. The idea of the predictor DTC follows the studies of Arnds et al. (2015) and Kotharkar et al. (2019), who observe a decreasing gradient of air temperature from the city center toward more rural areas, Yang et al. (2020), who observed a moderating effect of overall city size on the relationship between local development and urban heat island effects, and Núñez-Peiró et al. (2021), who concluded that information on distance to the city center is highly collinear with predictors of local climate zones and has high explanatory power for temperature trends, especially at night. For a more flexible yet less straightforward approach, we refer the interested reader to the work of Straub et al. (2019) who define multiple (district) centers based on land use data.

Fig. 1 (right) shows the distribution of the two spatial predictors. The values of the predictor IMP assigned to the available measurement sites range from 0% to 88.73%. DTC shows a maximum of 5814 meters from the city center, while the most central measurement point has a distance of 328 meters. For our sample, the correlation between IMP and DTC does not exceed the magnitude of −0.26. An evaluation of additional predictors for local climate zones and land cover categories is included in Appendix A.

To provide a first insight into the nature of the observed temperature process, Fig. 2 displays boxplots for each of the $N = 28$ in-

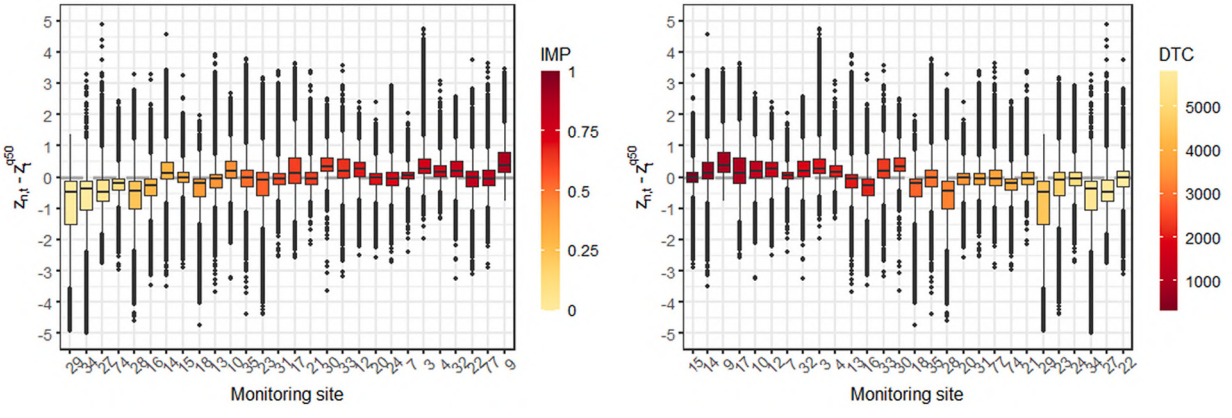

Fig. 2. Boxplots calculated over the differences $(z_{n,t} - z_t^{q50})$, $\forall t \in \{1, \ldots, T\}$ for each monitoring site; limits of y-axis restricted to [−5,5]. Order corresponding to increasing IMP (left) and DTC (right).

sample monitoring sites, calculated over the differences $(z_{n,t} - z_t^{q50})$, $\forall t \in \{1, \ldots, T\}$. The values resulting from the difference between each measurement at site n and time t, $z_{n,t}$, and the median of all measurements at time t, z_t^{q50}, can be interpreted as a local temperature anomaly. Thus, if a boxplot is centered around the dashed horizontal zero line, the corresponding measurement location coincides with an artificial median reference site. If a boxplot is below (above) the zero line, the corresponding measurement location tends to have lower (higher) temperature values than an artificial contemporary median reference site. A systematic deviation of boxplots from the zero line indicates the presence of spatial heterogeneity in the data. A possible explanation for systematic deviation is provided by the predictors IMP (DTC), as an increase in predictor values tends to be associated with a more positive (negative) deviation from the artificial median reference site.

A more detailed picture of urban temperature patterns is obtained by examining the diurnal cycle and the corresponding patterns in the annual cycle. This consideration is important for the assumption of temporally invariant spatial trends that underlies the NP-OK paradigm. In view of the complex cyclic patterns underlying the observed temperature series, we propose to construct twelve separate time series for each of the $N = 28$ in-sample sites: The four seasons, spring (March, April, May), summer (June, July, August), autumn (September, October, November), and winter (December, January, February), and three diurnal time windows that contain radiation periods with fairly distinct patterns of air temperature heterogeneity (see, e.g., Beck et al., 2018, Fig. 8): sunrise (defined as the three hourly observations spanning the average time of sunrise in that season), noon (the three hourly observations for 11 a.m. to 2 p.m.), and sunset (defined as the three hourly observations spanning the average time of sunset in that season). The selected time windows are also of particular interest because temperature readings are generally lowest around sunrise and the morning hours are a good indicator of whether nighttime cooling has occurred, especially during the summer months. During the midday hours, people generally go outside and are thus exposed to outdoor temperatures. At sunset, people also like to stay outside to shop, go home, or exercise.

Fig. 3 shows for the resulting twelve sub-samples the empirical density curve (dark red curve) of the temperature observations over all $N = 28$ in-sample monitoring sites with superimposed mean (orange line). For comparison purposes, each plot is supplemented by the overall empirical density curve (dashed gray curve). The distributions differ significantly over the sub-samples, and consequently the uncertainty can be reduced by looking at sub-samples instead of the whole data. A more detailed picture is provided by Fig. 4 and Fig. 5, which show rolling window estimates of $\mu(\boldsymbol{s}_n)$ and $\sigma(\boldsymbol{s}_n)$ for each observed time series $\boldsymbol{z}_n$ sub-sampled by the twelve selected time windows. Consequently, the rolling window size of $K = 42$ observations spans a total of 3 hours over 14 days. The assumption of time-invariant spatial heterogeneity requires that mean and standard deviation are reasonably constant over time. Visual inspection of the rolling mean estimates suggests stability of the means over time, at least for the summer and winter months. Spring shows a slight systematic increase in mean temperature, while the autumn sub-samples show the opposite trend. Although the mean behavior for

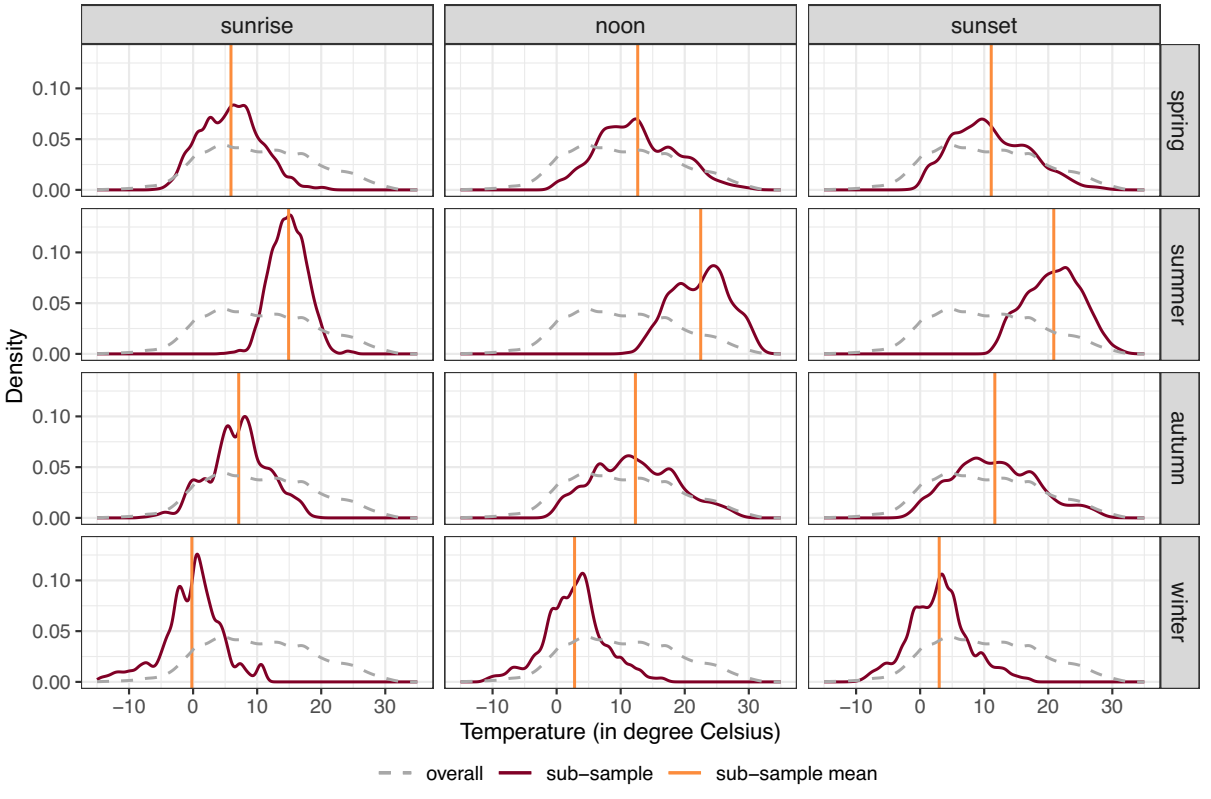

Fig. 3. Temperature density estimates for twelve sub-samples with superimposed mean and superimposed density estimate for all data.

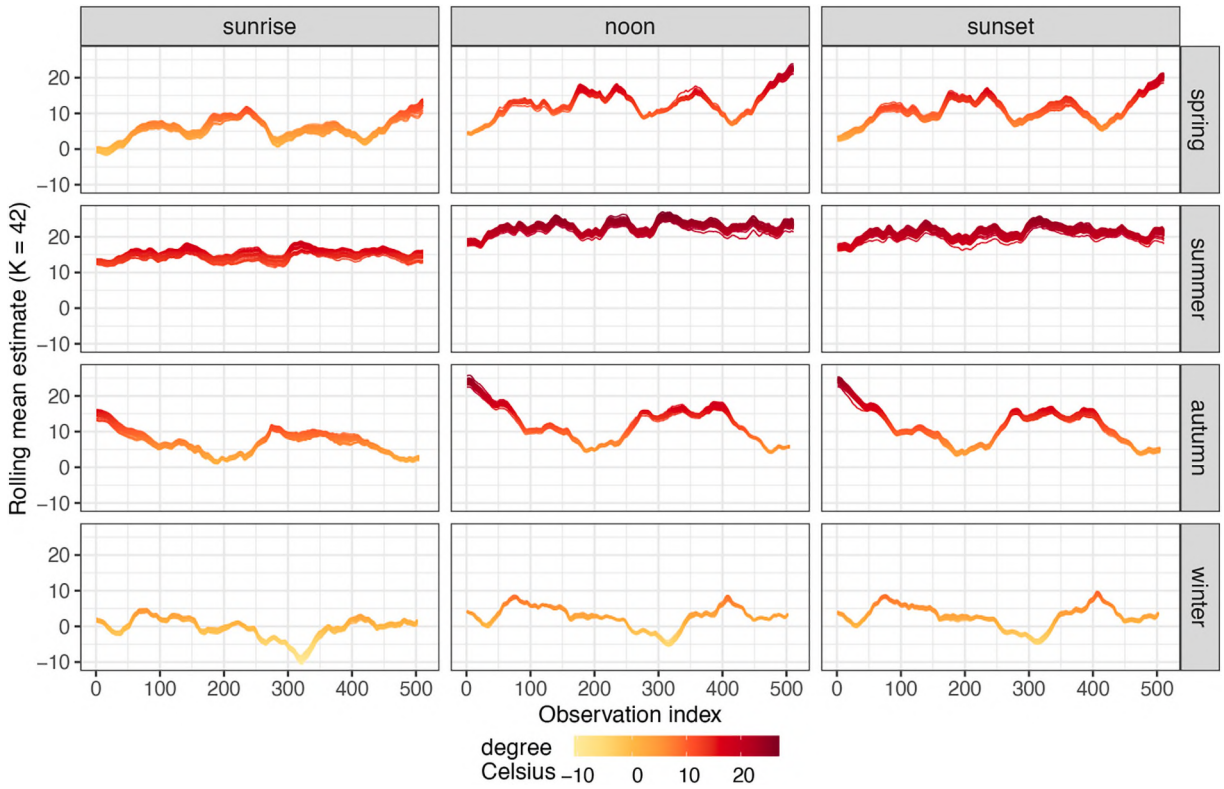

Fig. 4. Mean estimates using rolling windows of size $K = 42$, for each time series of observations $\mathbf{z}_n$, for twelve sub-samples.

these two seasons is not really stable, a finer split on the time axis does not suggest any improvement. The rolling window estimates of standard deviations show more erratic behavior, so the sub-sample construction does not appear to have a stabilising effect for this characteristic of the series.

By combining both dimensions, the twelve plots in Fig. 6 provide insight into the systematic nature of temporal and spatial heterogeneity of the underlying deviations from the median reference site. For each site and time window, the size (color) of the dots indicates the absolute value (sign) of the average of the deviations between the measurements and the values of the artificial median series. The plots show stable average deviation patterns for the sunrise hours across all seasons, while the summer noon and sunset hours show a higher degree of spatial variation compared to, for example, the winter noon hours. As expected, especially for sunrise hours, measurement sites in or near the city center tend to report higher temperature values than measurement sites in suburban areas.

4. Results

4.1. NP-OK on sub-samples of the data

We begin our evaluation of the NP-OK method by applying it to the twelve sub-samples. Fig. 7 shows the nonparametric trend function estimates for the spatial mean, based on the in-sample locations. Note that the fitted lines are drawn for three fixed values of DTC, namely the lower quartile (dark red), the median (orange), and the upper quartile (yellow). The flexible nature of nonparametric trend models includes any parametric linear or nonlinear functions as a special case. The estimated trend functions show the best cross-validated performance for large bandwidths and reveal approximately linear relationships between predictors and spatial mean for all twelve sub-samples. Predictor IMP provides a reasonable explanation of spatial mean, especially for summer months, when the effect of heat storage by impervious surfaces is largest. Although comparable in direction, the size of the effect is significantly minor for other seasons, especially for winter months. Predictor DTC, however, reveals no clear-cut effect for season summer and generally for the time around noon. On the contrary, the lower the impact of solar radiation, the higher the effect of predictor DTC: For the hours around sunrise and sunset, we observe systematically higher mean temperature at monitoring sites closer to city centre.

Fig. 8 shows an analogous plot of the nonparametric trend function estimates for the spatial standard deviation, based on the in-sample locations. As for the predictor DTC, the most notable effects are level shifts, which lead to higher deviations especially in more rural areas except for summer sunrise and autumn sunrise. The impact of IMP is ambiguous, as the estimated trend functions show a slightly positive dependence for spring and autumn, while revealing a negative impact on the standard deviation for the winter months.

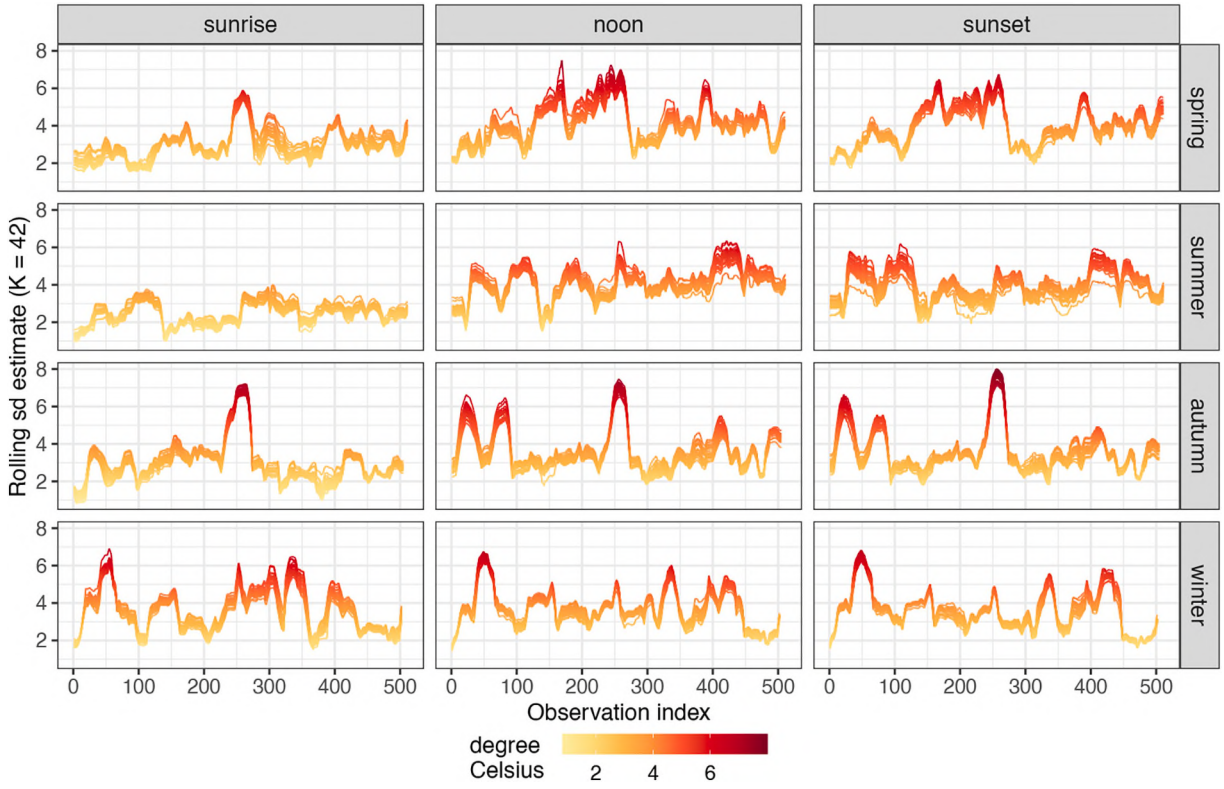

Fig. 5. Standard deviation estimates using rolling windows of size $K = 42$, for each time series of observations $\mathbf{z}_n$, for twelve sub-samples.

With respect to the spatial interpolation step, Fig. 9 shows empirical variogram estimates and variogram model fits based on the detrended data. Visual inspection reveals notable differences in the residual spatial dependence structure of the sub-samples. The estimates for summer noon and autumn noon yield nearly flat lines, indicating that the spatial correlation does not decrease further beyond a distance of a few hundred meters. For these sub-samples, the trends in the predictors IMP and DTC have already filtered out most of the systematic process variation, so the residual interpolation for these time windows is merely more than a simple average over all available contemporary observations. Other time windows, such as winter sunrise, show residual spatial dependence that can be used for more complex residual interpolation through kriging.

4.2. NP-OK on pooled data

In a next step, we apply the NP-OK method to the pooled data of all twelve sub-samples considered in the previous section. Comparing the results of such a pooled approach with those based on a sub-sample should provide insight into the usefulness of the sub-sample in meeting assumptions (a)–(c).

Fig. 10 depicts the estimated trend functions (top row) and variogram estimate (bottom). Although the estimates of $\widehat{m}_{\mu}(\mathsf{IMP}, \mathsf{DTC})$ and $\widehat{m}_{\sigma}(\mathsf{IMP}, \mathsf{DTC})$ do neglect the individual characteristics observed in Figs. 7 and 8, using a pooled approach still provides a clear-cut effect, especially for trending in spatial mean. The fitted variogram provides an estimate for the average spatial dependence structure of the individual random fields where the increasing availability of training data seems to stabilise the estimates. However, the pooled approach neglects temporal variability in spatial trends and therefore does not fulfill assumptions (a)–(c).

4.3. Prediction maps

Fig. 11 depicts prediction maps based on sub-sampling (denoted as NP-OK(I)), the pooled sub-sample data (denoted as NP-OK(II)), and inverse distance weighting (IDW) for two exemplary time instants. The latter serves as a basic, assumption-free benchmark method (see, e.g., DeGaetano and Belcher, 2007; Ninyerola et al., 2007; Wang et al., 2017 for more details on the IDW routine). For the 1st of August 2016, 12 p.m., the maps of NP-OK(I) and NP-OK(II) reveal a remarkably similar structure to the map of the predictor distribution (compare right plot of Fig. 1) which suggests that the mapping is highly driven by predictor IMP for both variants. Considering the maps referring to 1st of February 2016, 5 p.m., we observe a more prevalent gradient from centre to urban areas with predictor DTC interfering the effect of IMP. In contrast, IDW provides an over-smoothed mapping, discarding effects of potential spatial heterogeneity for summer noon hours as well as for winter sunset. Observations of high (low) temperature simply result in comparatively

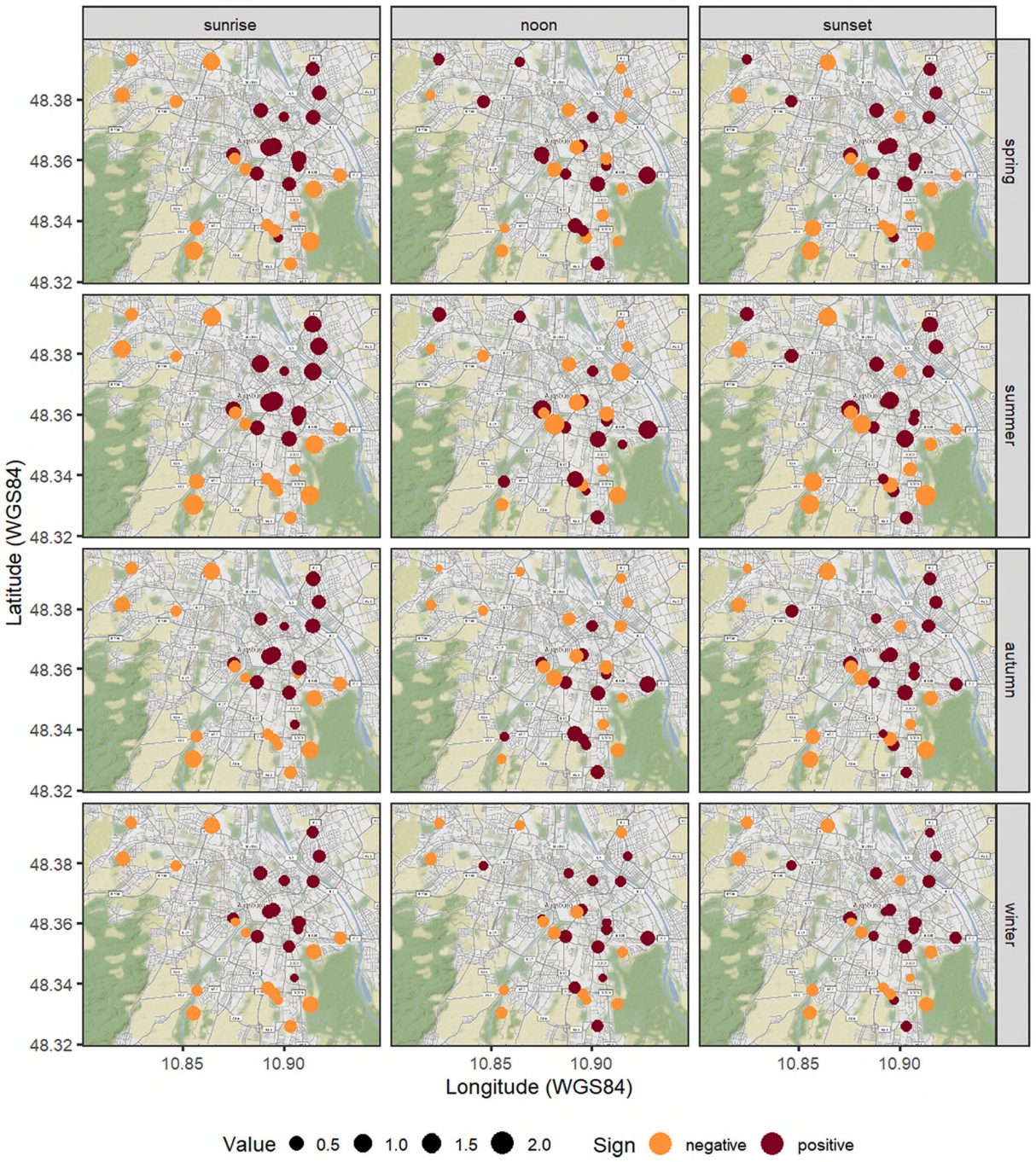

Fig. 6. Spatial distribution of each site's average differences to the artificial median series for twelve sub-samples: Red (orange) indicates positive (negative) average deviation from artificial median while dot size indicates strength of deviation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

high (low) prediction for the surrounding areas.

4.4. Model performance comparison

For evaluating the predictive performance of our method, we carry out a validation in the narrow sense by performing a leave-one-out cross-validation (LOOCV) on the in-sample sites. In detail, in each loop $n \in \{1, \ldots, N = 28\}$ of LOOCV, one time series of measurements $\boldsymbol{z}_n$, i.e. one in-sample monitoring site, is omitted. Then, NP-OK is applied to the remaining sites in order to predict $\widehat{Z}(s_n; t)$ for

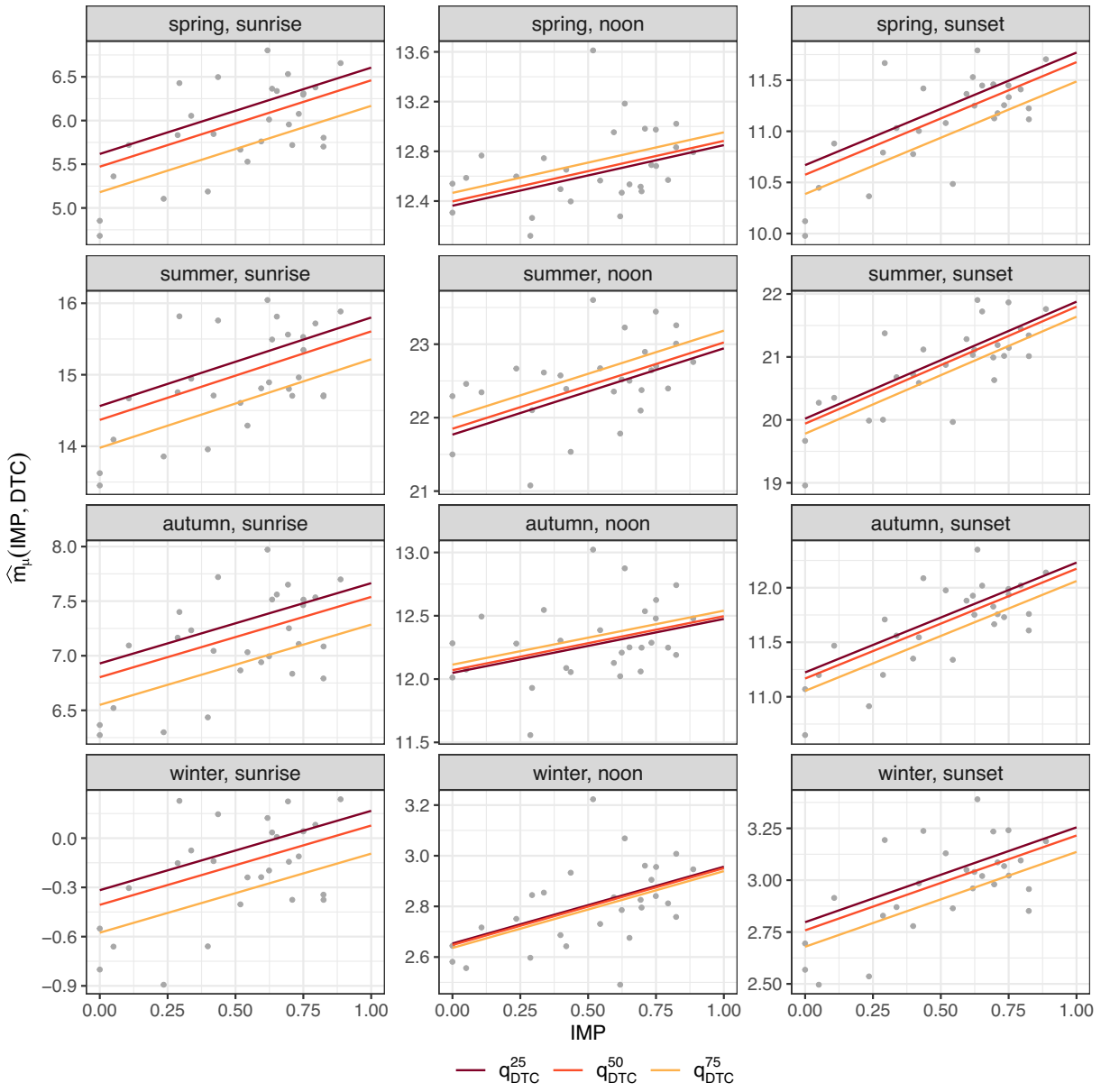

Fig. 7. Estimated nonparametric trend functions for spatial mean in dependence of predictor IMP and for different quartiles of DTC for twelve sub-samples.

each $t \in \{1, \ldots, T\}$ and to derive the corresponding out-of-sample prediction errors $\widehat{e}_{n,t} = z_{n,t} - \widehat{Z}(s_n; t)$. After $N = 28$ loops, two performance criteria are calculated:

$$\text{RMSE}_t = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\widehat{e}^2_{n,t}}, \qquad \text{BIAS}_n = \frac{1}{T}\sum_{t=1}^{T}\widehat{e}_{n,t}.$$

RMSE_t allows to interpret the general model performance in terms of undirected deviation from observed values while BIAS_n gives insight in potential systematic under- or overestimation due to e.g. omission of trend modelling, misspecification of the considered trend models and/or predictor variables. For a comprehensive overview over methods for evaluating the performance of environmental methods, we refer the interested reader to the position paper of Bennett et al. (2012). We evaluate NP-OK with and without subsampling into the twelve time windows as described in the preceding section, and complement the results by evaluating IDW. Further results on ordinary kirging without prior modelling of spatial trend or pooled semi-variance estimation can be found in Appendix B.

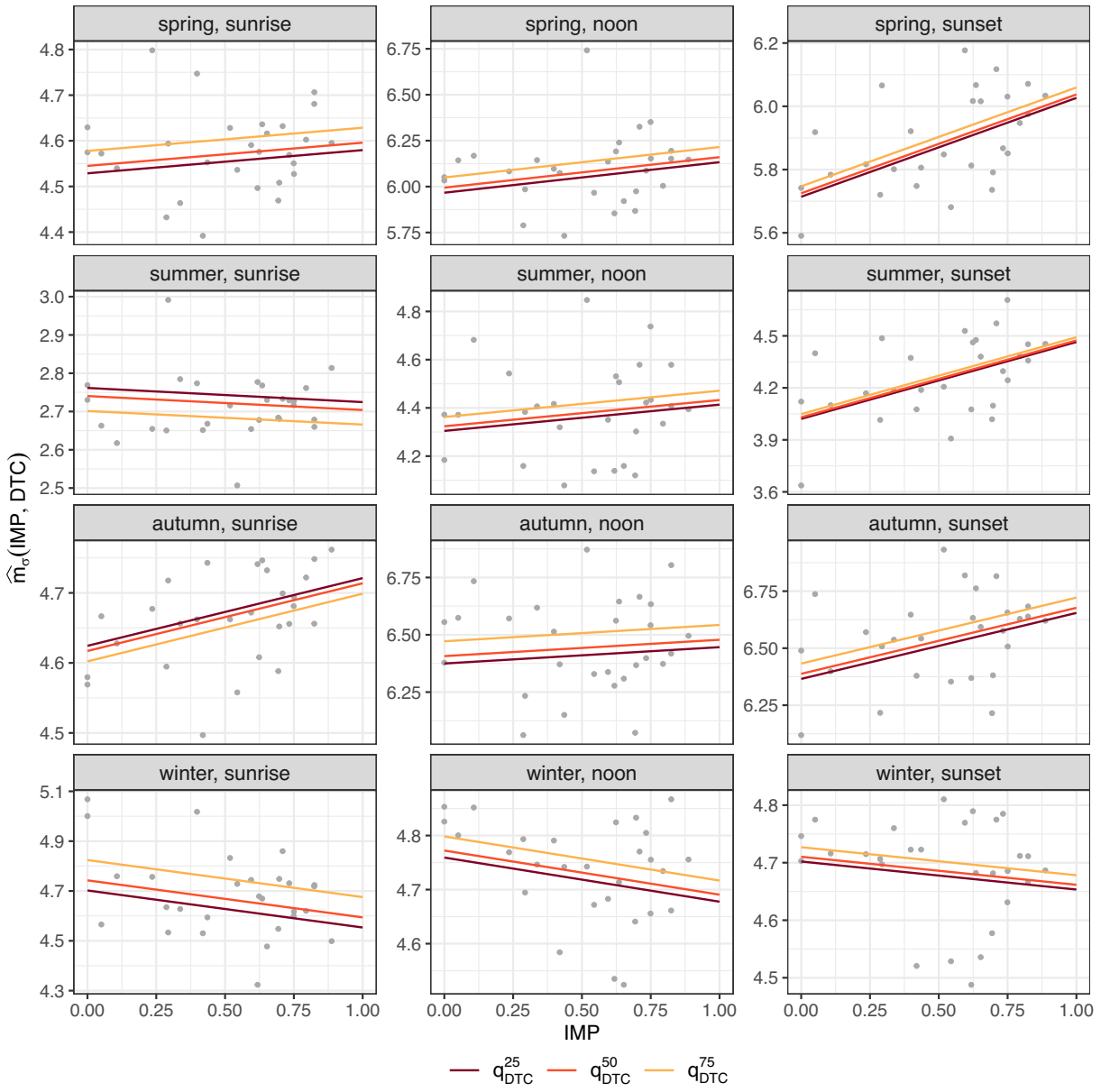

Fig. 8. Estimated nonparametric trend functions for spatial standard deviation in dependence of predictor IMP and for different quartiles of DTC for twelve sub-samples.

Fig. 12 aggregates the prediction errors and shows density curves over $RMSE_t$ derived from in-sample LOOCV in dependence of time of day and season for NP-OK(I), NP-OK(II), and IDW, where the vertical dashed lines indicate the respective average $RMSE_t$. For the time windows analysed, both variants of NP-OK outperform the benchmark approach IDW. While the $RMSE_t$ distributions hardly differ across models for the winter months, both variants of NP-OK clearly outrank IDW in summer. Remarkably, sub-sampling does not increase overall predictive performance. While NP-OK(I) provides small gains for noon hours of spring, autumn and winter as well as winter sunset, NP-OK(II) is slightly more accurate for summer sunset, spring sunrise and summer sunrise. A corresponding panel of density estimates over $BIAS_n$ is depicted in Fig. 13. Overall, the estimates referring to both variants of NP-OK are more concentrated around zero as compared to the estimates referring to IDW. Again, there seem to be no systematic deviations caused by neglecting efforts of sub-sampling.

4.5. Robustness check

Although results of Sections 4.1 to 4.4 provide comprehensive insights into implementation and performance of NP-OK, we

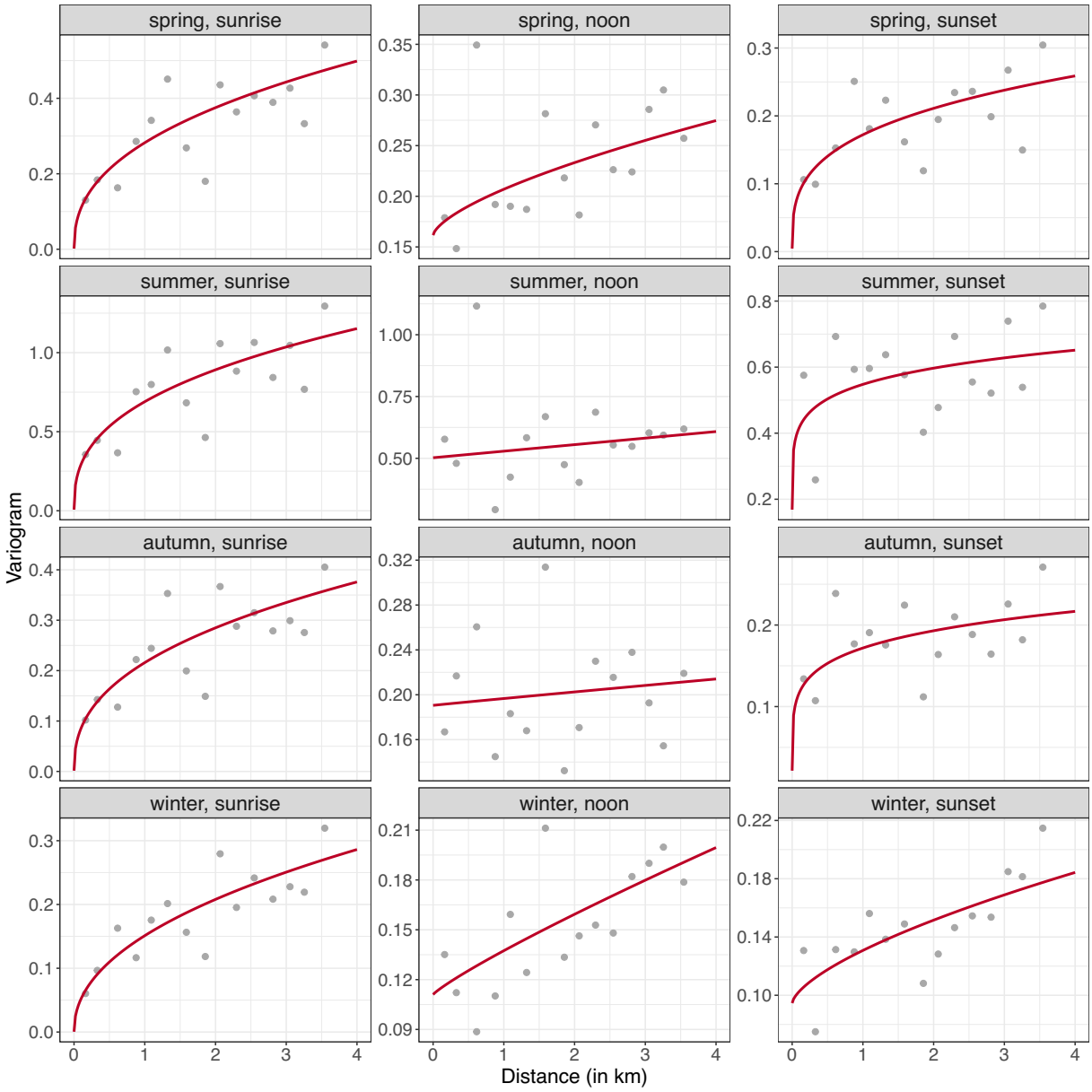

Fig. 9. Empirical variogram estimates and corresponding variogram model fits for twelve sub-samples.

complement the findings with a validation in the broad sense. It provides estimates of error measures for the randomly retained hold-out sample of size $H = 10$ utilising all available $N = 28$ in-sample sites for the purpose of model estimation. Corresponding results are presented in subtables b), d) and f) of Table 1. In subtables a), c) and e) of Table 1, the values corresponding to the dashed lines in the panels of Fig. 12 are summarised. Additionally, an average over sub-sample means of RMSE_t is given for each time of day, each season, and overall. The validation results indicate that both variants of NP-OK provide comparable overall accuracy, in terms of in-sample validation as well as in terms of hold-out validation. The hold-out results also support the choice of the predictors IMP and DTC, as performance does not drop systematically for the ten hold-out time series. On the contrary, we observe a deterioration of the results for the simple IDW benchmark, which again underlines the importance of spatial trend modelling for accurate predictions. This effect is especially noticeable for the summer months. In terms of LOOCV, the most significant improvements are obtained for summer and sunset, where NP-OK(I) provides an average RMSE_t reduction of ≈17% over IDW. For the hold-out stations, improvements of up to ≈36% (for summer and sunrise) can be observed. However, even for time windows where we find only modest differences in absolute performance between NP-OK(I) and IDW, e.g. for winter noon, the relative improvement still exceeds 7.5% for both LOOCV and the hold-out validation. This is also supported by the hold-out R_t^2, defined as

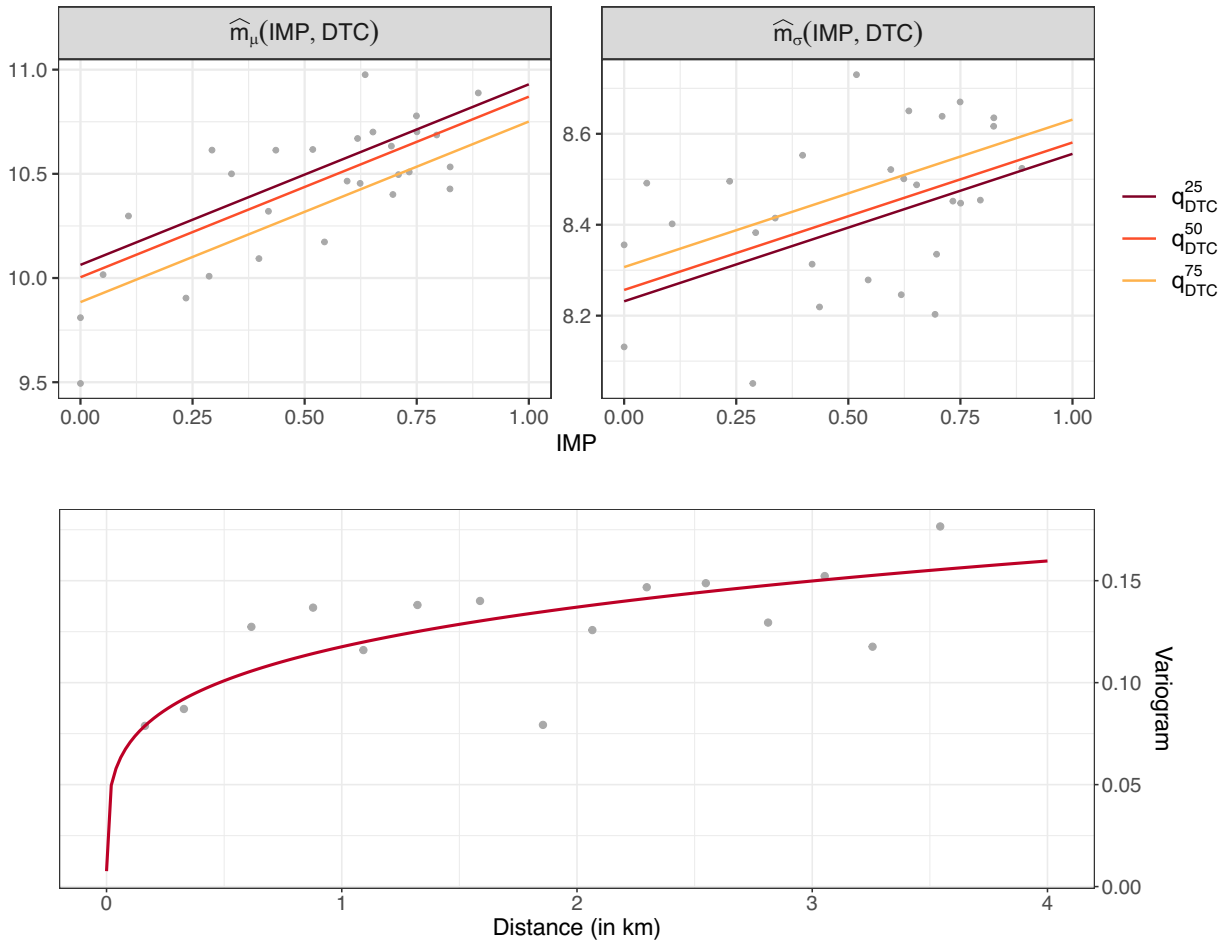

Fig. 10. Estimations based on pooled sub-samples: nonparametric trend functions for spatial mean and standard deviation in dependence of predictor IMP and for different quartiles of DTC (top row); empirical variogram estimates and corresponding variogram model fit (bottom).

$$R_t^2 = 1 - \frac{\sum_{h=1}^{H} \hat{e}_{h,t}^2}{\sum_{h=1}^{H} (z_{h,t} - \bar{z}_t)^2}$$

with $\bar{z}_t$ denoting the average over all hold-out observations at time instant t. At the median, R_t^2 is significantly higher for NP-OK(I) and NP-OK(II) (0.388 and 0.432, respectively) than for IDW (0.042), with the latter indicating performance only slightly superior to naive interpolation by $\bar{z}_t$.

To provide insight into the spatial error structure beyond the two exemplary prediction maps in Fig. 11, we visualise $BIAS_n$ over the hold-out sites for the exemplary time windows of summer noon and winter sunset, see Fig. 14. Comparison of the error structure shows that NP-OK(I) and NP-OK(II) again exhibit comparable behaviour. For the summer noon time window, there is a systematic overestimation of rural areas and a systematic underestimation of stations near the center, as is the case with IDW. However, the winter sunset time window shows a fairly smooth residual structure for NP-OK(I) and NP-OK(II), indicating a sufficiently well-specified trend modelling. The opposite is the case for IDW, where we observe a residual spatial trend for both time windows considered. In particular, $BIAS_n$ reaches its maximum for NP-OK(I), NP-OK(II), and IDW with $+0.90$, $+0.63$, and $+1.25$ for summer noon and $+0.57$, $+0.48$, and $+0.79$ for winter sunset, respectively.

5. Discussion

The empirical application of the proposed method reveals a number of interesting discussion points. First, our results based on cross-validation and hold-out validation suggest that sub-sampling does not increase overall prediction performance of NP-OK. These findings are of particular interest, since our former analyses reveal remarkably different patterns of heterogeneity and spatial dependence over the twelve time windows considered, indicating a loss in potentially useful information when pooling those sub-samples. The increase in the size of the training set seems to compensate for this loss of information. The choice of a particular

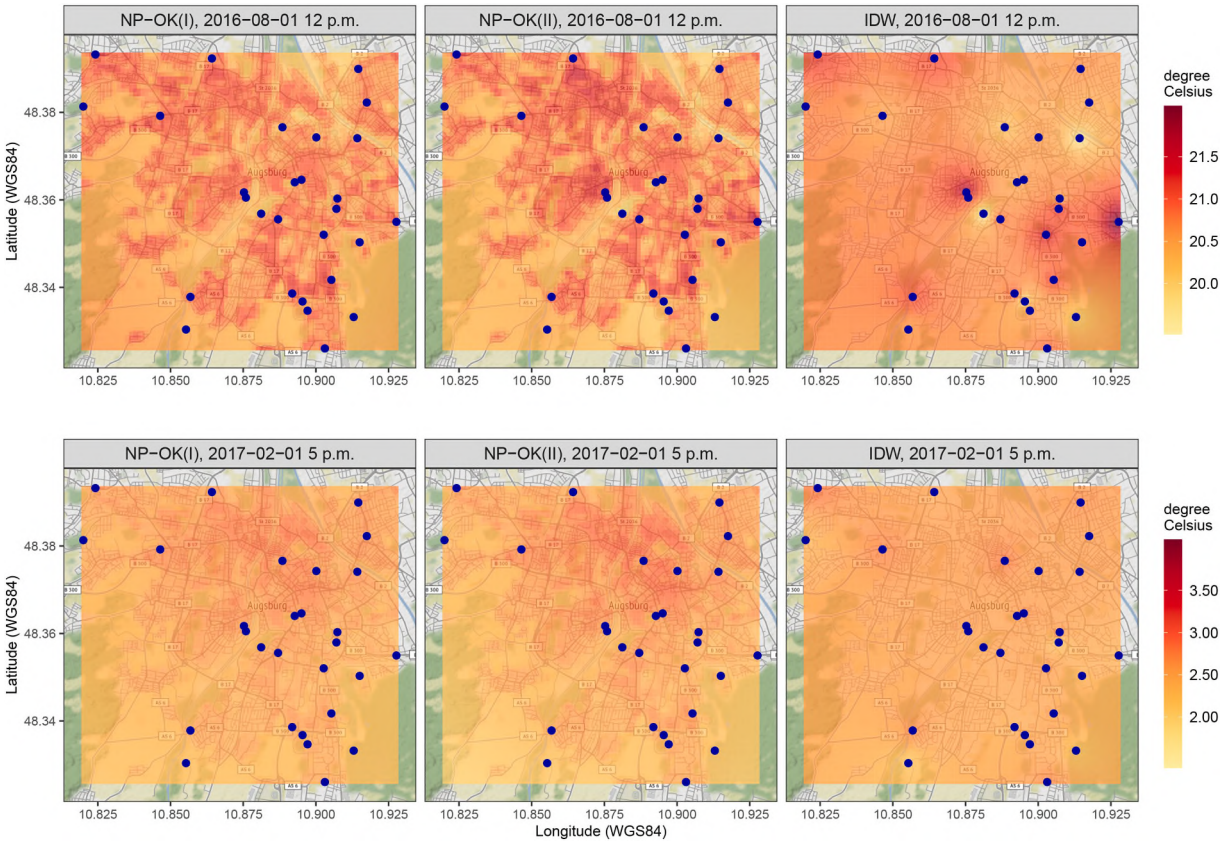

Fig. 11. Prediction maps for 1st of August 2016, 12 p.m., and 1st of February 2016, 5 p.m., on a resolution of 100 m × 100 m, derived from NP-OK (I), NP-OK(II), and IDW; locations of in-sample monitoring sites marked by blue dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

level of aggregation thus seems to be a trade-off in which one might willingly refrain from (generally reasonable) sub-sampling, especially in the case of sparse data. From a methodological point of view, one could also consider a completely different sampling scheme, e.g., combining day and night times (Amorim et al., 2021) or times corresponding to certain weather conditions, such as calm and clear weather, by introducing information on cloud cover or wind speed (Straub et al., 2019).

Second, we propose two predictors of spatial variation that include small-scale (percentage of imperviousness, IMP) and large-scale (distance to city center, DTC) information on urbanisation. As evaluated in detail in Appendix A, additional predictors for local climate zones and land cover categories do not improve predictive performance in our data setting. This is also consistent with the findings of Núñez-Peiró et al. (2021), who report strong collinearity between the predictors IMP and DTC and several additional predictors for local climate zones. In this respect, these two predictors already capture a substantial amount of the relevant information. However, this result does not mean that the additional predictors are generally unsuitable for modelling the spatial mean and standard deviation of urban temperature processes. A simple approach relying only on IMP and DTC can only approximate the spatial heterogeneity of air temperature processes. However, with only $N = 28$ training samples available, the additional model flexibility provided by more predictors risks overfitting and degrading generalisation performance, preventing the use of additional information. This is a well-known problem when highly flexible nonparametric models are applied to a relatively small sample of measurement points. From this point of view, it would certainly be desirable to have more measurement points and more detailed information about the site environment in question. However, in the present case, the simple approach achieves better prediction accuracy in terms of out-of-sample RMSE. For the same reason, the automated bandwidth selection leads to very conservative, approximately linear trend functions. Under such circumstances the full flexibility of nonparametric regression does not really come into play. Still, the use of a simple linear extrapolation is not necessarily disadvantageous, especially for support-remote prediction of environmental processes such as air temperature (see, e.g., Hsieh, 2020).

Third, an adjustment of the data cleaning routine might have some impact on our results. Restricting our analysis to approximately complete data is useful for an initial investigation of the proposed NP-OK method and the effects of sub-sampling, as it ensures a comparable number of observations for each sub-model. However, the NP-OK method itself is quite robust to missing data, as neither trend modelling nor pooled variogram estimation are technically affected by a higher percentage of missing data.

Fourth, the modularity and flexibility of the proposed NP-OK approach allow easy configuration of spatial trend models for other urban regions with different size, structure, or measurement network coverage. Furthermore, the approach allows predictions at

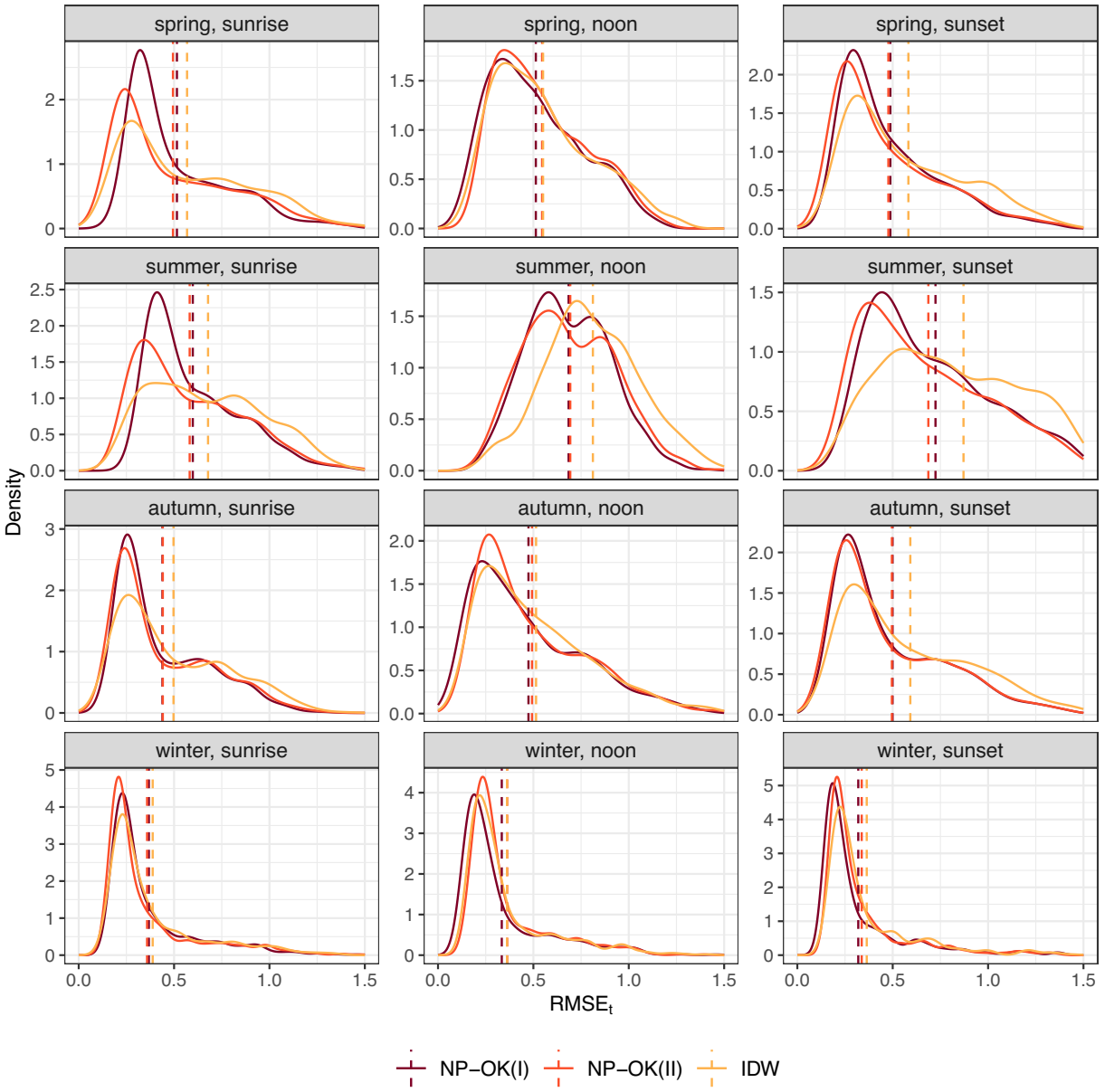

Fig. 12. Density curves for RMSE$_t$ derived from in-sample LOOCV in dependence of time of day and season for NP-OK(I), NP-OK(II), and IDW; dashed lines show the sub-sample averages of RMSE$_t$.

arbitrary spatial resolution, technically only limited by the resolution of the available predictors. We would like to highlight the relevance of such flexible high-resolution interpolators by illustrating two potential areas of application: In epidemiology, multilevel (ensemble) machine learning models have been developed nationwide to estimate daily air temperature with a spatial resolution of 1×1 km^2 from satellite-based land surface temperature, e.g. for France (Kloog et al., 2017), Israel (Rosenfeld et al., 2017; Zhou et al., 2020), Sweden (Jin et al., 2022) and the USA (Kloog et al., 2014). Although such spatial resolution is suitable for assessing the health effects of air temperature in large nationwide epidemiological cohorts, it is too coarse for e.g. smaller cohorts such as the KORA cohort (Cooperative Health Research in the Region Augsburg), which only covers a study area of 45×45 km^2 (Holle et al., 2005). And although a recent French model could increase the spatial resolution for the urban areas to 200×200 m^2 (Hough et al., 2020), a dense network of monitoring sites and a multitude of predictor data is necessary to develop similar models. In contrast, the geostatistical NP-OK approach is shown to provide reliable high-resolution prediction maps of urban air temperature based on relatively few measurement stations and only two spatial predictors. Such high-resolution prediction maps are also a relevant tool for urban planning with regard to different aspects. They enable the spatio-temporal detection and delineation of potential urban canopy hot and cold spots which can then be adequately considered in the development of mitigation and adaptation strategies as for example the

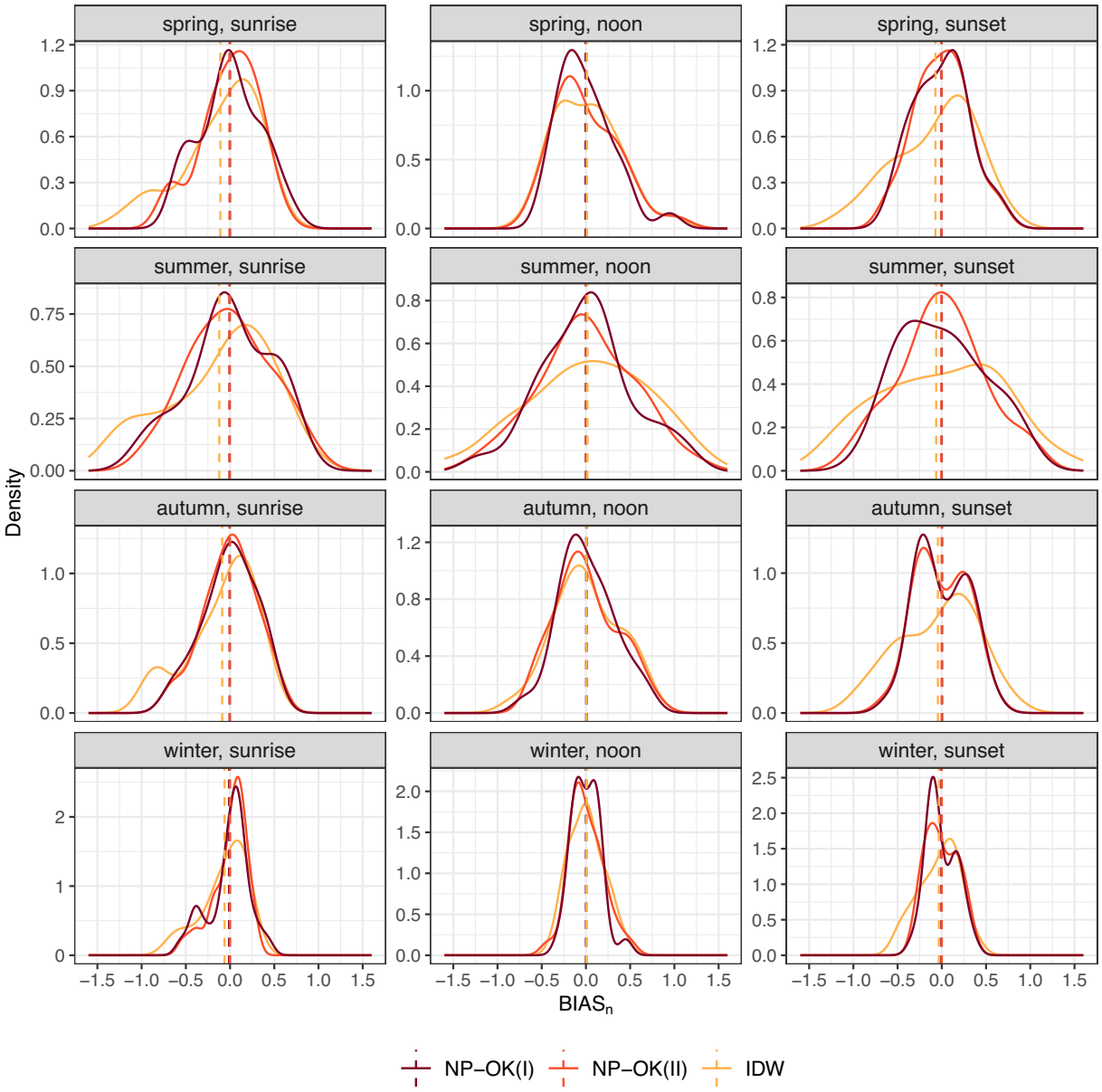

Fig. 13. Density curves for BIAS$_n$ derived from in-sample LOOCV in dependence of time of day and season for NP-OK(I), NP-OK(II), and IDW; dashed lines show the sub-sample averages of BIAS$_n$.

enhancement of urban vegetation in order to positively alter the local radiation and energy balance via shading and evapotranspiration (Demuzere et al., 2014; Emmanuel and Loconsole, 2015). Moreover, applying our mapping of spatio-temporal urban air temperatures to changing urban environments may help monitoring the local effects of urban climate mitigation-adaptation measures. In this context, the here presented NP-OK approach is neither intended nor suited to replace approaches for estimating the effects of urban planning scenarios on air temperature that are based on numerical models – e.g. ENVImet (Bruse and Fleer, 1998; Beck et al., 2021) or PALM-4U (Maronga et al., 2019; Straub et al., 2020). Rather, both approaches – numerical and geostatistical modeling - can complement each other for deriving reliable estimates of the urban canopy temperature variations under urban and climate change conditions.

And fifth, in further applications of the residual interpolation optimised paradigm, one could try to overcome the limitations of time-invariant spatial behaviour imposed on NP-OK not by means of sub-sampling, but by interpolation routines that do not require the assumptions (a)–(c) to be satisfied. For example, regression-tree based algorithms have recently gained popularity in the area of spatial and spatio-temporal interpolation (Hengl et al., 2018; Georganos et al., 2019). Such an adaptation would combine the data-driven and assumption-free nature of machine learning based interpolation with the flexibility and interpretability of a residual interpolation

Table 1
$RMSE_t$ results of in-sample LOOCV and hold-out validation for NP-OK(I), NP-OK(II), and IDW, by season and time of day.

	Sunrise	Noon	Sunset	Avg.
		(a) NP-OK(I), in-sample.		
Spring	0.5150	0.5128	0.4875	0.5051
Summer	0.5982	0.6846	0.7243	0.6690
Autumn	0.4390	0.4742	0.4992	0.4708
Winter	0.3678	0.3342	0.3195	0.3405
Avg.	0.4800	0.5015	0.5076	**0.4964**
		(b) NP-OK(I), hold-out.		
Spring	0.4923	0.5474	0.4210	0.4869
Summer	0.4913	0.6321	0.5798	0.5677
Autumn	0.4679	0.4555	0.5009	0.4748
Winter	0.4569	0.3680	0.3732	0.3994
Avg.	0.4771	0.5007	0.4687	**0.4822**
		(c) NP-OK(II), in-sample.		
Spring	0.4942	0.5456	0.4772	0.5057
Summer	0.5817	0.6933	0.6868	0.6539
Autumn	0.4410	0.4929	0.4955	0.4765
Winter	0.3585	0.3627	0.3376	0.3529
Avg.	0.4688	0.5236	0.4993	**0.4972**
		(d) NP-OK(II), hold-out.		
Spring	0.5290	0.5758	0.4185	0.5078
Summer	0.5383	0.6436	0.5243	0.5687
Autumn	0.4997	0.4756	0.4993	0.4915
Winter	0.4832	0.3901	0.3723	0.4152
Avg.	0.5125	0.5213	0.4536	**0.4958**
		(e) IDW, in-sample		
Spring	0.5680	0.5508	0.5821	0.5670
Summer	0.6781	0.8119	0.8712	0.7871
Autumn	0.4970	0.5141	0.5923	0.5344
Winter	0.3879	0.3617	0.3638	0.3711
Avg.	0.5327	0.5596	0.6024	**0.5649**
		(f) IDW, hold-out		
Spring	0.6776	0.5802	0.5872	0.6150
Summer	0.7624	0.7329	0.7477	0.7476
Autumn	0.6587	0.5135	0.6890	0.6204
Winter	0.6049	0.4063	0.4754	0.4955
Avg.	0.6759	0.5582	0.6248	**0.6197**

paradigm.

6. Conclusions

The processes that influence air temperature in cities have interrelated spatial and temporal characteristics. Spatial patterns such as anisotropic dependence structures may change over time, while temporal heterogeneities such as diurnal and seasonal cycles may change in space. Therefore, neglecting the temporal dimension in spatial modelling can have undesirable effects and potentially useful information can be lost. However, accounting for a three-dimensional spatio-temporal dependence structure is very costly and requires strong assumptions. In this paper, we demonstrate the time-varying spatial heterogeneity of urban air temperature using data for the Augsburg urban area and propose a two-stage spatial interpolation procedure, discussing the use of sub-sampling in terms of predictive performance and satisfaction of theoretical assumptions. The proposed NP-OK method takes into account the temporal dimension, but does not require the estimation of a complete spatio-temporal dependence structure. Based on a simple geostatistical model, NP-OK consists of nonparametric spatial detrending followed by residual-based interpolation using kriging, which provides a flexible and modular framework for spatio-temporal modelling. In our empirical application, we use two predictors for the nonparametric detrending, including small-scale (percentage of imperviousness, IMP) and large-scale (distance to city center, DTC) information on urbanisation. We show that NP-OK outperforms a simple inverse distance weighting interpolation with only $N = 28$ sampling sites, especially in the summer months when temperature interpolation tends to be more difficult. At the same time, the method appears to be robust to violations of theoretical assumptions, as convincing results are obtained even without prior subsampling of the data. The study of the resulting trend functions as well as the interpolation maps contributes to the understanding of the time-varying spatial heterogeneity of urban air temperature and the corresponding drivers of canopy layer urban heat islands. This understanding can form a basis for further investigating the impacts of urban heat and mitigation strategies (e.g. Manoli et al., 2019; Martilli et al., 2020). However, the proposed method is not limited to this specific application domain, but can be easily applied to other study regions and data due to its generic nature.

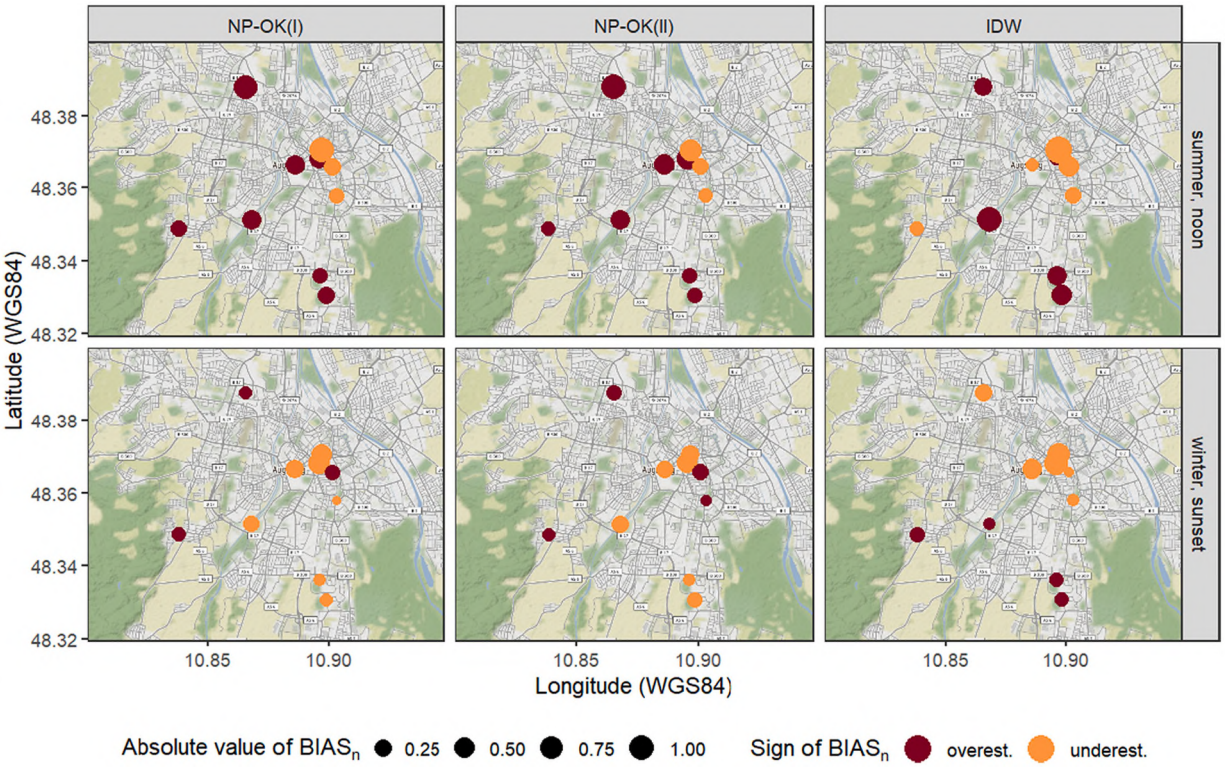

Fig. 14. Spatial distribution of each hold-out site's $BIAS_n$ for NP-OK(I), NP-OK(II), and IDW; for two exemplary sub-samples: summer noon and winter sunset.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The logger network was supported by the Helmholtz Climate Initiative REKLIM (Regional Climate Change and Humans). The authors express their thanks to Cornelius Hald, Benjamin Kühlbach, Thomas Kusch, Andreas Oswald, Katrin Schönhofer and Robin Umminger for technical maintenance of the logger network and to Uta Geruschkat for the management and processing of the logger data.

Appendix A. Additional predictors

Apart from predictors IMP and DTC, we also analyse the influence of information on land use and local climate zones as discussed by Beck et al. (2018). For our study region, we derive the share of the respective land use and local climate zone classes on a buffer of 100 m radius around a certain location of interest. The following table provides information on the resulting predictor variables, their sample average over the $N = 28$ in-sample sites and the corresponding average LOOCV-$RMSE_t$ when considered as additional third predictor in NP-OK(I):

Predictor name	Predictor description	Avg. sample buffer share	Avg. LOOCV- $RMSE_t$
UrbFab	Continuous and discontinuous urban fabric	57.7 %	0.5027
ArtiVeg	Green urban areas,sport and leisure facilities	16.1%	0.5207
IndusComerc	Industrial or commercial units,road, rail, (air)ports	14.1%	0.5081
Forest	Broad-leaved forest, coniferous forest, mixed forest	4.8%	-
Pastures	Agricultural use with significant share of natural vegetation	3.8%	0.5896
Arable	Non-irrigated arable land	2.2%	-

(*continued on next page*)

(*continued*)

Predictor name	Predictor description	Avg. sample buffer share	Avg. LOOCV- $RMSE_t$
Water	Water courses,water bodies	1.2%	0.5217
LCZ_2	Compact mid-rise development	4.2%	-
LCZ_5	Open mid-rise development	51.5%	0.4996
LCZ_6	Open low-rise development	14.8%	0.5021
LCZ_8	Large low-rise development	11.2%	0.4985
LCZ_{1XX}	Dense or scattered trees, plants, bare soil or sand	18.4%	0.5021

Note that predictors Forest, Arable and LCZ_2 cover less than 5 distinct values and are therefore excluded from this robustness check. Furthermore, nonparametric trend functions implemented in `R`-package `npreg` face issues when deriving predictions if predictor values are heavily out-of-support. Under such circumstances we fall back on a simple linear extrapolation (see also Hsieh, 2020). Considering the average LOOCV-$RMSE_t$ result of 0.4964 achieved by predictors IMP and DTC (Table 1), employing an additional predictor does not allow to improve the predictive performance of NP-OK(I). To complement this finding, Fig. A.15 examines the relationship between LOOCV-$BIAS_n$ and the respective monitoring site's predominant land use category (left) and local climate zone (right). While IDW reveals a clear systematic, NP-OK(I) and NP-OK(II) provide $BIAS_n$ results mostly centred around zero. This indicates that information already captured in IMP and DTC is a good proxy for potential influence of land use and local climate zones.

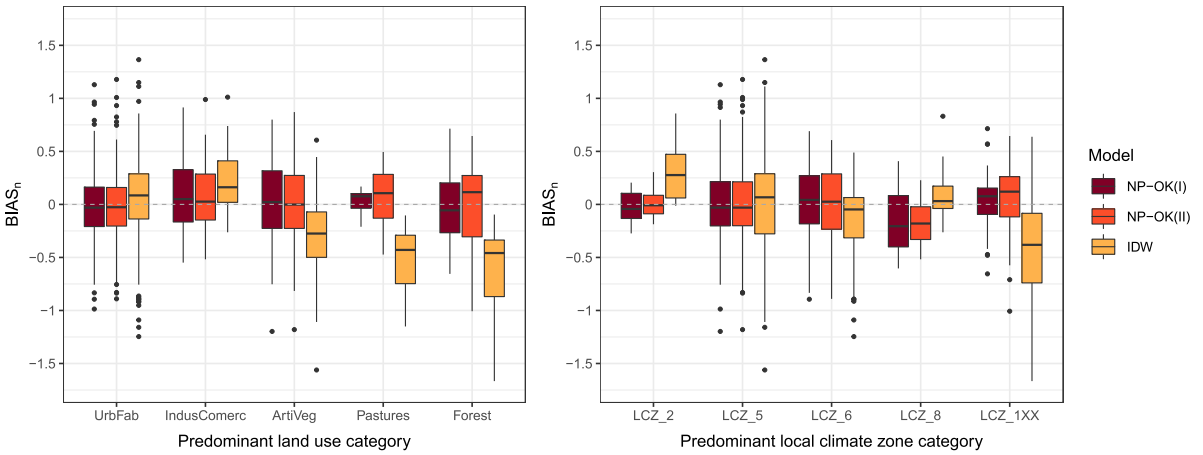

Fig. A.15. Boxplots over LOOCV-$BIAS_n$ results for NP-OK(I), NP-OK(II), and IDW grouped by predominant land use category (left) and predominant local climate zone category (right).

Appendix B. Additional results for ordinary kriging

To assess the impact of detrending and pooled semivariogram estimation on model performance, we report LOOCV and hold-out validation results for pure spatial interpolation using ordinary kriging (OK) in Table B.2. In this approach, the semivariances are estimated based only on the cross-sectional observations available at time $t \in D_t$, and just as in IDW, no detrending is performed in advance.

Table B.2
$RMSE_t$ results of in-sample LOOCV and hold-out validation for ordinary kriging, by season and time of day.

	Sunrise	Noon	Sunset	Avg.
		(a) OK, in-sample		
Spring	0.5557	0.5410	0.5812	0.5593
Summer	0.6836	0.7396	0.8402	0.7545
Autumn	0.4867	0.4875	0.5714	0.5152
Winter	0.3762	0.3556	0.3596	0.3638
Avg.	0.5256	0.5309	0.5881	**0.5482**
		(b) OK, hold-out		
Spring	0.6138	0.5701	0.5485	0.5774
Summer	0.6663	0.7117	0.7242	0.7007
Autumn	0.6148	0.5007	0.6710	0.5955
Winter	0.5659	0.4097	0.4662	0.4806
Avg.	0.6152	0.5480	0.6025	**0.5886**

The observed results are consistent with our earlier findings: OK leads to slightly better results than IDW. This is due to the explicit

modeling of the spatial dependence structures in terms of semivariance estimates, while IDW imposes a fixed structure independent of the data. Compared to NP-OK(I) and NP-OK(II) (see Table 1), the error for OK is higher for both in-sample LOOCV and hold-out validation, suggesting that the modeling of spatial heterogeneity and the pooled semivariogram estimation actually have a positive effect on predictive performance.

Appendix C. Software information

All empirical analysis is conducted with the statistical software `R` (R Core Team, 2013) using the packages `cowplot` (Wilke, 2020), `data.table` (Dowle and Srinivasan, 2019), `dplyr` (Wickham et al., 2021), `geosphere` (Hijmans et al., 2019), `ggmap` (Kahle and Wickham, 2013), `ggplot2` (Wickham, 2016), `GISTools` (Brunsdon and Chen, 2014), `gridExtra` (Auguie, 2017), `gstat` (Gräler et al., 2016), `lubridate` (Grolemund and Wickham, 2011), `metR` (Campitelli, 2021), `np` (Hayfield and Racine, 2008), `raster` (Hijmans, 2019), `RColorBrewer` (Neuwirth, 2014), `reshape2` (Wickham, 2007), `rgdal` (Bivand et al., 2019), `sf` (Pebesma, 2018), `sp` (Pebesma and Bivand, 2005), `spatstat` (Baddeley and Turner, 2005), `stars` (Pebesma, 2020), `tseries` (Trapletti and Hornik, 2020), `xlsx` (Dragulescu and Arendt, 2018) and `zoo` (Zeileis and Grothendieck, 2005).

References

Amorim, M.C.C.T., Dubreuil, V., Amorim, A.T., 2021. Day and night surface and atmospheric heat islands in a continental and temperate tropical environment. Urban Clim. 38, 100918. https://doi.org/10.1016/j.uclim.2021.100918.

Armstrong, M., 1998. Basic Linear Geostatistics. Springer, Berlin, Heidelberg.

Arnds, D., Böhner, J., Bechtel, B., 2015. Spatio-temporal variance and meteorological drivers of the urban heat island in a European city. Theor. Appl. Climatol. 128, 43–61. https://doi.org/10.1007/s00704-015-1687-4.

Aubrecht, C., Özceylan, D., 2013. Identification of heat risk patterns in the U.S. national capital region by integrating heat stress and related vulnerability. Environ. Int. 5, 65–77. https://doi.org/10.1016/j.envint.2013.03.005.

Auguie, B., 2017. gridExtra: Miscellaneous Functions for 'Grid' Graphics. https://CRAN.R-project.org/package=gridExtra. , R package version 2.3.

Baddeley, A., Turner, R., 2005. Spatstat: anr package for analyzing spatial point patterns. J. Stat. Software 12, 1–42. https://doi.org/10.18637/jss.v012.i06.R package version 2.2-0.

Basu, R., Ostro, B.D., 2008. A multicounty analysis identifying the populations vulnerable to mortality associated with high ambient temperature in California. Am. J. Epidemiol. 168, 632–637. https://doi.org/10.1093/aje/kwn170.

Beck, C., Buse, K., Fritsch, M., Irber, P., 2021. Numerical simulation of thermal effects of urban green enhancements in different urban environments in the city of Augsburg (Bavaria, SE Germany). EGU Gener. Assem. 2021 online, 19-30 Apr EGU21-3562.

Beck, C., Straub, A., Breitner, S., Cyrys, J., Philipp, A., Rathmann, J., Schneider, A., Wolf, K., Jacobeit, J., 2018. Air temperature characteristics of local climate zones in the Augsburg urban area (Bavaria, southern Germany) under varying synoptic conditions. Urban Clim. 25, 152–166. https://doi.org/10.1016/j.uclim.2018.04.007.

Behm, S., Haupt, H., Schmid, A., 2018. Spatial detrending revisited: modelling local trend patterns in NO_2-concentration in Belgium and Germany. Spatial Stat. 28, 331–351. https://doi.org/10.1016/j.spasta.2018.04.004.

Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., et al., 2012. Characterising performance of environmental models. Environ. Modell. Software 40, 1–20. https://doi.org/10.1016/j.envsoft.

Bivand, R., Keitt, T., Rowlingson, B., 2019. Rgdal: Bindings for the 'Geospatial' Data Abstraction Library. https://CRAN.R-project.org/package=rgdal. , R package version 1.5-27.

Brunsdon, C., Chen, H., 2014. GISTools: Some further GIS capabilities for R. https://CRAN.R-project.org/package=GISTools. , R package version 0.7-4.

Bruse, M., Fleer, H., 1998. Simulating surface-plant-air interactions inside urban environments with a three dimensional numerical model. Environ. Modell. Software 13, 373–384. https://doi.org/10.1016/S1364-8152(98)00042-5.

Calder, C., Cressie, N., 2009. Kriging and variogram models. In: Kitchin, R., Thrift, N. (Eds.), International Encyclopedia of Human Geography. Elsevier, pp. 49–55. https://doi.org/10.1016/B978-008044910-4.00461-2.

Campitelli, E., 2021. Metr: Tools for Easier Analysis of Meteorological Fields. https://doi.org/10.5281/zenodo.2593516.R package version 0.11.0.

Chilés, J., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. John Wiley and Son Inc., Hoboken, New Jersey.

Chun, B., Guldmann, J.M., 2014. Spatial statistical analysis and simulation of the urban heat island in high-density central cities. Landscape Urban Plann. 125, 76–88. https://doi.org/10.1016/j.landurbplan.2014.01.016.

Coseo, P., Larsen, L., 2014. How factors of land use/land cover, building con?.guration, and adjacent heat sources and sinks explain Urban Heat Islands in Chicago. Landscape Urban Plann. 125, 117–129. https://doi.org/10.1016/j.landurbplan.2014.02.019.

Cressie, N.A.C., 1993. Statistics for Spatial Data Wiley Series in Probability and mathematical Statistics. Revised. Wiley, New York. https://doi.org/10.1111/j.1365-3121.1992.tb00605.x.

EEA, 2016. Copernicus Land Monitoring Service - High Resolution Layer Imperviousness: Product Specifications Document. https://land.copernicus.eu/user-corner/technical-library/hrl-imperviousness-technical-document-prod-2015. Accessed on May 6, 2020.

De Iaco, S., Myers, D., Posa, D., 2002. Nonseparable space-time covariance models: some parametric families. Mathemat. Geol. 34, 23–42. https://doi.org/10.1023/A:1014075310344.

DeGaetano, A.T., Belcher, B.N., 2007. Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. J. Appl. Meteorol. Climatol. 46, 1981–1992. https://doi.org/10.1175/2007JAMC1536.1.

Demuzere, M., Orru, K., Heidrich, O., Olazabal, E., Geneletti, D., Orru, H., Bhave, A., Mittal, N., Feliu, E., Faehnle, M., 2014. Mitigating and adapting to climate change: multi-functional and multi-scale assessment of green urban infrastructure. J. Environ. Manage. 146, 107–115. https://doi.org/10.1016/j.jenvman.2014.07.025.

Dowle, M., Srinivasan, A., 2019. Data.Table: Extension of 'Data.Frame'. https://CRAN.R-project.org/package=data.table. , R package version 1.14.2.

Dragulescu, A.A., Arendt, C., 2018. Xlsx: Read, Write Format Excel 2007 and Excel 97/2000/XP/2003 Files. https://CRAN.R-project.org/package=xlsx. , R package version 0.6.5.

Emmanuel, R., Loconsole, A., 2015. Green infrastructure as an adaptation approach to tackling urban overheating in the glasgow clyde valley region. UK. Landscape Urban Plann. 138, 71–86. https://doi.org/10.1016/j.landurbplan.2015.02.012.

Esau, I., Miles, V., Varentsov, M., Konstantinov, P., Melnikov, V., 2019. Spatial structure and temporal variability of a surface urban heat island in cold continental climate. Theor. Appl. Climatol. 137, 2513–2528.

Fenner, D., Meier, F., Scherer, D., Polze, A., 2014. Spatial and temporal air temperature variability in Berlin. Germany During years 2001–2010. Urban Clim. 10, 308–331. https://doi.org/10.1016/j.uclim.2014.02.004 iCUC8: the 8th international conference on urban climate and the 10th symposium on the urban environment.
Fischer, E.M., Oleson, K.W., Lawrence, D.M., 2012. Contrasting urban and rural heat stress responses to climate change. Geophys. Res. Lett. 39 https://doi.org/10.1029/2011GL050576.
Georganos, S., Grippa, T., Gadiaga, A.N., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2019. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int. https://doi.org/10.1080/10106049.2019.1595177.
Gilabert, J., Ventura, S., Segura, R., Martilli, A., Badia, A., Llasat, C., Corbera, J., Villalba, G., 2021. Abating heat waves in a coastal Mediterranean city: what can cool roofs and vegetation contribute? Urban Clim. 37, 100863. https://doi.org/10.1016/j.uclim.2021.100863.
Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. R J. 8, 204–218. https://doi.org/10.32614/RJ-2016-014.R package version 2.0-8.
Gräler, B., Gerharz, L., Pebesma, E., 2012. Spatio-Temporal Analysis and Interpolation of PM10 Measurements in Europe. ETC/ACM Technical Paper 2011/10. ETC/ACM, Bilthoven, Netherlands.
Grolemund, G., Wickham, H., 2011. Dates and times made easy with 'lubridate'. J. Stat. Software 40, 1–25. https://doi.org/10.18637/jss.v040.i03.R package version 1.8.0.
Hayfield, T., Racine, J.S., 2008. Nonparametric econometrics: the np package. J. Stat. Software 27. http://www.jstatsoft.org/v27/i05/. , R package version 0.60-11.
He, C., Ma, L., Zhou, L., Kan, H., Zhang, Y., Ma, W., Chen, B., 2019. Exploring the mechanisms of heat wave vulnerability at the urban scale based on the application of big data and artificial societies. Environ. Int. 127, 573–583. https://doi.org/10.1016/j.envint.2019.01.057.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518.
Hijmans, R., Williams, E., Vennes, C., 2019. Geosphere: Spherical Trigonometry. https://CRAN.R-project.org/package=geosphere. , R package version 1.5-14.
Hijmans, R.J., 2019. raster: Geographic Data Analysis and Modeling https://CRAN.R-project.org/package=raster.R package version 3.5-2.
Holle, R., Happich, M., Löwel, H., Wichmann, H., 2005. Kora - a research platform for population based health research. Gesundheitswesen 67 Suppl. 1, 19–25. https://doi.org/10.1055/s-2005-858235.
Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., 2006. Spatial interpolation of ambient ozone concentrations from sparse monitoring points in Belgium. J. Environ. Monit. 8, 1129–1135. https://doi.org/10.1039/b612607n.
Hough, I., Just, A.C., Zhou, B., Dorman, M., Lepeule, J., Kloog, I., 2020. A multi-resolution air temperature model for France from MODIS and Landsat thermal data. Environ. Res. 183 https://doi.org/10.1016/j.envres.2020.109244.
Hsieh, W.W., 2020. Improving Predictions by Nonlinear Regression Models from Outlying Input Data. http://arxiv.org/abs/2003.07926arXiv:2003.07926.
Hurvich, C., Simonoff, J., Tsai, C., 1998. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. J. Royal Stat. Soc. 60, 271–293. https://doi.org/10.1111/1467-9868.00125.
Janssen, S., Dumont, G., Fierens, F., Mensink, C., 2008. Spatial interpolation of air pollution measurements using CORINE land cover data. Atmos. Environ. 42, 4884–4903. https://doi.org/10.1016/j.atmosenv.2008.02.043.
Jin, Z., Ma, Y., Chu, L., Liu, Y., Dubrow, R., Chen, K., 2022. Predicting spatiotemporally-resolved mean air temperature over Sweden from satellite data using an ensemble model. Environ. Res. 204, 111960. https://doi.org/10.1016/j.envres.2021.111960.
Kahle, D., Wickham, H., 2013. Ggmap: spatial visualization with 'ggplot2'. R J. 5, 144–161. https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf. , R package version 3.0.0.
Kauermann, G., Haupt, H., Kaufmann, N., 2012. A hitchhiker's view on spatial statistics and spatial econometrics for lattice data. Stat. Modell. 12, 419–440. https://doi.org/10.1177/1471082X12457493.
Kloog, I., Chudnovsky, A., Koutrakis, P., Schwartz, J., 2012. Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts. USA. Sci. Total Environ. 432, 85–92. https://doi.org/10.1016/j.scitotenv.2012.05.095.
Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2014. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. Remote Sens. Environ. 150, 132–139. https://doi.org/10.1016/j.rse.2014.04.024.
Kloog, I., Nordio, F., Lepeule, J., Padoan, A., Lee, M., Auffray, A., Schwartz, J., 2017. Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data. Int. J. Climatol. 37, 296–304. https://doi.org/10.1002/joc.4705.
Kotharkar, R., Bagade, A., Ramesh, A., 2019. Assessing urban drivers of canopy layer urban heat island: a numerical modeling approach. Landscape Urban Plann. 190, 103586. https://doi.org/10.1016/j.landurbplan.2019.05.017.
Lazzarini, M., Molini, A., Marpu, P.R., Ouarda, T.B., Ghedira, H., 2015. Urban climate modifications in hot desert cities. Role Land Cover. Loc. Clim. Season. 42, 9980–9989. https://doi.org/10.1002/2015GL066534.
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.
Maronga, B., Gross, G., Raasch, S., Banzhaf, S., Forkel, R., Heldens, W., Kanani-SLtdc:ühring, F., Matzarakis, A., Mauder, M., Pavlik, D., Pfafferott, J., Schubert, S., Seckmeyer, G., Sieker, H., Winderlich, K., 2019. Development of a new urban climate model based on the model PALM - project overview, planned work, and first achievements. Meteorol. Zeitschrift 28, 105–119. https://doi.org/10.1127/metz/2019/0909.
Martilli, A., Krayenhoff, E.S., Nazarian, N., 2020. Is the urban heat island intensity relevant for heat mitigation studies? Urban Clim. 31, 100541. https://doi.org/10.1016/j.uclim.2019.100541.
McCarthy, M.P., Best, M.J., Betts, R.A., 2010. Climate change in cities due to global warming and urban effects. Geophys. Res. Lett. 37 https://doi.org/10.1029/2010GL042845.
Mirzaei, P.A., 2015. Recent challenges in modeling of urban heat island. Sustain. Cities Soc. 19, 200–206. https://doi.org/10.1016/j.scs.2015.04.001.
Montero, J.M., FernLtdc:ández-AvilLtdc:és, G., Mateu, J., 2015. Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118762387
Mueller, N., Rojas-Rueda, D., Khreis, H., Cirach, M., Andrés, D., Ballester, J., Bartoll, X., Daher, C., Deluca, A., Echave, C., Milá, C., Márquez, S., Palou, J., Pérez, K., Tonne, C., Stevenson, M., Rueda, S., Nieuwenhuijsen, M., 2020. Changing the urban design of cities for health: the superblock model. Environ. Int. 134, 105132. https://doi.org/10.1016/j.envint.2019.105132.
Neuwirth, E., 2014. RColorBrewer: ColorBrewer Palettes. https://CRAN.R-project.org/package=RColorBrewer. , R package version 1.1-2.
Núñez-Peiró, M., Sánchez-Guevara Sánchez, C., Neila González, F.J., 2021. Hourly evolution of intra-urban temperature variability across the local climate zones. Case Madrid. Urban Clim. 39, 100921. https://doi.org/10.1016/j.uclim.2021.100921.
Nieuwenhuijsen, M.J., 2020. Urban and transport planning pathways to carbon neutral, liveable and healthy cities; a review of the current evidence. Environ. Int. 140, 105661. https://doi.org/10.1016/j.envint.2020.105661.
Ninyerola, M., Pons, X., Roure, J.M., 2007. Objective air temperature mapping for the iberian peninsula using spatial interpolation andGIS. Int. J. Climatol: J. Royal Meterol. Soc. 27, 1231–1242. https://doi.org/10.1002/joc.1462.
Oleson, K.W., Bonan, G.B., Feddema, J., 2010. Effects of white roofs on urban temperature in a global climate model. Geophys. Res. Lett. 37 https://doi.org/10.1029/2009GL042194.
Ortiz Porangaba, G.F., Teixeira, D.C.F., de Costa Trindade Amorim, M.C., da Silva, M.H.S., Dubreuil, V., 2021. Modeling the urban heat island at a winter event in Três Lagoas. Brazil. Urban Clim. 37, 100853. https://doi.org/10.1016/j.uclim.2021.100853.
Pebesma, E., 2018. Simple features for R: standardized support for spatial vector data. R J. 10, 439–446. https://doi.org/10.32614/RJ-2018-009.R package version 1.0-4.
Pebesma, E., 2020. Stars: spatiotemporal arrays. Raster Vect. Dat. Cub. https://CRAN.R-project.org/package=stars , R package version 0.5-4.

Pebesma, E.J., Bivand, R.S., 2005. Classes and methods for spatial data in R. R News 5, 9–13. https://CRAN.R-project.org/doc/Rnews/. , R package version 1.4.6.
Pham, T.T.H., Apparicio, P., Séguin, A.M., Landry, S., Gagnon, M., 2012. Spatial distribution of vegetation in montreal: an uneven distribution or environmental inequity? Landscape Urban Plann. 107, 214–224. https://doi.org/10.1016/j.landurbplan.2012.06.002.
Przybylak, R., Kejna, M., Uscka-Kowalkowska, J., Kunz, M., Maszewski, R., Arazny, A., 2015. Spatial distribution of air temperature in Torun (Central Poland) and its causes. Theor. Appl. Climatol. 127, 441–463. https://doi.org/10.1007/s00704-015-1644-2.
Core Team, R., 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
Rizwan, A., Dennis, L., Liu, C., 2008. A review on the generation, determination and mitigation of urban heat island. J. Environ. Sci. 20, 120–128. https://doi.org/10.1016/S1001-0742(08)60019-4.
Rosenfeld, A., Dorman, M., Schwartz, J., Novack, V., Just, A.C., Kloog, I., 2017. Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel. Environ. Res. 159, 297–312. https://doi.org/10.1016/j.envres.2017.08.017.
Saaroni, H., Amorim, J., Hiemstra, J., Pearlmutter, D., 2018. Urban green infrastructure as a tool for urban heat mitigation: survey of research methodologies and findings across different climatic regions. Urban Clim. 24, 94–110. https://doi.org/10.1016/j.uclim.2018.02.001.
Stewart, I.D., 2011. A systematic review and scientific critique of methodology in modern urban heat island literature. Int. J. Climatol. 31, 200–217. https://doi.org/10.1002/joc.2141.
Straub, A., Beck, C., Philipp, A., 2020. Simulation of Critical Urban Climate Load Situations in Augsburg Southern Germany, using PALM-4U. EGU General Assembly 2020, online, 4-8 May. EGU202018723.
Straub, A., Berger, K., Breitner, S., Cyrys, J., Geruschkat, U., Jacobeit, J., Kühlbach, B., Kusch, T., Philipp, A., Schneider, A., Umminger, R., Wolf, K., Beck, C., 2019. Statistical modelling of spatial patterns of the urban heat island intensity in the urban environment of Augsburg. Germany Urban Clim. 29, 100491. https://doi.org/10.1016/j.uclim.2019.100491.
Trapletti, A., Hornik, K., 2020. Tseries: Time series Analysis and Computational Finance. https://CRAN.R-project.org/pacage=tseries. , R package version 0.10-49.
Van Hove, L., Jacobs, C., Heusinkveld, B., Elbers, J., Van Driel, B., Holtslag, A., 2015. Temporal and spatial variability of urban heat island and thermal comfort within the Rotterdam agglomeration. Build. Environ. 83, 91–103. https://doi.org/10.1016/j.buildenv.2014.08.029.
Wang, M., He, G., Zhang, Z., Wang, G., Zhang, Z., Cao, X., Wu, Z., Liu, X., 2017. Comparison of spatial interpolation and regression analysis models for an estimation of monthly near surface air temperature in China. Rem. Sens. 9, 1278. https://doi.org/10.3390/rs9121278.
Wickham, H., 2007. Reshaping data with the 'reshape' package. J. Stat. Software 21, 1–20. https://doi.org/10.18637/jss.v021.i12.R package version 1.4.4.
Wickham, H., 2016. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-98141-3.R package version 3.3.5.
Wickham, H., François, R., Henry, L., Müller, K., 2021. Dplyr: A Grammar of Data Manipulation. https://CRAN.R-project.org/package=dplyr. , R package version 1.0.7.
Wilke, C., 2020. Cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. https://CRAN.R-project.org/package=cowplot. , R package version 1.1.1.
Yang, J., Zhan, Y., Xiao, X., Xia, J.C., Sun, W., Li, X., 2020. Investigating the diversity of land surface temperature characteristics in different scale cities based on local climate zones. Urban Clim. 34, 100700. https://doi.org/10.1016/j.uclim.2020.100700.
Yi, C., Kwon, H.G., Yang, H., 2022. Spatial temperature differences in local climate zones of Seoul metropolitan area during a heatwave. Urban Clim. 41, 101012. https://doi.org/10.1016/j.uclim.2021.101012.
Zeileis, A., Grothendieck, G., 2005. Zoo: S3 infrastructure for regular and irregular time series. J. Stat. Software 14, 1–27. https://doi.org/10.18637/jss.v014.i06.R package version 1.8-9.
Zhang, Y., Murray, A.T., Turner, B., 2017. Optimizing green space locations to reduce daytime and nighttime urban heat island effects in Phoenix. Arizona Landscape Urban Plann. 165, 162–171. https://doi.org/10.1016/j.landurbplan.2017.04.009.
Zhou, B., Erell, E., Hough, I., Rosenblatt, J., Just, A.C., Novack, V., Kloog, I., 2020. Estimating near-surface air temperature across Israel using a machine learning based hybrid approach. Int. J. Climatol. 40, 6106–6121. https://doi.org/10.1002/joc.6570.
Zimmerman, D.L., Stein, M., Gelfand, A., Diggle, P., 2010. Classical geostatistical methods. Handbook Spat. Stat. 29–44.