Statistical prediction of the nocturnal urban heat island intensity based on urban morphology and geographical factors - An investigation based on numerical model results for a large ensemble of French cities

Thomas Gardes, Robert Schoetter, Julia Hidalgo, Nathalie Long, Eva Marques, Valéry Masson

Version of Record: https://www.sciencedirect.com/science/article/pii/S0048969720327704
Manuscript_5271d3ae7b94c37c63bcd3622e793ce1

Statistical prediction of the nocturnal urban heat island intensity based on urban morphology and geographical factors - An investigation based on numerical model results for a large ensemble of French cities

Thomas GARDES[a)], Robert SCHOETTER[a)], Julia HIDALGO[b)], Nathalie LONG[c)], Eva Marquès[a),] Valéry MASSON[a)]

a) CNRM UMR 3589, Université Fédérale de Toulouse, Météo-France/CNRS, 42, avenue Gaspard Coriolis, 31057 Toulouse, France.

b) LISST, Université Fédérale de Toulouse – CNRS, 5, allées Antonio Machado, 31058 Toulouse, France.

c) UMR LIENSs, La Rochelle Université – CNRS, 2 rue Olympe de Gouges, 17000 La Rochelle, France.

Corresponding Author:
Thomas GARDES
42, avenue Gaspard Coriolis, 31057 Toulouse, France
+33 6 21 86 23 74
thomas.gardes@meteo.fr

Abstract

Taking into account meteorological data in urban planning increases in relevance in the context of changing climate and enhanced urbanisation. The present article focusses on the nocturnal urban heat island intensity (UHII) simulated with a physically based atmospheric model for more than 200000 Reference Spatial Units (RSU), which correspond to building patches delimited by roads or water bodies in 42 French urban agglomerations. First are investigated the statistical relationships between the UHII and six predictors: Local Climate Zone, distance to the agglomeration centre, population, distance to the coast, climatic region, and elevation differences. It is found that the maximum UHII of an agglomeration increases proportional to the logarithm of its population, decreases for cities closer than 10 km to the coast, and is shaped by the regional climate. Secondly, a Random Forest model and a regression-based model are developed to predict the UHII based on the predictors. The advantage of the regression-based model is that it is easier to understand than the *black box* Random Forest model. The Random Forest model is able to predict the UHII with less than 0.5 K absolute error for 54% of the RSU. The regression-based model performs slightly worse than the Random Forest model and predicts the UHII with less than 0.5 K absolute error for 52% of the RSU. A future challenge is to conduct a similar investigation at global scale, which is to date limited by the availability of a robust description of urban form and functioning.

Keywords: Urban heat island intensity; Urban morphology; Local Climate Zones; Regression-based models; Random Forest

1. Introduction

In the context of enhanced urbanisation (UN, 2018), and changing climate (Collins et al., 2013) it is important to quantify the influences of urban areas on the local meteorological

conditions to better support public policies. Such are the urban heat island (UHI) effect (Arnfield et al., 2003), the urban impact on moisture (Unger, 1999), precipitation (Shepherd, 2005), or the modification of the wind field by buildings and urban vegetation (Moonen et al., 2012). The UHI is characterised by higher nocturnal air temperature in urban areas compared to the surrounding rural areas. Its intensity depends on the characteristics of the urban agglomeration, geographical factors, and the prevailing meteorological conditions. Oke (1973) found that the maximum UHI intensity during clear nights with low wind speed is proportional to the logarithm of the number of inhabitants of the agglomeration. Such a finding is valuable, but more detailed information on the spatial distribution of the UHI intensity is required for the purpose of urban planning, the quantification of population exposure to the UHI, or the development of urban meteorological and climate services.

Observation based studies established that the UHI intensity depends on the degree of urbanisation with larger values in denser urban settings (e.g. Hidalgo et al., 2008). Stewart and Oke (2012) introduced the Local Climate Zones (LCZ) with the aim to provide a classification of urban morphology and functioning that is applicable at global scale and relevant for the local thermal climatic conditions. Some recent studies found that the LCZ classification is useful to discriminate the UHI intensity in different urban settings. These studies are based on station observations (Stewart et al., 2014; Alexander and Mills, 2014; Lehnert et al., 2015; Skarbit et al., 2017; Fenner et al., 2017; Beck et al., 2018), mobile measurements (Leconte et al., 2015), and numerical model simulations (Stewart et al., 2014; Verdonck et al., 2018; Kwok et al., 2019).

Geographic Information Systems (GIS) approaches have been developed in the framework of the WUDAPT project (World Urban Database and Access Portal Tools; Wang et al., 2018) to produce LCZ maps of urban areas using freely available satellite images. These can be employed to produce maps characterising the thermal climatic conditions under the

assumption that the LCZ is the main driver for the local thermal climate, e.g. Zheng et al., 2017 for Hong Kong; Kotharkar and Bagade, 2017 for Nagpur, India. The major drawback of this approach is that it cannot capture drivers of the UHI related to atmospheric dynamics like the horizontal advection of air temperature by the prevailing wind, sea breezes for cities close to the coast, katabatic flows for cities with elevation differences, and more general the prevailing regional climate.

Several previous studies employed more enhanced methods to statistically predict the UHI intensity using predictors characterising the local meteorological conditions and/or the urban morphology. Bernard et al. (2017) use wind speed and direction, cloud cover, Normalised Difference Vegetation Index (NDVI), and building density to predict the spatial and temporal UHI intensity variation with empirical models for three French cities. They find that for all seasons, wind speed and cloud cover are good predictors for the UHI intensity. The relation between NDVI (building density) and the UHI intensity is largest in the summer (autumn and winter) season. A similar study has been conducted for the Brazilian city of Paranavai (Piffer Dorigon and Amorim, 2019) based on a linear regression applied on NDVI, land use, elevation, and surface temperature. **They are able to statistically predict the spatial distribution of the UHI, with a value of the adjusted R^2 of 0.49 (summer situation) and 0.47 (winter situation).**

Ho et al. (2014) compare the ability of three statistical methods (Least Square Regression, Support Vector Machine, Random Forest) to model the UHI intensity of Vancouver (Canada) using the predictors elevation, Sky View Factor (SVF), land surface temperature, and NDVI. Makido et al. (2016) investigate the quality of three different statistical models (Ordinary Least Squares, Regressions Tree, and Random Forest) to predict the UHI intensity in Doha (Qatar) using the NDVI, building density, albedo, and distance to the coast as predictors. They find that the distance to the coast is the most important predictor. Straub et al. (2019) tested Multiple Linear Regression (MLR) and Random Forest to

statistically model the UHI intensity for the city of Augsburg (Germany) based on the predictors distance to the city centre, Sky View Factor (SVF), elevation, and land cover. The studies of Ho et al. (2014), Bernard et al. (2017), and Straub et al. (2019) are based on observations of near-surface air temperature from an urban station network, whereas Makido et al. (2016) used air temperature observations from sensors placed on vehicles. All mentioned studies find that Random Forest is the best approach for the statistical prediction of the UHI intensity, minimising global Root Mean Square Error (RMSE).

The literature survey shows that most previous studies statistically predict the UHI intensity only for one or few cities. They could therefore not systematically investigate factors like the distance to the coast, city size or regional climate. Furthermore, the studies based on observations from stations or mobile measurements are limited in the spatial coverage of the investigated urban agglomeration.

The present study is conducted in the framework of the MApUCE project (Applied Modelling and Urban Planning Law: Climate and Energy), which aims to incorporate data on urban morphology, urban climate, and building energy consumption into urban planning documents. It investigates the nocturnal UHI intensity simulated with a physically based atmospheric model at the scale of so-called Reference Spatial Units, (RSU), which correspond to building patches delimited by roads or water bodies (Plumejeaud et al., 2015). More than 200000 RSU in 42 French cities (Figure 1) of different size, morphology, geographical situation, and regional climate are investigated. The focus is on the summer season and meteorological situations which favour a strong UHI.

The main objectives of the present study are :

- The quantification of the relationships between the UHI intensity and predictors characterising the RSU (Local Climate Zone; distance to the centre of the urban agglomeration), the total population of the agglomeration, and geographical factors (distance of the agglomeration to the coast, elevation differences, French climatic region).

The relationships between the UHI intensity and the predictors are described via regression-based models.

- The construction of a Random Forest model to predict the nocturnal UHI intensity based on all predictors and to evaluate the results against those of the physically based numerical model.
- The construction and evaluation of a regression-based model to predict the UHII intensity. The regression-based model is easier to understand than the *black box* Random Forest model and easier to transfer to potential users.

Section 2 presents the data and the Random Forest model, Section 3 the regression-based models. Results are discussed in Section 4. Discussion is made in Section 5, and conclusions are drawn in Section 6.

(a) (b)

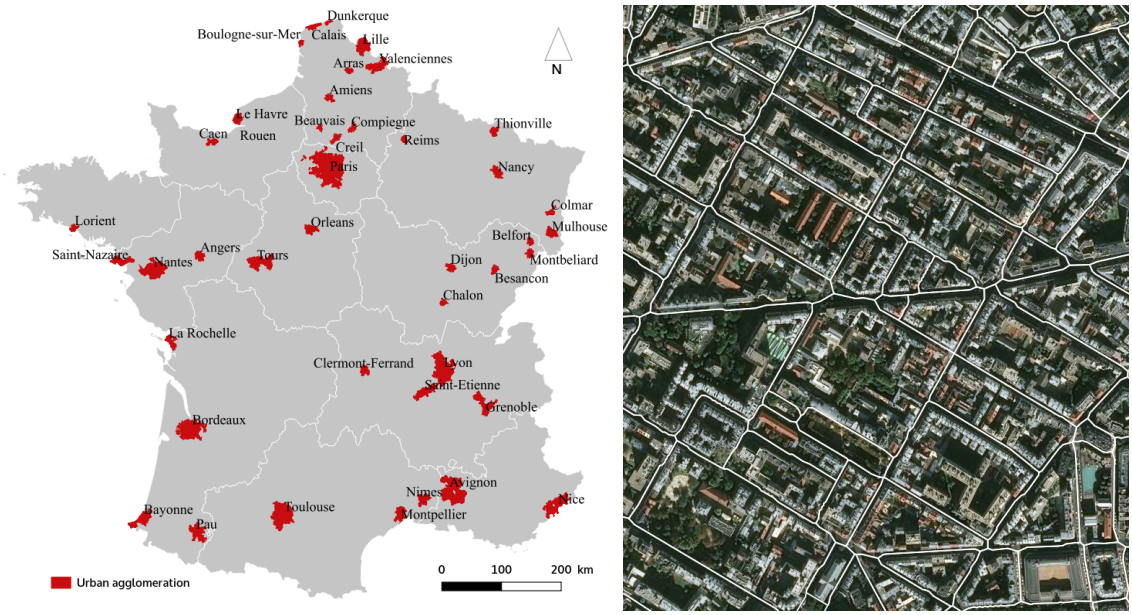


Figure 1: (a): The 42 investigated French urban agglomerations; (b): Spatial delimitation of Reference Spatial Units for the agglomeration of Paris.

1. Data

The present study relies on simulations with the mesoscale atmospheric model Meso-NH of the nocturnal urban heat island intensity for 42 French cities. This numerical modelling approach is enabled thanks to a previously constructed database on the urban morphology (Bocher et al., 2018), construction materials (Tornay et al., 2017) and building energy consumption (Schoetter et al., 2017) in France. The employed Meso-NH configuration has been evaluated by Schoetter et al. (2020) for the cities of Toulouse and Dijon, but due to the lack of high-quality observation data, it cannot be evaluated for all the cities investigated in the present study. An overarching assumption is therefore that these numerical model simulations can be used as a reference for computationally cheap statistical models for the prediction of the UHII that will be developed in Section 3. We want to stress, that despite this overarching assumption made here, more studies relying on high-quality observations are still needed to improve the monitoring and quantification of urban climate processes. This section presents the data that will serve as predictands and predictors of the statistical models developed in Section 3. A summary of the data and methodology is given in Figure 2.

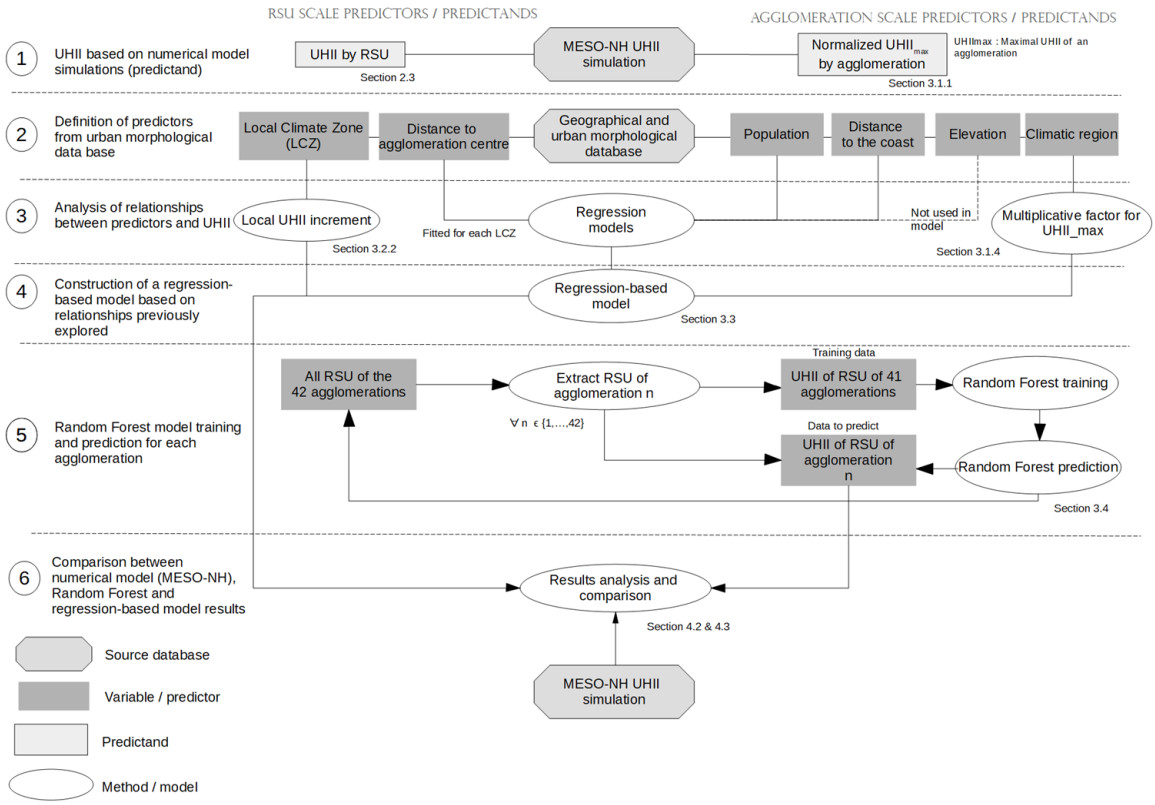


Figure 2: The datasets and the methodology employed for the statistical prediction of the nocturnal urban heat island intensity.

2.1 Local weather types favourable to a strong urban heat island effect

It is well known that the UHI intensity depends on the local meteorological conditions. It tends to be higher for lower wind speed, cloud cover, and relative humidity (Wilby, 2003; Oke et al., 2017; Hoffmann et al., 2018). This is plausible since for sunny conditions, the urban materials store more heat during the day and release it during the night. For situations with low wind speed, there is less advection to the urban area of cooler air from the adjacent rural areas. The wind direction governs the advection of the UHI and is therefore relevant for its spatial pattern. It is therefore interesting to quantify the UHI and its governing processes for different meteorological situations and seasons. Hidalgo et al.

(2014) and Hidalgo and Jougla (2018) developed a local weather type classification that can be applied to the UHI characterisation. Such a Local Weather Type approach allows to make explicit the meteorological variability of a place and allows to identify weather situations relevant for the UHI development. This is in contrast to simulate and analyse a seasonal or climatological average UHI, which is computationally more expensive and will mask all the UHI specificities linked to the weather situation itself. The local weather types (LWT) are selected based on a PAM classification (Partitioning Around Medoids) of daily values of the temperature amplitude, specific humidity, precipitation, wind speed and direction. Jougla and Hidalgo (2019) applied this method to 50 French cities using a 2.5 km resolution re-analysis with the Météo-France model AROME (Seity et al., 2011) for the period 2000 to 2009. The number of LWT is about 10, but differs slightly for each agglomeration. In the present study, only the summer season is investigated and for each urban agglomeration a LWT favourable to the development of a strong UHI is selected. Schoetter et al. (2020) show that not one single day, but at least 3 to 6 days should be simulated to quantify the UHI pattern for a urban given agglomeration and LWT. Therefore, 6 days are simulated for the selected LWT.

2.2 Homogeneous data on urban form and function

The database on urban form and function has been compiled in the framework of the MApUCE project (http://mapuce.orbisgis.org/). Urban agglomerations are defined following the French Institute on Economics and Statistics (INSEE) definition of "a municipality or a group of municipalities of at least 2000 inhabitants which includes a continuously built-up zone where constructions are not more than 200 m apart"[1]. Urban morphological indicators, calculated on regular grids are usually used to describe the urban fabric in meteorological models (Ching et al., 2009).

However, this method to partition the urban fabric does not respect the irregular structure

[1] https://www.insee.fr/en/metadonnees/definition/c1501

of cities, especially French cities, since it simplifies its complex shapes. According to Berghauser-Pont and Haupt (2005), buildings can be considered as an elementary object of the urban fabric but the building scale is inappropriate for meteorological models. An aggregation of buildings, called building block, allows respecting the *natural* delimitation of the city structure and representing a well-defined geographical entity. This scale is considered as a Reference Spatial Unit (RSU, Figure 1b), which is delimited by the building block's boundaries like roads, rail tracks, large water bodies, urban parks or rural areas (Bocher et al., 2018, Masson et al, 2020). The following centralised data on building outlines, construction practices, demography, and household equipment have been used.

- The BD Topo® digital basic map provided by the French Geographical Institute (IGN) (http://professionnels.ign.fr/bdtopo) includes information on building outlines and height with a precision of 1 m as well as on building use (industrial, commercial, educational, administration, and so on). Bocher et al. (2018) developed a geoprocessing chain to compute indicators on urban morphology (e.g. building surface fraction, mean building height) at the RSU scale based on the 2014 version of the BD TOPO, the effective resolution of the urban morphology dataset is around 100 m in the city centres and around 250 m in suburban areas.

- Bibliographical references on building construction practices in different French regions have been used by Tornay et al. (2017) to provide a detailed description of building construction materials for characteristic buildings at RSU scale taking into account the temporal evolution of building construction practices from vernacular buildings to the most recent construction (after 2013).

- The INSEE compiles the census of the French population (https://www.insee.fr/fr/information/2008354; ~20 million individuals), which contains information on demography (e.g. age of inhabitants), household characteristics (e.g. the type of the heating system). Bourgeois et al. (2017) developed statistical models to predict

indicators related to building energy consumption based on the complete census valid in 2011 To protect privacy, the census data are aggregated over about 2000 individuals, which leads to an effective resolution of about 500 m in the city centres and a coarser resolution in suburban areas.

2.3 Numerical simulations of the nocturnal urban heat island intensity

For the selected days, numerical simulations are performed with the mesoscale atmospheric model Meso-NH (Lac et al., 2018). Meso-NH is applied in hindcast mode and used to dynamically downscale the ECMWF-IFS[2] high-resolution operational forecast analysis via three intermediate nesting steps to a horizontal resolution of 250 m. The horizontal grid resolution is 8 km (D1), 2 km (D2), 1 km (D3), and 250 m (D4). The horizontal extent of the domains is at least 1000 km x 1000 km for D1, 300 km x 300 km for D2, 150 km x 150 km for D3, and 50 km x 50 km for D4. However, the domain extents and positions are slightly adapted for each agglomeration to take into account the local topography or the spatial extent of the agglomeration. For urban areas close to the coast/mountains, D4 is enlarged to include a sufficiently large part of the sea/mountainous area. Furthermore, a larger D4 has been chosen for some very extended cities like Paris to cover them entirely. The employed model configuration including the grid nesting and the physical parametrisations is identical to the one described in Section 3 and Table 2 of Kwok et al. (2019). Model output is hourly.

Meso-NH is coupled with the urban canopy parametrisation Town Energy Balance (TEB; Masson, 2000), which solves the urban surface energy balance as a function of the meteorological conditions simulated by Meso-NH. TEB assumes a simplified urban morphology with buildings aligned along street canyons and solves the surface energy budget of a representative roof, wall and road to take into account their different physical

[2] European Centre for Medium-Range Weather Forecasts Integrated Forecasting System

properties, orientations, and positions in the urban canopy layer. In-canyon urban vegetation is taken into account with the approach of Lemonsu et al. (2012), the Building Energy Model (BEM; Bueno et al., 2012; Pigeon et al., 2014; Schoetter et al., 2017) is employed to solve the energy budget of a representative building at district scale by taking into account the characteristics of the building envelope, building use, and practices related to heating and air conditioning. The building energy model simulates the anthropogenic heat flux due to the buildings as a function of the prevailing meteorological conditions. The anthropogenic heat flux due to traffic and industrial activities is neglected in the present study since we do not possess detailed maps of these fluxes. The traffic heat fluxes are usually lower than the building heat fluxes in French cities (Pigeon et al., 2007), and will have a relevant influence only at grid points with major road networks. Industrial facilities might exhibit large anthropogenic heat fluxes, but they are usually not located inside the urban agglomerations, but rather at the outskirts.

The Surface Boundary Layer (SBL) scheme of Hamdi and Masson (2008) is employed to calculate vertical profiles of meteorological parameters in the urban canopy layer. TEB is part of the Externalised Surface (SURFEX; Masson et al., 2013), which considers four different surface cover types as tiles within a single grid, namely urban areas, rural areas, oceans, and lakes. The input parameters for TEB describing the urban morphology are directly taken from the MApUCE database; they can be visualised on http://mapuce.orbisgis.org/. None of the model parameters is initialised based on an LCZ map, the model results are therefore not directly influenced by the LCZ. The land cover maps and physical parameters for the rural areas are taken from the 1 km resolution ECOCLIMAP-I database (Masson et al., 2003). Model results for these rural areas will not directly influence the results of the present study, since only model results in the urban areas will be analysed. The simulated values of air temperature at 2 m above ground (*T2M*) in the urban environment are taken from the second level of the TEB SBL scheme, which is

placed exactly at 2 m above ground. A similar approach is made for the other tiles of SURFEX.

In the present study, the objective is to attribute to which degree the urbanisation modifies the local near-surface air temperature and to compare these influences for cities with different morphology and a different geographical situation. For this reason, two simulations are conducted to quantify the local influence of an urban agglomeration on *T2M*. One for the reference surface cover (*refer*) and one no-urban simulation (*nourb*) for which all urban land use is replaced by a type of cropland that frequently occurs in the surroundings of this agglomeration. This approach is reasonable for France, since most cities are mainly surrounded by cropland. The main reason for using a *refer* and a *nourb* simulation instead of only one *refer* simulation is that for some cities it is not easy to define the rural grid cells. They might differ from the urban grid cells in elevation, distance to the coast, or other factors not directly representing the degree of urbanisation. This might then bias the calculated UHII and make it more difficult to compare in between cities. There are two major drawbacks from the use of a *nourb* simulation to quantify the UHI. First, with this approach, the UHI is defined locally based on the difference between two numerical simulations instead of using the classical definition of the UHI intensity as air temperature difference between an urban and an adjacent rural area. Secondly, the UHI slightly depends on the somewhat artificial choice of the land cover type that replaces the urban land cover. A study investigating the physical processes governing the UHI of one given urban agglomeration should therefore be based on only one reference simulation representing the actual land cover.

The local UHI intensity (*UHII*) at 2 m above ground is defined following Equation (1).

$$UHII(x, d, h) = T2M \qquad (1)$$

In Equation (1), *x* denotes the space coordinate, *d* a day, and *h* the hour of the day. Here the temporal average *UHII* is calculated for the time period 4 to 6 local time, which

corresponds to assume that the maximum UHI development occurs during this period for all cities. Previous analysis of the nocturnal urban heat island of Paris (Lemonsu and Masson, 2002) and Toulouse (Hidalgo et al., 2008) show that this is a good assumption for these cities, but it might not be the case for all the cities. The average *UHII* is calculated for the days following the 6 days with the selected LWT (Equation 2), since the nocturnal UHI is strongly shaped by the meteorological conditions of the preceding day (Hoffmann et al., 2012).

$$-\sum \qquad \sum \qquad)(2)$$

In Equation (2), N_t corresponds to the number of simulated days per LWT (6), multiplied by the number of hourly model outputs (3) in the selected time period (N_t = 6 x 3 = 18).

The present study is entirely based on maps of the average nocturnal *UHII*. The temporal evolution of the *UHII* is not investigated.

The simulated *UHII* calculated using the model output for D4 is available on a regular grid with 250 m horizontal resolution. To calculate the average *UHII* at RSU scale, the weighted average of the *UHII* values simulated for all model grid points, which spatially intersect an RSU is calculated (Equation 3). The weights are the areas *A* of intersection between the RSU and the model grid points.

$$\frac{\Sigma_x}{\Sigma_x} \qquad (3)$$

Due to the lack of high-quality observations it is not possible to evaluate the simulated *UHII* for all cities. High-quality long-term dense observations are available for two of the simulated cities, Toulouse (CAPITOUL campaign; Masson et al., 2008; March 1 2004 to February 28 2005) and Dijon (MUSTARDijon campaign; Richard et al., 2018; since June 2014) although not for the time period 2000 to 2009, which has been used for the selection of the LWT. Schoetter et al. (2020) evaluate the simulated *UHII* of these two cities for different seasons and LWT using the same Meso-NH-TEB configuration than in the present

study. For Toulouse, the seasonal average nocturnal *UHII* in June, July and August (JJA) is slightly overestimated (1.9 K simulated instead of 1.6 K observed), for the LWT with the highest average *UHII*, the simulated *UHII* is also slightly overestimated (2.2 K simulated instead of 2.0 K observed). For Dijon, the seasonal average nocturnal *UHII* in JJA is simulated well (simulated *UHII* 1.5 K, observed *UHII* 1.6 K), however the simulated *UHII* for the LWT most favourable to a strong UHI is underestimated by 0.7 K (simulated 1.9 K instead of observed 2.6 K). The evaluation for Dijon shows that the simulated *UHII* for a given agglomeration and LWT can be biased. In the following, the simulated *UHII* will be used to analyse the relationship between the *UHII* and predictors describing urban morphology, geography, and so on. These statistical relationships will be derived for a large sample of cities and RSU which means that non-systematic errors of the simulated *UHII* will cancel. Furthermore, the simulated *UHII* will be used as reference for statistical models that predict the *UHII*. This is justified since the atmospheric numerical model takes into account physical processes like advection, turbulence in the planetary boundary layer, and TEB is also a physically based model describing the urban surface energy balance.

2.4 Predictors of the urban heat island intensity

The present study investigates the relationships between the *UHII* and the six predictors presented in Table 1.

Predictor	**Scale**

Total population of the urban agglomeration	Urban agglomeration
Distance of the urban agglomeration to the coast	Urban agglomeration
French climatic region	Urban agglomeration
Elevation differences in and around the urban agglomeration	Urban agglomeration
Local Climate Zone of RSU	Reference Spatial Unit
Distance of the RSU to the urban agglomeration centre	Reference Spatial Unit

Table 1: The six predictors of the UHII.

The centre of each urban agglomeration needs to be defined. For this purpose, the outlines of the so-called IRIS defined by INSEE as 'aggregated units for statistical information' (https://www.insee.fr/en/metadonnees/definition/c1523) are taken. The IRIS cover areas with a nighttime residential population of about 2000 inhabitants. The agglomeration centre is defined as the centre of the IRIS with the largest building surface fraction. With this definition, it is not possible that an industrial or commercial area with high building surface fraction at the outskirts of an urban agglomeration is selected as agglomeration centre, since the nighttime residential population is low in such areas and the IRIS will also cover adjacent low density residential areas. Manual adjustments have been made for 9 agglomerations for which the automatically determined centre did not match well with the actual centre. Furthermore, 4 agglomerations are characterised by a distinct sub-centre with a population of at least 30000. A second centre has been added for these 4 agglomerations.

The six predictors of the *UHII* are calculated as follows.

- The total population of an urban agglomeration (*Pop*) is calculated based on the population at RSU scale (Equation 4), which has been derived from the 2015 INSEE census of the French population (https://www.insee.fr/fr/information/3561862).

Σ)(4)

- The distance of an urban agglomeration to the coast (*DistCoast*) is calculated using OpenStreetMap data on the French coastline (https://osmdata.openstreetmap.de/data/coastlines.html, 09/05/2019). A point is placed each 5~m along the coast line. Then, the distance between the agglomeration centre and the nearest coast point is computed.

- The French climatic region is taken from the classification of Joly et al. (2010). They define eight climatic regions in France: Pure, Altered and Degraded Oceanic Climate; Pure and Altered Mediterranean Climate; Semi-Continental climate; Mountain climate; and the South-West (of France) Basin climate. All of these climatic regions are represented in the sample of cities. However, only Toulouse (Saint-Nazaire) is located in the South-West Basin (Degraded Mediterranean) climatic region. To avoid too low sample size, the climatic region attribute is changed to Altered Oceanic (Pure Oceanic) for Toulouse (Saint-Nazaire), which are the most similar to the actual regional climatic zones.

- The elevation difference for a given urban agglomeration (*ElDiff*) is defined as the difference between the highest and lowest elevation in a buffer of 10 km around the urban agglomeration (EL_{10km}) (Equation 5). Data on elevation is taken from the BD ALTI® 75~m raster data provided by IGN (http://professionnels.ign.fr/bdalti). The *ElDiff* predictor indicates whether there are hills or mountains inside or surrounding the urban agglomeration. These can be responsible for channelling of the wind or katabatic flows, which can influence the UHII.

)(5)

- The Local Climate Zones (LCZ; Stewart and Oke, 2012) characterising the urban morphology at RSU scale are determined as described in Hidalgo et al. (2019) via a semi-automatic classification of the parameters on urban morphology in the dataset produced by

Bocher et al. (2018) and the dominant urban typology at RSU scale defined by Tornay et al. (2017). Hidalgo et al. (2019) determine thresholds of building height and building density more adapted to French cities than the parameter ranges included in the original LCZ classification by Stewart and Oke (2012). Due to the shortcomings of the available input datasets it is not possible to identify LCZ 10 (heavy industry) or to distinguish different rural vegetation types. All rural vegetation is therefore represented as LCZ D (low plants) in the statistical analysis, whereas in the numerical model simulations it is described via the ECOCLIMAP-I database, which distinguishes rural land cover types like forests, cropland, grassland, and so on.

- The distance of an RSU to the centre of the urban agglomeration (*DistCentre)* is defined as the distance between the RSU centre and the urban agglomeration centre.

3. Statistical modelling of the urban heat island intensity

Regressions between the *UHII* and each single predictor are derived with the objective to eliminate the influence of the other predictors as far as possible. A separation is made between the predictors at the urban agglomeration scale, which are used to statistically predict the maximum *UHII* of the agglomeration ($UHII_{max}$, Equation 6) and those at RSU scale which are used to predict the *UHII* at RSU scale as a function of $UHII_{max}$.

)(6)

3.1 Predictors at the urban agglomeration scale

3.1.1 Total population

Linear regression models have been tested between $UHII_{max}$ and the total population (*Pop*) or the natural logarithm of *Pop*. The highest values of the adjusted R^2 are found for the natural logarithm of *Pop* (Equation 7). The linear model has been tested for all agglomerations and secondly separately for each French climatic region. Results will be discussed in Section 4.

)(7)

The residuals of the regressions are denoted withϵ.

For the analysis of the remaining three predictors at urban agglomeration scale, the $UHII_{max}$ is corrected for the effect of the population using the relation in Equation (7) to remove spurious correlations that might be introduced due to the strong influence of the total population on $UHII_{max}$. The $UHII_{max}$ *corrected for the influence of population (*$UHIIC_{max}$*)* is the $UHII_{max}$ a given agglomeration would, on average, experience if its population would correspond to the median of the population values in the sample (*MedPop* = 202484 inhabitants). It is calculated following Equation (8).

———)(8)

3.1.2 Distance of the urban agglomeration to the coast

Linear regression models have been adjusted to relate $UHIIC_{max}$ and *DistCoast* or the natural logarithm of *DistCoast.* The selection of the tested regression models has been made based on visual inspection of the data. The best performing relationship measured by the adjusted R^2 is given in Equation (9).

)(9)

3.1.3 Elevation differences

Linear regression models have been adjusted to relate $UHIIC_{max}$ and *ElDiff* as well as the natural logarithm of *ElDiff*. The selection of the tested regression models has been made based on visual inspection of the data. The values of the adjusted R^2 are low for all of the tested regression models, and the p-values indicate that the relationships are not statistically significant.

3.1.4 French climatic region

The impact of the French climatic region (*CR*) on $UHIIC_{max}$ is quantified by calculating a multiplicative factor ($\Delta UHIIC_{max}^{CR}$) between the $UHIIC_{max}$ averaged per climatic region and the average $UHIIC_{max}$ (Equations 10, 11, 12)

$$\frac{\overline{UHIIC_{max}}^{CR}}{\overline{UHIIC_{max}}} \quad (10)$$

$$\overline{UHIIC_{max}} \quad \text{—} \sum \quad)(11)$$

$$\overline{UHIIC_{max}} \quad \text{———} \sum \quad)(12)$$

N_{UA} is the number of urban agglomerations.

3.2 Predictors of the urban heat island intensity at Reference Spatial Unit scale

3.2.1 Distance to the centre of the urban agglomeration

The relationship between the *UHII* and the distance to the centre of the urban agglomeration is quantified using *UHII* values that are normalised by $UHII_{max}$ (*UHIIN*) to reduce the influence of different absolute values of the *UHII* for different agglomeration sizes or geographical situations (Equation 13).

——————(13)

To quantify the relationship between *UHIIN and DistCentre*, the effects of different spatial

extents of different urban agglomerations as well as the different morphology need to be filtered. The characteristic radius (*R*) of an urban agglomeration is defined as the radius of a circle with an area corresponding to the sum of the areas of all RSU with an urban LCZ (1 to 9, but excluding LCZ 7 "lightweight low-rise") (Equation 14).

$$\sqrt{\frac{\Sigma_R}{\qquad\qquad}} \quad (14)$$

For a given urban agglomeration, the distance of each RSU to the agglomeration centre is normalised by *R* (Equation 15).

$$\frac{\qquad\qquad}{\qquad\qquad} \quad (15)$$

Specific LCZ occur at systematically different frequencies in a different distance to the agglomeration centre. For example LCZ 2 (dense mid-rise) is more frequent in the centre than at the outskirts of French cities. This could severely bias the regression since a different value of *UHIIN* is expected for different LCZ. For this reason, the relationship between *UHIIN* and *DISTN* is adjusted separately for each LCZ. Different regression models have been tested including linear regression adjusted on *DISTN* and its logarithm. Based on these tests, an exponential relationship matches best (Equation 16).

(16)

3.2.2 Local Climate Zone

To quantify the impact of the LCZ on *UHII*, a 2 km x 2 km raster is defined for each agglomeration. For every point of this raster, the average *UHII* is calculated (Equation 17).

$$\overline{UHII} \qquad \frac{\Sigma_R}{\Sigma_R} \quad (17)$$

Furthermore, the RSU with the same LCZ in the raster are grouped and the average *UHII* per LCZ is calculated (Equation 18).

$$\overline{UHII} \quad \frac{\sum_R}{\sum_R} \quad (18)$$

The local increment of the *UHII* due to the LCZ ($\Delta UHII^{LCZ}$) is then defined following Equation (19).

$$\text{———}\sum \quad \overline{UHII} \quad \overline{UHII} \quad (19)$$

This local increment quantifies how much the LCZ alter the local thermal environment. The use of the 2 km x 2 km raster aims to filter the potential influences of the size of the agglomeration, the distance to the centre, or the climatic region. A comparison has been made to investigate whether the size of the raster has a relevant influence on the results by repeating the analysis using a 1 km x 1 km and a 4 km x 4 km raster. Results do not change in a relevant manner (not shown).

3.3 Combination of regressions models

The relationships between the *UHII* and the different predictors derived in the Sections 3.1 and 3.2 are combined to obtain an intuitive formula to predict the *UHII* at RSU scale based on all the predictors. First, the regression-based relationships derived in Section 3.1 are combined to obtain a formula to predict $UHII_{max}$ of an agglomeration based on the predictors at agglomeration scale (Equation 20). The elevation difference predictor is excluded, since no statistically significant relationship between $UHIIC_{max}$ and this predictor is found (Section 4.1.5). The influence of the climatic region and the distance to the coast are included as a multiplicative factor in Equation (20), which is coherent with their definition in Sections 3.1.2 and 3.1.4. The distance to the coast is only considered for agglomerations closer than the maximum distance for which the ocean impacts the *UHII* ($DIST^{Max}_{Coast}$, Section 4.1.4), and a normalisation is made to enforce the continuity of the equation. Secondly, the relationships to predict the *UHII* at RSU scale based on $UHII_{max}$ derived in Section 3.2 are combined to yield Equation (21). The combination of Equations

(20) and (21) gives the regression-based model to predict the *UHII* at RSU scale based on all the predictors (Equation 22).

(20)

(21)

(22)

3.4 Modelling of the UHI intensity with Random Forest models

Random Forest models are employed to model the *UHII* at RSU scale using the predictors described in Section 2.4. The Random Forest algorithm, first introduced by Breiman (2001), is a predictive model based on bootstrap method applied to Classification And Regression Trees (CART). It consists in building a forest of *n* CART with *n* random

samples with replacement and to consider the majority vote of all trees as a prediction tool. In addition to the bootstrap, Random Forest is improved by subsampling the predictors used to build each CART. This way, it brings more independence between the trees because they are not all built with the same predictors so that the variance of prediction is reduced. Random Forest models allow to combine both categorical and continuous predictors. First tests showed that randomly splitting the 217162 RSU into a training dataset consisting of 70% of the RSU and a validation dataset (30% of the RSU) leads to optimistic bias because the RSU from the same urban agglomeration are used for both training and validation. To avoid such overfitting, we choose to train a Random Forest model for each urban agglomeration. For a given agglomeration n, all RSU located in this specific agglomeration are excluded from the training dataset. The trained Random Forest model is then applied only to agglomeration n. This methodology is applied to each of the 42 agglomerations. The main metric used to quantify the quality of the Random Forest models is the percentage of RSU for which the absolute value of the prediction error is lower than a defined threshold (0.2 K; 0.5 K; 0.8 K; and 1 K). Furthermore, the Random Forest model predictions of the *UHII* are mapped for cities of different sizes and geographical situations and compared with those of the physically based Meso-NH-TEB model.

4. Results

In Section 4.1, the relationships between the *UHII* and the different predictors are presented. The *UHII* predicted by the Random Forest (combined regression-based) model is evaluated against the physically based numerical model in Section 4.2 (Section 4.3).

4.1 Regressions between the urban heat island intensity and the predictors

4.1.1 Urban heat island increment attributed to LCZ

The values of the local increment of the *UHII* due to the LCZ ($\Delta UHII^{LCZ}$) are given in Table 2. The results are consistent with the characteristics of the LCZ. The LCZ with high

building density and height display the largest positive values of $\Delta UHII^{LCZ}$, the LCZ sparsely built, low plants, and water display a negative value. The only counter-intuitive result is that LCZ 1 (compact high-rise) displays a lower value of $\Delta UHII^{LCZ}$ than the other compact LCZ (2 and 3). This might be due to the fact that in French cities compact high-rise LCZ are usually located outside the historical dense mid-rise city centres and characterised by a lower building surface fraction than compact mid-rise or compact low-rise LCZ, since the classification of Hidalgo et al. (2019) used mainly building height to attribute the high-rise LCZ.

LCZ	Name	$\Delta UHII^{LCZ}$ [K]
1	compact high-rise	0.17
2	compact mid-rise	0.33
3	compact low-rise	0.35
4	open high-rise	0.12
5	open mid-rise	0.16
6	open low-rise	0.15
7	lightweight low-rise	-0.06
8	large low-rise	-0.02
9	sparsely built	-0.05
D	low plants	-0.08
E	bare rock or paved	0.09
G	water	-0.14

Table 2: Local increment of the urban heat island intensity due to the Local Climate Zones (LCZ) averaged for the 42 French urban agglomerations.

4.1.2 Distance to the centre of the urban agglomeration

The relationships between the normalised *UHII* at RSU scale and the normalised distance to the city centre are displayed in Figure 3 for the different LCZ. The normalised *UHII* is, on average, highest close to the city centre and decreases towards very low values for normalised distances of about twice the effective radius of the agglomeration. However, a large scattering around the average is found for the individual RSU. This is due to the large

variety of agglomerations shapes, which can strongly differ from a circle, or the presence of sub-centres within a given urban agglomeration. We conclude that the distance to the city centre is a predictor that cannot be neglected, but is difficult to capture with a simple statistical relationship.

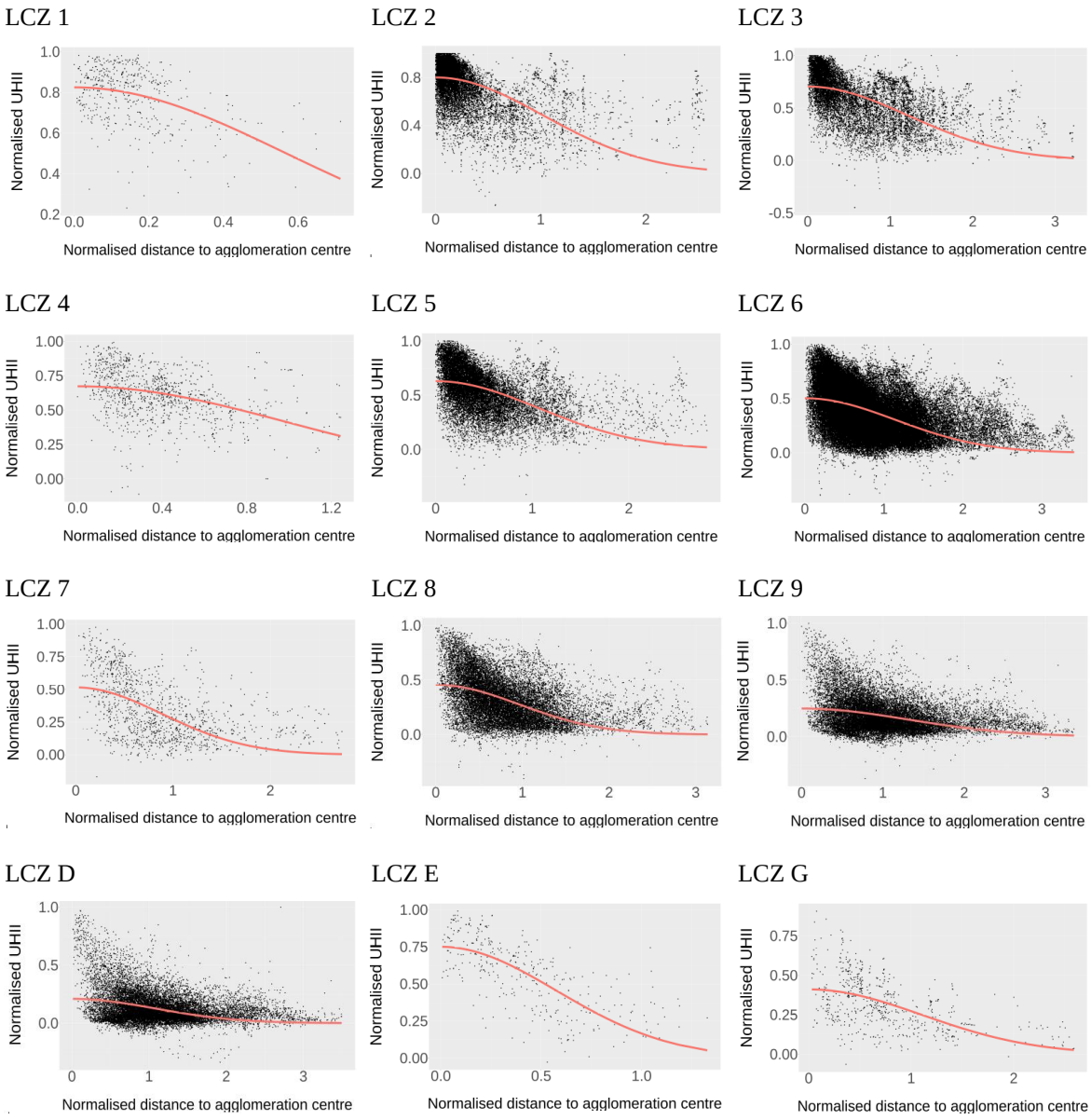


Figure 3: Relationship between the normalised urban heat island intensity and the normalised distance to the city centre, for each Local Climate Zone (LCZ). The red lines display the adjusted exponential model.

4.1.3 Total population of the urban agglomeration

The linear relation between $UHII_{max}$ and the natural logarithm of the total population of the agglomeration (Figure 4) has different values of the explained variance for different French climatic regions. Adjusting the relation for all agglomerations, gives the relation in Equation (23) with a value of the adjusted R^2 of 0.29. The p-value of 0.0002 indicates that this relationship is statistically significant. The statistical relationship is only valid for French urban agglomerations with a population between 50000 and 10 million.

7(23)

The statistical relationship between $UHII_{max}$ and the natural logarithm of *Pop* is less pronounced for the climatic regions Pure Mediterranean (Med.-Pure) and Pure Oceanic (Oce.-Pure) than for the other climatic regions; the value of the adjusted R^2 is only 0.11. The statistical relationship is stronger for the cities in the other climatic regions (adjusted R^2 of 0.39). However, the relationship in Equation (23) is used for all climatic regions for the combined regression-based model (Section 4.3), since the best performance is achieved with this choice.

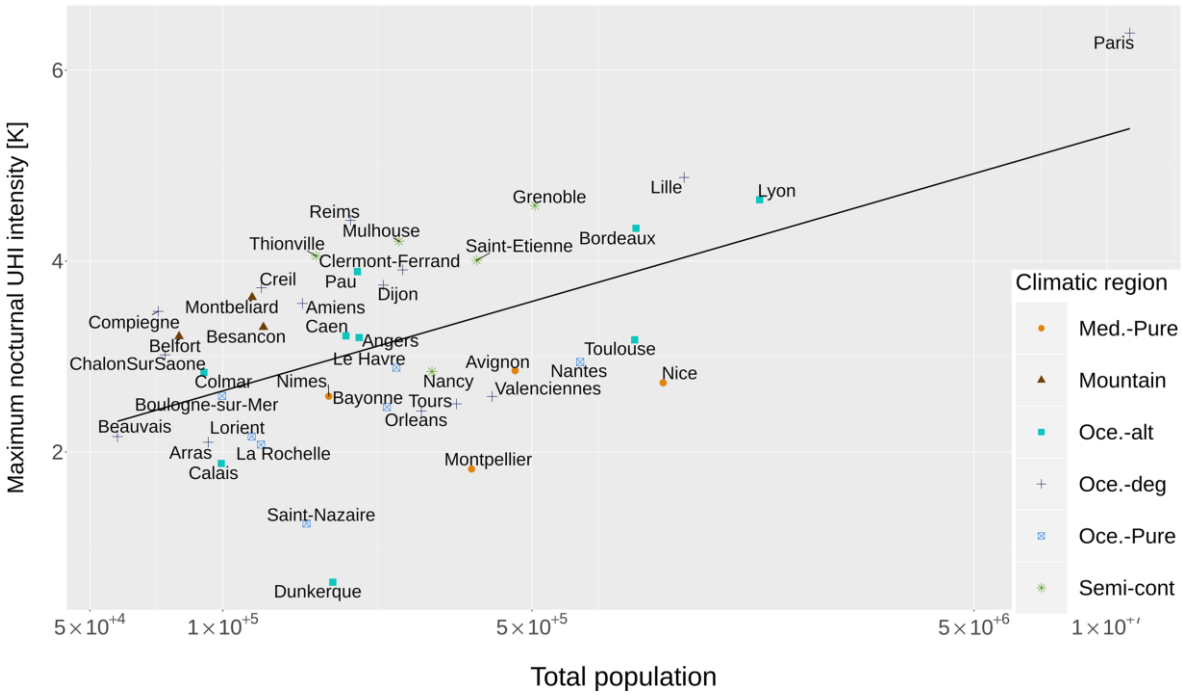


Figure 4: Relationship between the maximum nocturnal urban heat island intensity and the total population of an urban agglomeration. Med.-Pure (Oce.-Pure) is the Pure Mediterranean (Oceanic) climatic region, Oce.-deg (Oce.-alt) the Degraded (Altered) Oceanic climatic region.

4.1.4 Distance of the urban agglomeration to the coast

The relationship between $UHIIC_{max}$ and the distance of the urban agglomeration to the coast is only pronounced if the distance to the coast is less than 10 km (Figure 4). For this reason, a linear regression between $UHIIC_{max}$ and the natural logarithm of *DistCoast* is adjusted for agglomerations whose centre is less than 10 km from the coast. The logarithmic model performs best and yields an adjusted R^2 of 0.32 (p-value = 0.05) (Equation 24).

$$km \quad (24)$$

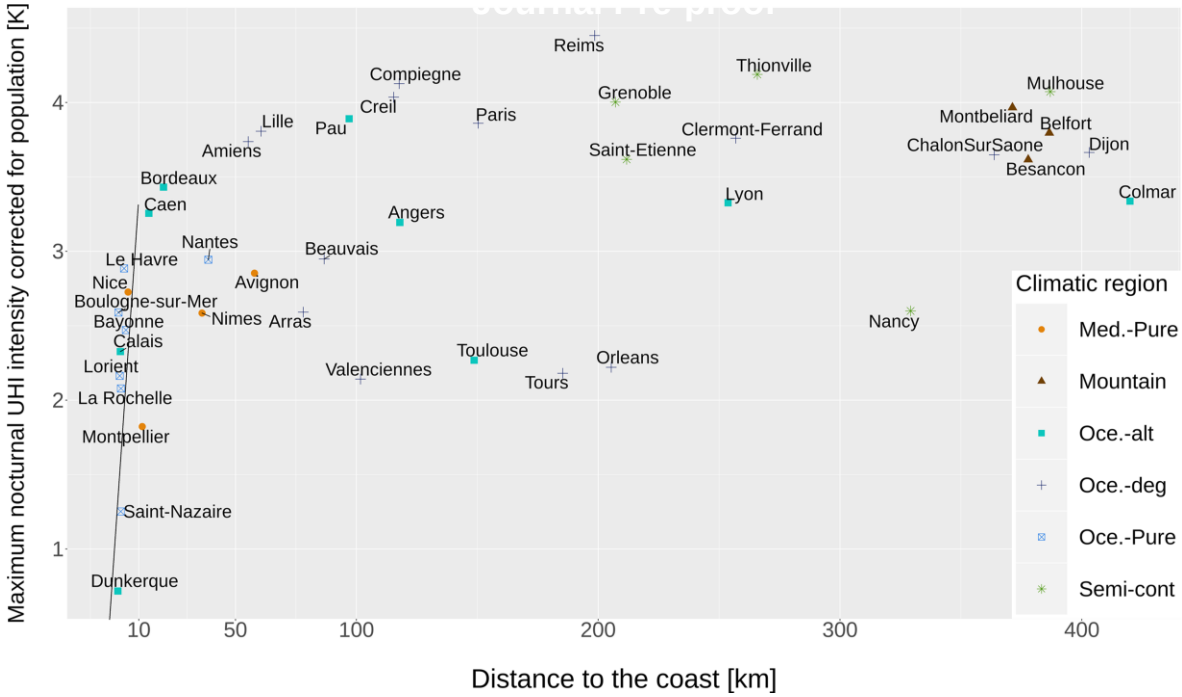


Figure 5: Relationship between the maximum nocturnal urban heat island intensity corrected for the population of the agglomeration and the distance of the centre of the urban agglomeration to the coast.

4.1.5 Elevation differences

Linear regression models have been adjusted between $UHIIC_{max}$ and *ElDiff* as well as the natural logarithm of *ElDiff*. However, no statistically significant relationship has been found for the investigated regression models. As a result, elevation differences are not considered as predictor for the Random Forest model (Section 4.2) and the regression-based model (Section 4.3).

4.1.6 French climatic region

The multiplicative effect of the French climatic region on $UHIIC_{max}$ is given in Table 3. $UHIIC_{max}$ is on average highest in the Semi-continental and Mountain climatic regions, and

lowest in the Pure Oceanic and Pure Mediterranean climatic regions. These results are plausible, since wind speed is on average higher in the climatic regions closer to the coast, for example due to the presence of sea breezes. Soil moisture can also be higher, dampening the cooling of the rural areas at night.

To quantify the relevance of the French climatic region for $UHIIC_{max}$, the explained variance (*EV*) due to the climatic region is computed following Equation (25). The value of *EV* of 0.25 indicates that the climatic region influences $UHIIC_{max}$ to a slightly smaller degree than the total population influences the $UHII_{max}$.

$$\frac{RC \sum_{UA} \quad -\overline{UHIIC^{RC}})}{\sum_{UA} \quad -\overline{UHIIC_{max}})} \qquad 25(25)$$

French regional climatic region	
Degraded Oceanic	1.29
Altered Oceanic	0.76
Pure Oceanic	0.26
Pure Mediterranean	0.42
Mountain	1.72
Semi-continental	1.62

Table 3: Influence of the French climatic region on the maximum nocturnal urban heat island intensity.

4.2 Random Forest model prediction of the urban heat island intensity

The *UHII* at RSU scale is statistically predicted with Random Forest models based on the five predictors population, climatic region, distance to the coast, distance to the agglomeration centre, and LCZ. A different Random Forest model is trained for each urban agglomeration, by excluding all the RSU of the concerned agglomeration from the training

dataset. Tests with different Random Forest model configurations characterised by the number of trees (ntree; tested from 30 to 500) and the number of predictors used at each node (mtry, tested from 2 to 4), show that these parameters do not strongly influence the results in terms of prediction quality. As final configuration of the Random Forest model we set ntree = 30 and mtry = 3, which improves computation time without deterioration of the results.

The relative weight of the different predictors used by the Random Forest model cannot be assessed using the R^2 like for the linear regression models. Instead, the Random Forest algorithm provides the *node purity*, which allows to assess which predictors are the most relevant for the classification (Table 4). The average *node purity* for the 42 trained Random Forest models indicates that the total population is the most important predictor, followed by the distance of the RSU to the agglomeration centre and the Local Climate Zone. These results appear to be plausible when compared to the results of the relationships established via the linear regression models (Section 4.1).

Predictor	**Node Purity**
Total population of the urban agglomeration	121059
Distance of the RSU to the urban agglomeration centre	74195
Local Climate Zone of RSU	71478
Distance of the urban agglomeration to the coast	54253
French climatic region	25502

Table 4: Average value of the node purity for each predictor used by the Random Forest models.

The total RMSE for all RSU is 0.85 K. The RMSE per urban agglomeration varies between 0.25 and 1.69 K. Paris is the agglomeration with the highest RMSE (1.69 K). Without

Paris, the RMSE for all remaining RSU decreases to 0.61 K. The RMSE for all RSU averaged for a specific agglomeration increases with the value of the agglomeration (Figure 6). The Random Forest model performs particularly poorly for Nice, which might be due to its specific local climatic conditions with both land/sea and mountain breezes, which does not occur for the other cities.

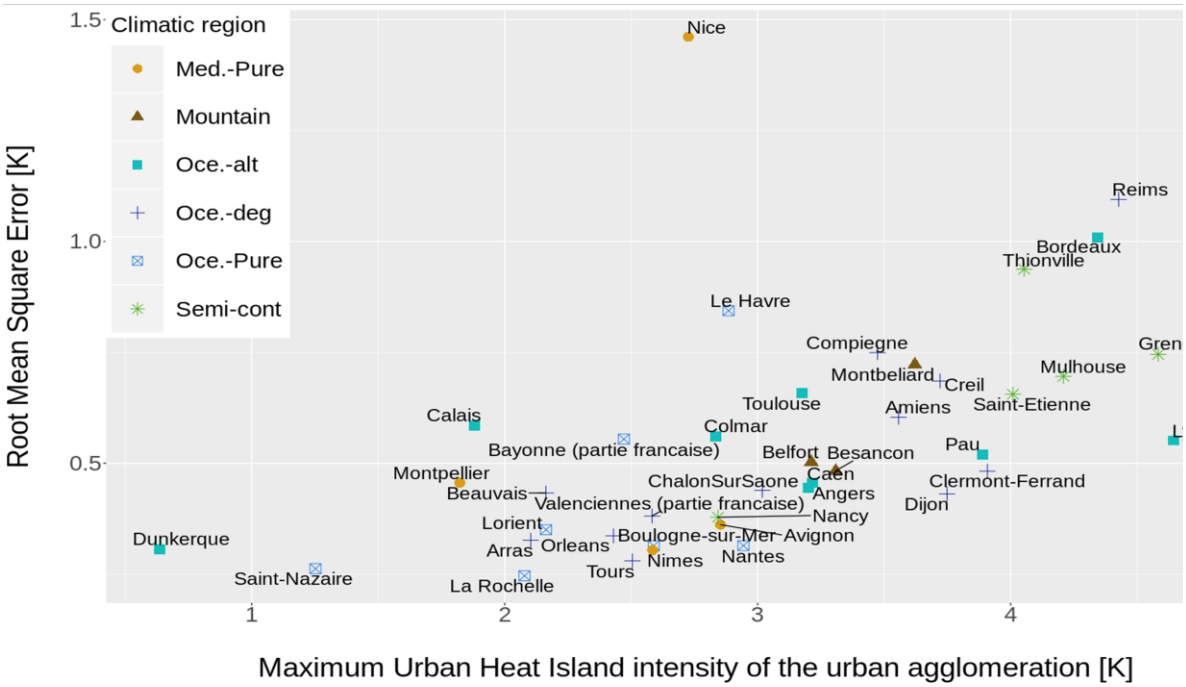


Figure 6 : Relationship between the root mean square error of the Random Forest model prediction and the maximum nocturnal urban heat island intensity of an urban agglomeration..

Table 5 displays the percentage of all RSU for which the Random Forest model predicts the UHII with a given range of the absolute error. These values will be compared with those of the regression-based model in Section 4.3.

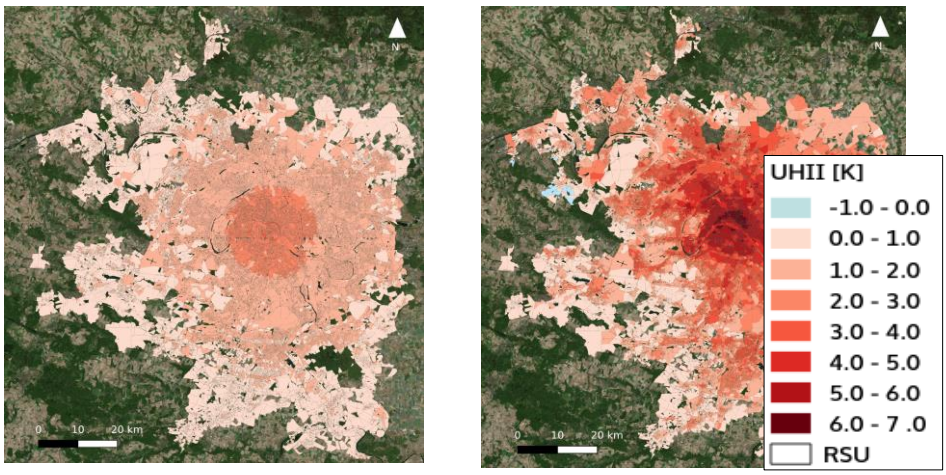
Absolute prediction error [K]	< 0.2	0.2-0.5	0.5-0.8	0.8-1.0	> 1.0
% of RSU (all cities)	21	24.4	15.3	7.3	32
% of RSU (without Paris)	25.4	29	17.5	7.6	20.5

Table 5: Percentage of Reference Spatial Units (RSU) for which the Random Forest model predicts the nocturnal urban heat island intensity with a given range of the absolute error.

Without Paris, more than half (54%) of the RSU are predicted with less than 0.5 K absolute error. The UHII values predicted with the Random Forest model as well as the results from the physically based atmospheric model Meso-NH are displayed for the megalopolis of Paris (Figure 7), the coastal city of La Rochelle (Figure 8), and the medium sized city of Angers (Figure 9). For La Rochelle and Angers, the Random Forest model captures the spatial pattern of the UHI, but underestimates the UHII values, especially around the city centre. More distant RSU are quite well predicted. The Random Forest prediction for Paris is of low quality, since the 41 French agglomerations used for the training are too different from Paris (e.g. in terms of population).

(a) Random Forest

(b) Meso-NH-TEB



(c) Error of Random Forest model

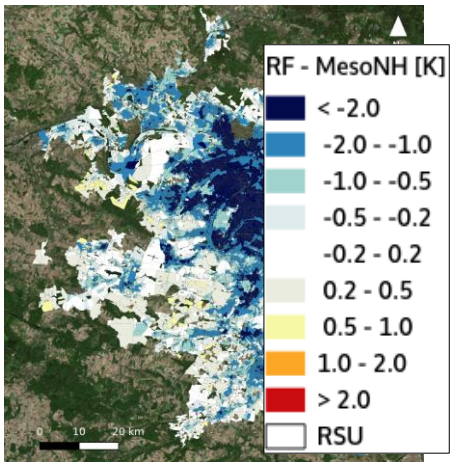
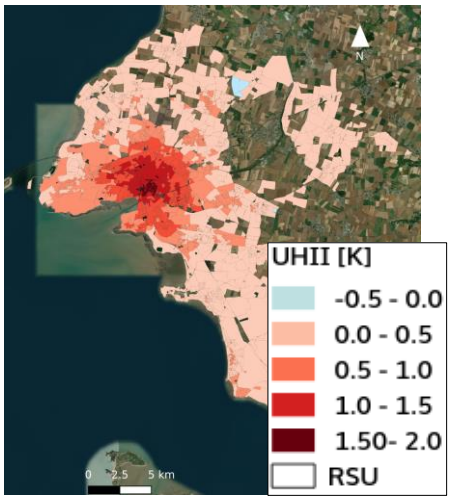
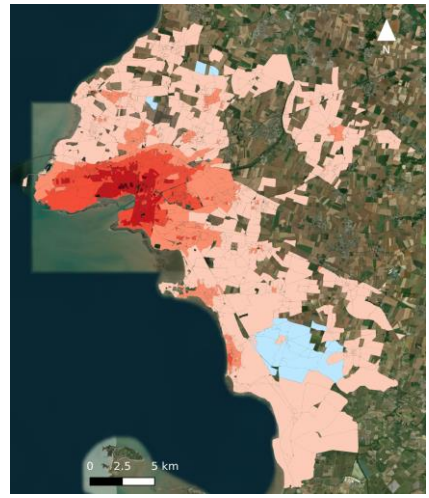


Figure 7: Nocturnal urban heat island intensity (UHII) in the agglomeration of Paris. (a): Statistically predicted with the Random Forest model, (b): Dynamically simulated with the physically based atmospheric model Meso-NH, and (c): Difference between the results of the Random Forest model and Meso-NH-TEB. Background data source: Bing Satellite, from QGIS QuickMapServices.

(a) Random Forest



(b) Meso-NH-TEB



(c) Error of Random Forest model

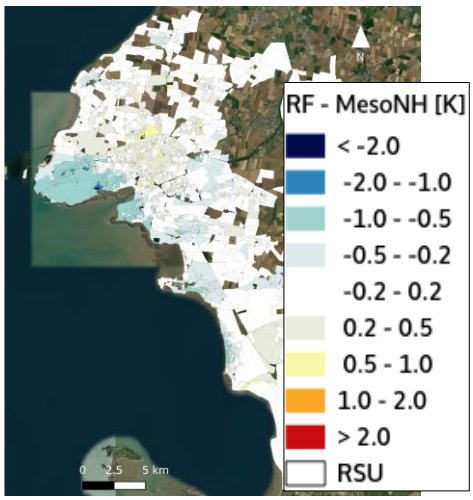
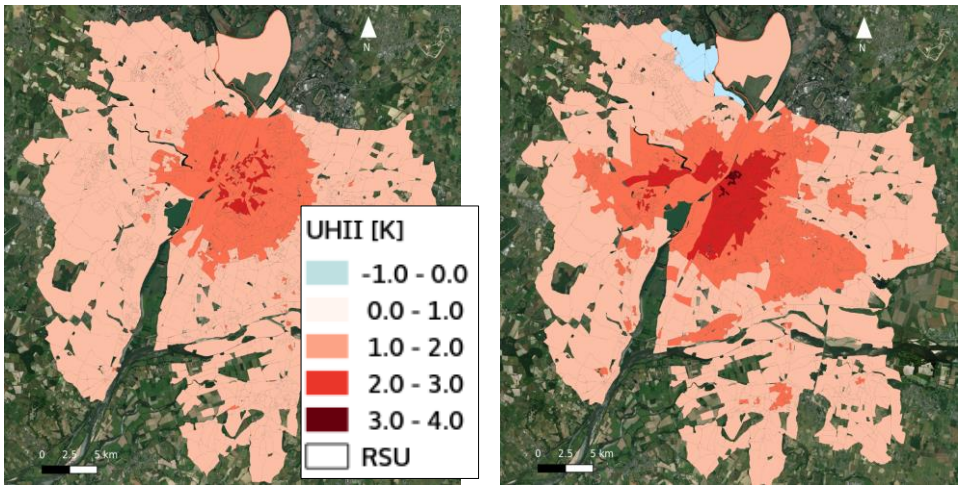


Figure 8: Same as Figure 7, but for the agglomeration of La Rochelle.

(a) Random Forest

(b) Meso-NH-TEB



(c) Error of Random Forest model

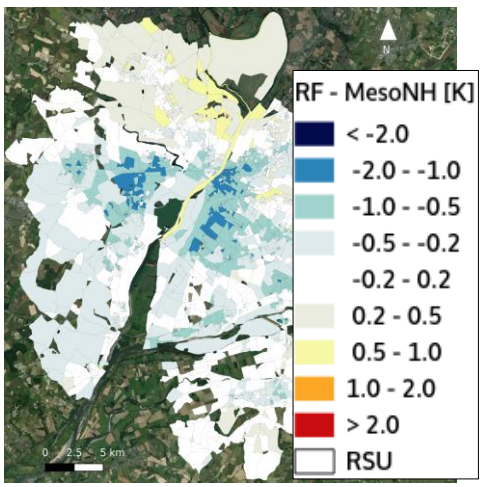


Figure 9: Same as Figure 7, but for the agglomeration of Angers.

4.3 Regression-based statistical model

The *UHII* is calculated using the combined linear regressions derived in Section 3 (Equation 26). The contribution of the distance to the coast is not taken into account for agglomerations located in more than 10 km distance to the coast. The contribution of the distance to the coast is normalised to 1 for cities located in 10 km distance to the coast to enforce the continuity of the equation. The formula cannot be used for agglomerations outside of France (due to the climatic region) and with less than 50000 or more than 10 million inhabitants.

(26)

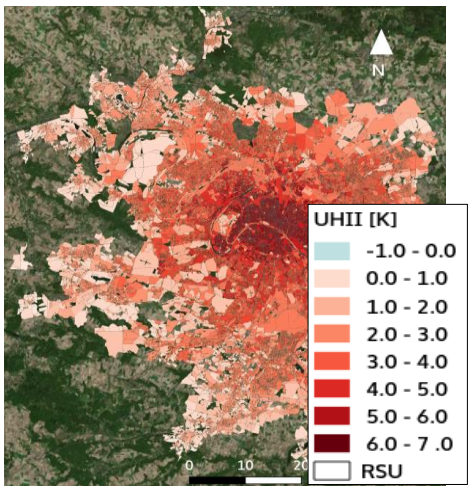
The RMSE for all RSU is 0.64 K, and 0.61 K without Paris, which is comparable to the value obtained with the Random Forest model. The RMSE per urban agglomeration varies between 0.19 and 1.34 K. The agglomeration with the highest prediction error is Calais (1.34 K). The percentage of the RSU for which the absolute value of the prediction error is within specific thresholds is given in Table 6. The regression-based model is able to predict the *UHII* with an absolute error of less than 0.5 K for 52% of the RSU.These results are comparable to those obtained by the Random Forest model, but the regression-based model is not directly limited by a training on the dataset, allowing it to give better results on particular cities like Paris.

Absolute prediction error [K]	< 0.2	0.2-0.5	0.5-0.8	0.8-1.0	> 1.0
% of RSU (all cities)	24.1	26.4	18.6	8.9	22
% of RSU (without Paris)	25.9	26.8	18	8.6	20.7

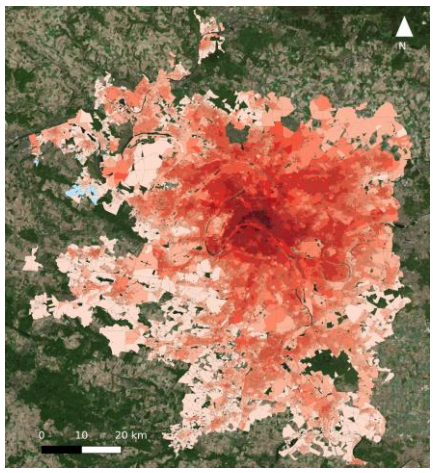
Table 6: Percentage of Reference Spatial Units for which the regression-based statistical model predicts the urban heat island intensity with a given range of the absolute error.

The *UHII* predicted by the regression-based model is displayed for the agglomerations Paris (Figure 10), La Rochelle (Figure 11), and Angers (Figure 12). For La Rochelle and Angers, the regression-based model presents results close to those of the Random Forest model. The spatial pattern of the UHII is captured, the UHII values are slightly underestimated around the city centre. The results for Paris obtained from the statistical model are better than those obtained with the Random Forest model; especially the spatial pattern of the UHII is better represented. An important shortcoming of the regression-based model is that it overestimates the effect of the distance to the centre of the agglomeration. This result is consistent with the relatively poor quality of the relationships between *UHIIN* and the normalised distance to the city centre (Figure 3). It cannot capture the variety of the forms of urban agglomerations with some being strongly different from a circle or possessing sub-centres. However, neglecting the influence of the distance to the city centre is not an alternative.

(a) Regression-based model



(b) Meso-NH-TEB



(c) Error of the regression-based model

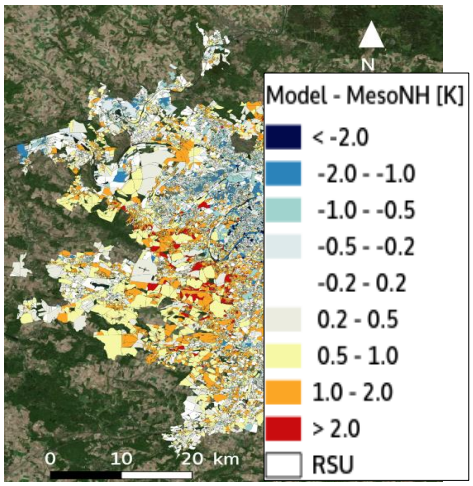
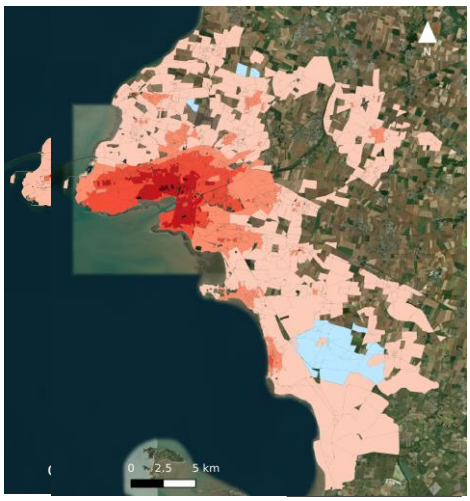


Figure 10: Nocturnal urban heat island intensity (UHII) in the agglomeration of Paris. (a): Statistically predicted using the regression-based model, (b): simulated with the physically based atmospheric model Meso-NH, (c): difference between the regression-based model and Meso-NH. *Background data: Bing Satellite, from QGIS QuickMapServices.*

(a) Regression-based model

(b) Meso-NH-TEB



(c) Error of the regression-based model

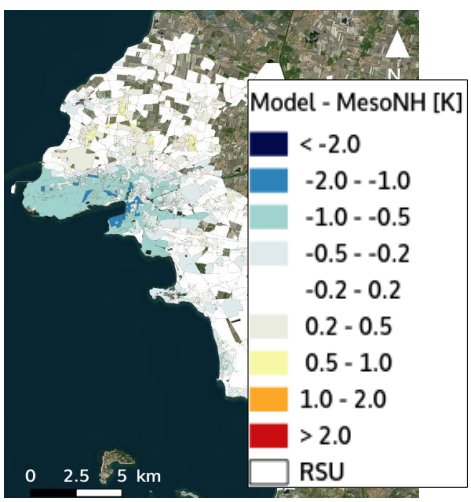
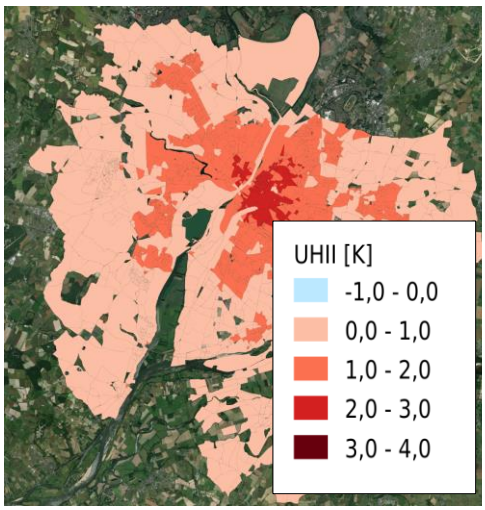
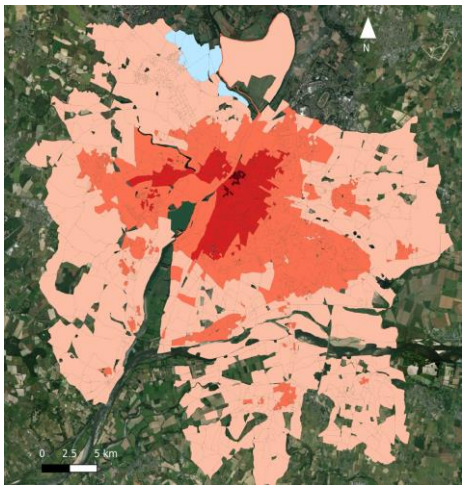


Figure *11*: Same as Figure 10, but for the agglomeration of La Rochelle.

(a) Regression-based model



(b) Meso-NH-TEB



(c) Error of the regression-based model

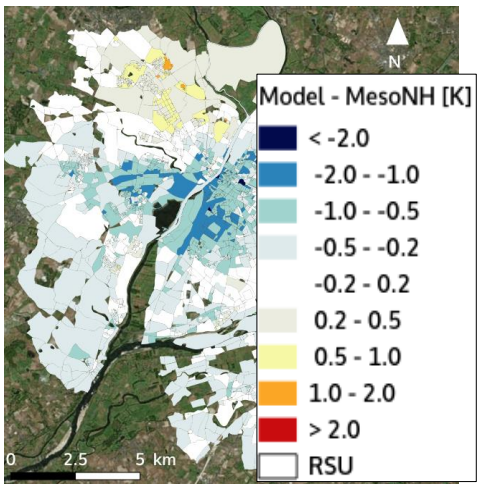


Figure 12: Same as Figure 10, but for the agglomeration of Angers.

5. Discussion

An important restriction of our study is that the statistical models for the UHII are derived based on the results of the numerical atmospheric model Meso-NH, which might be biased due to model shortcomings. Further validation using observation based data is required, but the required data are not yet available for a sufficient number of urban agglomerations.

However, the results obtained for the statistical relationships between the UHII and the different predictors are in line with what can be expected based on previous studies or physical reasoning.

- The maximum UHII of French agglomerations increases with the logarithm of the total population, which is consistent with the findings of Oke (1973), Sakakibara and Matsui (2005), Zhou et al. (2017), and other similar studies.
- The French climatic region also influences the maximum UHII with, on average, the largest values found in the Mountain and Semi-Continental climatic regions and the lowest values in the Pure Atlantic and Pure Mediterranean climatic regions. This is physically

plausible, especially since the higher values of wind speed in the climatic regions dominated by the Atlantic Ocean or the Mediterranean Sea lead to lower values of the UHII.

- The distance of the RSU to the centre of the agglomeration influences the UHII at RSU scale. This result is in slight contrast with Straub et al. (2019) who find that the distance to the city centre has only a low impact on the R^2 and the Mean Squared Error of their models. Based on the present study, it can be concluded that the distance to the city centre is a predictor that cannot be neglected, but is difficult to handle. For agglomerations with a circular shape and no large sub-centres at the periphery, the concept of the 'distance to the city centre' works quite well, but it cannot deal with cities whose shape strongly differs from a circle, for example those aligned along the coast or mountain valleys.

- The results obtained in the present study for the influence of the Local Climate Zone on the UHII at RSU scale are mainly consistent with what can be expected based on physical plausibility considerations. The largest positive local UHII increments are found in dense mid-rise and dense low-rise settings, whereas the largest negative local UHII increments occur for the LCZ lightweight low-rise, low plants, and water.

- Concerning the distance to the coast, the fine structure of the air temperature is governed by the local surface features as well as the turbulent structure of the atmospheric boundary layer above. The latter quickly sets in equilibrium with the surface below, in just a few km of distance from the shore of the sea (Pigeon et al 2007). This is why, 10 km from the shore, there is no direct effect of the sea on the anomaly of temperature between countryside and city (the UHI), while there is still an effect on the temperature itself since the proximity to the ocean influences the regional climate, but identical both for rural and urban areas.

- No statistically significant relationship is found in the present study between the maximum UHII of an urban agglomeration and elevation differences in and around the urban agglomeration.

An important restriction of the regression-based statistical models is that they depend on the *a priori* assumptions of which relationships might potentially exist. Therefore, not all possible forms of statistical relationships between the UHII and the predictors have been tested. It is therefore possible that other relationships exist which explain a higher degree of variance (e.g. for elevation difference). Furthermore, for the analysis of the statistical relationships for the maximum UHII at urban agglomeration scale, the UHII has been corrected for the effect of the total population. Uncertainties might arise since there might be collinearity between the total population and the other agglomeration scale predictor variables (distance to the coast, climatic region, and elevation differences). A strong collinearity appears however not plausible to the authors.

The present study also confirms Random Forest models as a valid method to predict the UHII based on predictors related to urban morphology and geographical factors. The RMSE of the predicted UHII is 0.85 K for all cities and varies from one agglomeration to another (0.25-1.69 K). For 30 out of the 42 agglomerations the RMSE of the predicted UHII is lower than 0.65 K, the value obtained by Makido et al. (2016) for the city of Doha using similar predictors. If we exclude the particular case of Paris from the analysis, we obtain an RMSE of 0.61 K, which is similar to the result obtained using the regression-based model. The regression-based model performs better on “outlier” agglomerations like Paris.

The RMSE obtained with the regression-based model (0.61 K, and ranging between 0.19 and 1.34 K depending on the agglomeration) is comparable to results from other statistical methods tested in the literature, like Ordinary Least Squares or Regression Tree (Makido et

al, 2016; RMSE of 1.25 K and 0.96 K) or results from empirical approaches like Zhang et al. (2019) who obtained an RMSE of 1.69 K.

A particularity of the present study is that it focusses only on the simulated *UHII* for one weather type favourable to a strong *UHII* instead of the climatological average *UHII*. The statistical models investigated in the present study do not take wind direction into account and might therefore perform even better for the climatological average *UHII* than for the *UHII* corresponding to one single weather type.

6. Conclusions and outlook

The present study investigated the nocturnal urban heat island intensity (UHII) of more than 200000 Reference Spatial Units in 42 French urban agglomerations simulated with the physically based atmospheric model Meso-NH coupled to the urban climate model TEB for a meteorological situation favourable to a strong UHII. The description of the form and functioning of the cities used in the numerical models is taken from administrative datasets on building outlines, demography, and building characteristics. Statistical relationships between the dynamically simulated UHII and the predictors total population, distance to the coast, French climatic zone, elevation differences, distance to the city centre, and Local Climate Zone have been derived. Furthermore, a Random Forest model and a regression-based model have been developed to calculate the UHII based on all the predictors.

The results for the relationships between the UHII and the predictors are mainly intuitive based on *a priori* knowledge of involved physical processes. The maximum UHII of an urban agglomeration increases with the logarithm of the number of inhabitants and with the logarithm of the distance to the coast for cities closer than 10 km to the coast. The maximum UHII is highest in the Semi Continental and Mountain climatic regions and lowest in the Pure Atlantic and Pure Mediterranean climatic regions. The Local Climate Zone alters the RSU-scale UHII, the dense mid-rise and dense low-rise LCZ exhibiting the largest positive and the LCZ lightweight low-rise, low plants, and water the largest negative

local increment of the UHII. However, the LCZ alone is not sufficient to predict the UHII of a given RSU. The most challenging predictor is the distance to the city centre, which strongly influences the UHII, but for which it is difficult to derive a universally applicable statistical relationship due to the variety of the shapes of the agglomerations.

The Random Forest model is able to predict the UHII with an RMSE of 0.85 K, but the prediction quality varies a lot depending on the considered urban agglomeration. The prediction is of low quality for Paris, since this megalopolis differs too much from the other French agglomerations, e.g. in terms of total population. The results for the other French agglomerations are close to those obtained in comparable studies. The disadvantage of the Random Forest model is that it is a *black box* since it is very difficult to understand how the predictors are actually used to predict the UHII. Furthermore, it is not easy to transfer to potential users. For this reason, the regression-based model has been developed as an alternative. It performs better than the Random Forest model for Paris, and gives similar prediction quality for the other agglomerations. Such a model might therefore be useful for operational applications, considering that most of the required data are easily accessible for a lot of agglomerations. The present study has investigated only large urban agglomerations (50000 to 10 million inhabitants). The developed statistical models cannot be used to quantify the UHII of villages or small towns, which should be the focus of future studies.

An important restriction of the present study is that the influence of meteorological factors on the UHII has not been taken into account by the statistical models. Especially the neglect of the wind velocity, which is found in the literature to be an essential predictor of the UHII (Bernard et al, 2017; Zhang et al., 2019) can lead to prediction errors. Future work could therefore focus on a larger number of meteorological situations (e.g. different values of the wind direction) and include the wind speed and direction as additional predictors in the Random Forest model. The results of the regression-based model and the Random Forest should also be validated using actual observation data and not only based on numerical

model simulations.

A similar investigation than in the present study, but at global scale would be of high interest for the urban planning community or the community dealing with climate change mitigation and adaptation. It would allow to sample a much larger range of climatic regions, urban morphologies, and socio-economical situations. However, a relatively homogeneous dataset on urban morphology, building construction practices, demographical variables, and human behaviour related to building and traffic energy consumption still needs to be compiled before such an investigation can be conducted.

Ackowlegements

This work was funded by the project applied Modelling and urbAn Planning laws: Urban Climate and Energy (MApUCE) funded by ANR with reference ANR-13-VBDU-0004, the project URCLIM funded by ERA4CS, an ERA-NET initiated by JPI Climate with co-funding from the European Union (Grant n° 690462) and the PAENDORA project funded by the ADEME (grant convention number 1717C0002

References

Alexander, P.J., and G. Mills, 2014: Local climate classification and Dublin's urban heat island. Atmosphere, 5 (4), 755-774.

Arnfield, A.J., 2003: Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. International Journal of Climatology. 23, 1–26. https://doi.org/10.1002/joc.859

Beck, C., A. Straub, S. Breitner, J. Cyrys, A. Philipp, J. Rathmann, A. Schneider, K. Wolf, and J. Jacobeit, 2018: Air temperature characteristics of local climate zones in the Augsburg urban area (Bavaria, southern Germany) under varying synoptic conditions. Urban Climate, 25, 152-166.

Berghauser-Pont, I., and P. Haupt, 2005: The spacemate: density and the typomorphology of the urban fabric. Nordisk Arkitekturforskning (Nordic Journal of Architectural Research) 4, 55–68.

Bernard, J., M. Musy, I. Calmet, E. Bocher, and P. Keravec, 2017: Urban heat island temporal and spatial variations: Empirical modeling from geographical and meteorological data. Building and Environment, 125, 423-438, 10.1016/j.buildenv.2017.08.009.

Bocher E., G. Petit, J. Bernard, and S. Palominos, 2018: A geoprocessing framework to compute urban indicators: The MApUCE tools chain. Urban Climate, 24, 153–174. https://doi.org/10.1016/j.uclim.2018.01.008.

Bourgeois A., M. Pellegrino, and J.-P. Lévy, 2017: Modeling and mapping domestic energy behavior: Insights from a consumer survey in France. Energy Research & Social Science, 32, 180-192.

Breiman, L., 2001: Random forests. Machine learning, 45(1), 5-32.

Bueno, B., G. Pigeon, L.K. Norford, K. Zibouche, and C. Marchadier, 2012: Development and evaluation of a building energy model integrated in the TEB scheme. Geoscientific Model Development, 5 (2), 433–448, doi:10.5194/gmd-5-433-2012.

Ching, J., M. Brown, S. Burian, F. Chen, R. Cionco, A. Hanna, T Hultgren, T. McPherson, D. Sailor,H. Taha, D. Williams, 2009: National Urban Database and access portal tool. Bulletin of the American Meteorological Society, 90, 1157–1168. https://doi.org/10.1175/2009BAMS2675.1.

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver, and M. Wehner, 2013: Long-term Climate Change: Projections, Commitments and Irreversibility. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley (editors)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Fenner, D., F. Meier, B. Bechtel, M. Otto, and D. Scherer, 2017: Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. Meteorologische Zeitschrift, 26, 525-547.

Hamdi, R., and V.Masson, 2008: Inclusion of a drag approach in the Town Energy Balance (TEB) scheme: Offline 1d evaluation in a street canyon. Journal of Applied Meteorology and Climatology, 47 (10), 2627–2644, doi:10.1175/2008JAMC1865.1.

Hidalgo, J., G. Pigeon, and V. Masson, 2008: Urban-breeze circulation during the CAPITOUL experiment: Observational data analysis approach. Meteorology and Atmospheric Physics, 102 (3-4), 223-241.

Hidalgo, J., V. Masson, and C. Baehr, 2014: From daily climatic scenarios to hourly atmospheric forcing fields to force soil-vegetation-atmosphere transfer models. Frontiers in Environmental Science, 2 (40), 1-13. doi:10.3389/fenvs.2014.00040.

Hidalgo, J., and R. Jougla, 2018: On the use of local weather types classification to improve climate understanding: An application on the urban climate of Toulouse. PLOS One, 13 (12), e0208138.

Hidalgo, J., G. Dumas, V. Masson, G. Petit, B. Bechtel, E. Bocher, M. Foley, R. Schoetter, and G. Mills, 2019: Comparison between local climate zones maps derived from administrative datasets and satellite observations. Urban Climate, 27, 64-89.

Ho, H., A. Knudby, P. Sirovyak, Y. Xu, M. Hodul, and S. Henderson, 2014: Mapping Maximum Urban Air Temperature on Hot Summer Days. Remote Sensing of Environment, 154, 38-45. 10.1016/j.rse.2014.08.012.

Hoffmann, P., O. Krueger, and K.H. Schluenzen, 2012: A statistical model for the urban heat island and its application to a climate change scenario. International Journal of Climatology, 32 (8), 1238–1248, doi:10.1002/joc.2348, URL https://rmets.onlinelibrary.wiley.com/doi/abs/.

Hoffmann, P., R. Schoetter, and K.H. Schlünzen, 2018: Statistical-dynamical downscaling of the urban heat island in Hamburg, Germany. Meteorologische Zeitschrift, 27 (2), 89-109.

Joly, D., T. Brossard, H. Cardot, J.N. Cavailhes, M. Hilal, and P. Wavresky, 2010: Les types de climats en France, une construction spatiale, Cybergeo: European Journal of Geography [Online Journal], Cartographie, Imagerie, SIG, document 501, http://journals.openedition.org/cybergeo/23155. 10.4000/cybergeo.23155.

Jougla R., J. Hidalgo, and B. Pouponneau, 2019: Identification des situations météorologiques locales pour une cinquantaine de villes françaises, La Météorologie, 106, 59-68, http://documents.irevues.inist.fr/handle/2042/70370

Kotharkar, R., and A. Bagade, 2017: Local Climate Zone classification for Indian cities: A case study of Nagpur. Urban Climate, 24, 369-392. http://dx.doi.org/10.1016/j.uclim.2017.03.003.

Kwok, Y.T., R. Schoetter, K.K.-L. Lau, J. Hidalgo, C. Ren, G. Pigeon, and V. Masson, 2019: How well does the Local Climate Zone scheme discern the thermal environment of Toulouse (France)? An analysis using numerical simulation data. International Journal of Climatology, 1-24, https://doi.org/10.1002/joc.6140.

Lac, C., J.-P. Chaboureau, V. Masson, J.-P. Pinty, P. Tulet, J. Escobar, M. Leriche, C. Barthe, B. Aouizerats, C. Augros, P. Aumond, F. Auguste, P. Bechtold, S. Berthet, S. Bielli, F. Bosseur, O. Caumont, J.-M. Cohard, J. Colin, F. Couvreux, J. Cuxart, G. Delautier, T. Dauhut, V. Ducrocq, J.-B. Filippi, D. Gazen, O. Geoffroy, F. Gheusi, R. Honnert, J.-P. Lafore, C. Lebeaupin Brossier, Q. Libois, T. Lunet, C. Mari, T. Maric, P. Mascart, M. Mogé, G. Molinié, O. Nuissier, F. Pantillon, P. Peyrillé, J. Pergaud, E. Perraud, J. Pianezze, J.-L. Redelsperger, D. Ricard, E. Richard, S. Riette, Q. Rodier, R. Schoetter, L. Seyfried, J. Stein, K. Suhre, M. Taufour, O. Thouron, S. Turner, A. Verrelle, B. Vié, F. Visentin, V. Vionnet, and P. Wautelet, 2018: Overview of the Meso-NH model version 5.4 and its applications. Geoscientific Model Development, 11, 1929–1969.

Leconte, F., J. Bouyer, R. Claverie, and M. Pétrissans, 2015: Using local climate zone scheme for UHI assessment: Evaluation of the method using mobile measurements. Building and Environment, 83, 39-49.

Lehnert, M., J. Geletič, J. Husák, and M. Vysoudil, 2015: Urban field classification by “local climate zones” in a medium-sized central European city: The case of Olomouc (Czech Republic). Theoretical and Applied Climatology, 122 (3-4), 531-541.

Lemonsu, A., and V. Masson, 2002: Simulation of a summer urban breeze over Paris. Boundary-Layer Meteorology, 104 (3), 463-490.

Lemonsu, A., V. Masson, L. Shashua-Bar, E. Erell, and D. Pearlmutter, 2012: Inclusion of vegetation in the Town Energy Balance model for modelling urban green areas. Geoscientific Model Development, 5 (6), 1377–1393, doi:10.5194/gmd-5-1377-2012.

Makido, Y., V. Shandas, S. Ferwati, and D.J. Sailor, 2016: Daytime Variation of Urban Heat Islands: The Case Study of Doha, Qatar. Climate. 4 (2) [32]. 10.3390/cli4020032.

Masson, V., 2000: A Physically-based scheme for the Urban Energy Budget in atmospheric models. Boundary-Layer Meteorology, 94, 357-397.

Masson, V., J.-L. Champeaux, F. Chauvin, C. Mériguet, and R. Lacaze, 2003: A global database of land surface parameters at 1-km resolution in meteorological and climate models. Journal of Climate, 16 (9), 1261-1282.

Masson, V., L. Gomes, G. Pigeon, C. Liousse, V. Pont, J.-P. Lagouarde, J. Voogt, J. Salmond, T.R. Oke, J. Hidalgo, D. Legain, O. Garrouste, C. Lac, O. Connan, X. Briottet, S. Lachérade, and P. Tulet, 2008: The Canopy and Aerosol Particles Interactions in TOulouse Urban Layer (CAPITOUL) experiment, Meteorology and Atmospheric Physics, 102, 135–157, https://doi.org/10.1007/s00703-008-0289-4.

Masson, V., P. Le Moigne, E. Martin, S. Faroux, A. Alias, R. Alkama, **S. Belamari, A. Barbu, A. Boone, F. Bouyssel, P. Brousseau, E. Brun, J.-C. Calvet, D. Carrer, B. Decharme, C. Delire, S. Donier, K. Essaouini, A.-L. Gibelin, H. Giordani, F. Habets, M. Jidane, G. Kerdraon, E. Kourzeneva, M. Lafaysse, S. Lafont, C. Lebeaupin Brossier, A. Lemonsu, J.-F. Mahfouf, P. Marguinaud, M. Mokhtari, S. Morin, G. Pigeon, R. Salgado, Y. Seity, F. Taillefer, G. Tanguy, P. Tulet, B. Vincendon, V. Vionnet, and A. Voldoire**, 2013: The SURFEXv7. 2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. Geoscientific Model Development, 6, 929-960.

Masson, V., W. Heldens, E. Bocher, M. Bonhomme, B. Bucher, C. Burmeister, C. de Munck, T. Esch, J. Hidalgo, F. Kanani-Sühring, Y.-T. Kwok, A. Lemonsu, J.-P. Lévy, B. Maronga, D. Pavlik, G. Petit, L. See, R. Schoetter, N. Tornay, A. Votsis, and J. Zeidler, 2020: City-descriptive input data for urban climate models: Model requirements, data sources and challenges, Urban Climate, 31, 100536, ISSN 2212-0955, https://doi.org/10.1016/j.uclim.2019.100536.

Moonen, P., T. Defraeye, V. Dorer, B. Blocken, and J. Carmeliet, 2012: Urban Physics: Effect of the micro-climate on comfort, health and energy demand. Frontiers of Architectural Research, 1, 197–228.

Oke, T.R., 1973: City size and the urban heat island. Atmospheric Environment, 7, 769-779. https://doi.org/10.1016/0004-6981(73)90140-6.

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

Piffer Dorigon, L., and M. Amorim, 2019: Spatial modeling of an urban Brazilian heat island in a tropical continental climate. Urban Climate, 28, 100461. DOI: 10.1016/j.uclim.2019.100461.

Pigeon, G., D. Legain, P. Durand, and V. Masson, 2007: Anthropogenic heat release in an old European agglomeration (Toulouse, France), International Journal of Climatology, 27, 1969–1981, https://doi.org/10.1002/joc.1530, 2007.

Pigeon, G., A. Lemonsu, C. Grimmond, P. Durand, O. Thouron, and V. Masson, 2007: Divergence of turbulent fluxes in the surface layer: case of a coastal city. Boundary-Layer Meteorology, 124, 269-290, 10.1007/s10546-007-9160-2.

Pigeon, G., K. Zibouche, B. Bueno, J. Le Bras, and V. Masson, 2014: Improving the capabilities of the Town Energy Balance model with up-to-date building energy simulation algorithms: an application to a set of representative buildings in Paris. Energy and Buildings, 76, 1–14, doi:https://doi.org/10.1016/j.enbuild.2013.10.038.

Plumejeaud-Perreau, C., C. Poitevin, C. Pignon-Mussaud, and N. Long, 2015: Building Local Climate Zones by using socio-economic and topographic vectorial databases. Proceeding of the 9th International Conference on Urban Climate, July 20-24, Toulouse, France.

Richard Y., J. Emery, J. Dudek, J. Pergaud, C. Chateau-Smith, S. Zito, M. Rega, T. Vairet, T. Castel, T. Thévenin T., and B. Pohl, 2018: How relevant are Local Climate Zones, Urban Climate Zones, and USGSDijon for urban climate research? Dijon (France) as a case study. Urban Climate, 26, 258-274.

Sakakibara, Y., and E. Matsui, 2005: Relation between heat island intensity and city size indices/urban canopy characteristics in settlements of Nagano basin, Japan. Geographical review of Japan, 78(12), 812-824.

Schoetter, R., V. Masson, A. Bourgeois, M. Pellegrino, and J.-P. Lévy, 2017: Parametrisation of the variety of human behaviour related to building energy consumption in the Town Energy Balance (SURFEX-TEB v. 8.2). Geoscientific Model Development, 10, 2801–2831.

Schoetter, R., J. Hidalgo, R. Jougla, V. Masson, M. Rega, and J. Pergaud, 2020 A statistical-dynamical downscaling for the urban heat island and building energy consumption - Analysis of its uncertainties. Published online by Journal of Applied Meteorology and Climatology, https://doi.org/10.1175/JAMC-D-19-0182.1.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France Convective-Scale Operational Model. Monthly Weather Review, 139, 976-991.

Shepherd, J. M., 2005: A Review of Current Investigations of Urban-Induced Rainfall and Recommendations for the Future, Earth Interactions, 9, 1–27, https://doi.org/10.1175/EI156.1.

Skarbit, N., I.D. Stewart, J. Unger, and T. Gál, 2017: Employing an urban meteorological network to monitor air temperature conditions in the 'local climate zones' of Szeged, Hungary. International Journal of Climatology, 37 (S1), 582-596.

Stewart, I.D., and T.R. Oke, 2012: Local climate zones for urban temperature studies. Bulletin of the American Meteorological Society, 93, 1879–1900. http://dx.doi.org/10.1175/bams-d-11-00019.1.

Stewart, I.D., T.R. Oke, and E.S. Krayenhoff, 2014: Evaluation of the 'local climate zone' scheme using temperature observations and model simulations. International Journal of Climatology, 34 (4), 1062-1080.

Straub, A., K. Berger, S. Breitner, J. Cyrys, U. Geruschkat, J. Jacobeit, B. Kühlbach, T. Kusch, A. Philipp, A. Schneider, R. Umminger, K. Wolf, and C. Beck, 2019: Statistical modelling of spatial patterns of the urban heat island intensity in the urban environment of Augsburg, Germany. Urban Climate, 29, 100491, 10.1016/j.uclim.2019.100491.

Tornay, N., R. Schoetter, M. Bonhomme, S. Faraut, and V. Masson, 2017: GENIUS: A methodology to define a detailed description of buildings for urban climate and building energy consumption simulations. Urban climate, 20, 75-93.

United Nations (UN) - Department of Economic and Social Affairs, 2018: 2018 Revision of World Urbanization Prospects. Available: https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html, last access on June 13 2019.

Unger, J., 1999: Urban-rural air humidity differences in Szeged, Hungary, International Journal of Climatology, 19, 1509–1515, https://doi.org/10.1002/(SICI)1097-0088(19991115)19:13<1509::AID-JOC453>3.0.CO;2-P.

Verdonck, M.-L., M. Demuzere, H. Hooyberghs, C. Beck, J. Cyrys, A. Schneider, R. Dewulf, and F. van Coillie, 2018: The potential of local climate zones maps as a heat stress assessment tool, supported by simulated air temperature data. Landscape and urban planning, 178, 183-197.

Wang, R., C. Ren, Y. Xu, K.K.-L. Lau, and Y. Shi, 2018: Mapping the local climate zones of urban areas by GIS-based and WUDAPT methods: A case study of Hong Kong. Urban Climate, 24, 567-576. https://doi.org/10.1016/j.uclim.2017.10.001.

Wilby, R.L., 2003: Past and projected trends in London's urban heat island. Weather, 58, 251-260.

Zhang, X., G.-J. Steeneveld, D. Zhou, C. Duan, and A.A.M. Holtslag, 2019: A diagnostic equation for the maximum urban heat island effect of a typical Chinese city: A case study for Xi'an. Building and Environment, 158, 39-50. 10.1016/j.buildenv.2019.05.004.

Zheng, Y., C. Ren, Y. Xu, R. Wang, J. Ho, K. Lau, and E. Ng, 2018: GIS-based mapping of Local Climate Zone in the high-density city of Hong Kong. Urban Climate, 24, 419-448. http://dx.doi.org/10.1016/j.uclim.2017.05.008

Zhou, B., D. Rybski, and J.P. Kropp, 2017: The role of city size and urban form in the surface urban heat island. Scientific Reports, 7, 4791, https://doi.org/10.1038/s41598-017-04242-2.

CRediT author statement

Thomas Gardes : Conceptualization, Methodology, Formal analysis, Writing – Original Draft

Robert Schoetter : Conceptualization, Methodology, Resources, Writing – Original Draft, Writing – Review & Editing

Julia Hidalgo : Methodology, Ressources, Writing – Review & Editing

Nathalie Long : Methodology, Ressources, Writing – Review & Editing

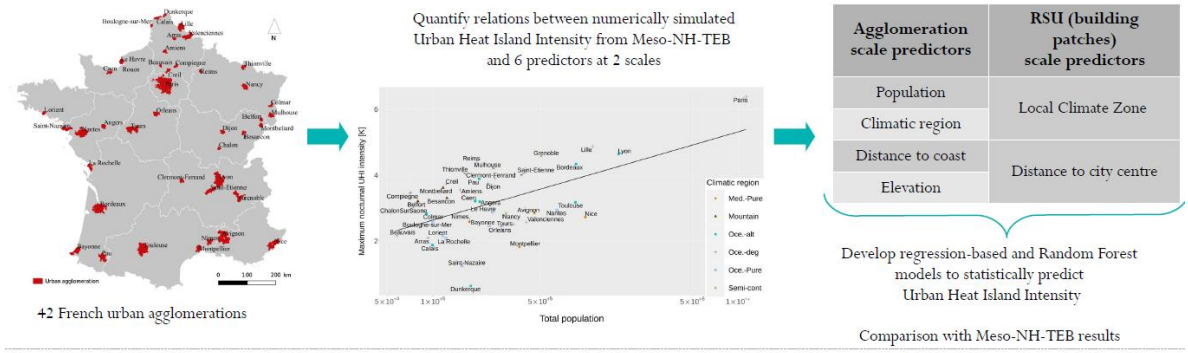
Eva Marquès : Methodology, Validation, Writing – Review & Editing

Valéry Masson : Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project administration, Funding acquisition

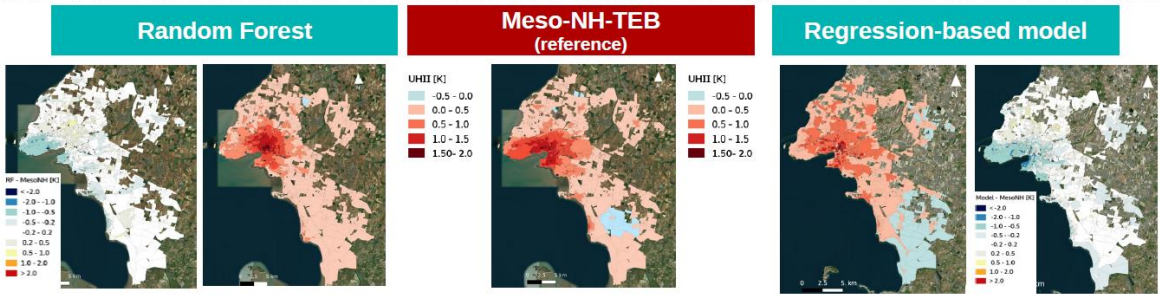
Declaration of Competing Interest

The authors declared that they have no conflict of interest that could have appeared to influence the work reported in this paper.

Graphical abstract



Comparison with Meso-NH-TEB results



Research highlights

- Physically-based simulation of the urban heat island (UHI) for 42 French cities.

- Quantification of the relationships between the UHI and geographical factors.

- Regression-based (RB) and Random Forest (RF) model developed to predict the UHI.

- RB andRF models predict the UHI with less than 0.5 K absolute error for about 50% of the building blocks.

- The RB model is easier to transfer to practitioners than the *black box* RF.