



Original research article

Urban multi-model climate projections of intense heat in Switzerland

Annkatrin Burgstall^{a,*}, Sven Kotlarski^a, Ana Casanueva^{a,b}, Elke Hertig^c, Erich Fischer^d,
Reto Knutti^d

^a Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

^b Meteorology Group, Department of Applied Mathematics and Computer Science, University of Cantabria, Santander, Spain

^c Faculty of Medicine (Regional Climate Change and Health), University of Augsburg, Augsburg, Germany

^d Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland



ARTICLE INFO

Keywords:

Climate change
Climate projections
Urban climate
Urban heat island
Heat extremes
Climate services

ABSTRACT

This paper introduces a straightforward approach to generate multi-model climate projections of intense urban heat, based on an ensemble of state-of-the-art global and regional climate model simulations from EURO-CORDEX. The employed technique entails the empirical-statistical downscaling method quantile mapping (QM), which is applied in two different settings, first for bias correction and downscaling of raw climate model data to rural stations with long-term measurements and second for spatial transfer of bias-corrected and downscaled climate model data to the respective urban target site. The resulting products are daily minimum and maximum temperatures at five urban sites in Switzerland until the end of the 21st century under three emission scenarios (RCP2.6, RCP4.5, RCP8.5). We test the second-step QM approach in an extensive evaluation framework, using long-term observational data of two exemplary weather stations in Zurich. Results indicate remarkably good skill of QM in present-day climate. Comparing the generated urban climate projections with existing climate scenarios of adjacent rural sites allows us to represent the urban heat island (UHI) effect in future temperature-based heat indices, namely tropical nights, summer days and hot days. Urban areas will be more strongly affected by rising temperatures than rural sites in terms of fixed threshold exceedances, especially during nighttime. Projections for the end of the century for Zurich, for instance, suggest more than double the number of tropical nights (Tmin above 20 °C) at the urban site (45 nights per year, multi-model median) compared to the rural counterpart (20 nights) under RCP8.5.

Practical implications

- Many users of science-driven climate information need scenario data specifically to understand and adapt to future climatic changes for urban areas. The urban climate generally differs from the climate in the rural surrounding; particularly during nighttime, temperatures are markedly higher than outside of the densely built urban areas, a phenomenon known as urban heat island (UHI e.g. Oke, 1982; Oke et al., 1991; Vogt and Parlow, 2011). As a consequence of global climate change (IPCC, 2013), the additional heat experienced in cities will become increasingly problematic and makes people living in urban agglomerations especially predisposed to heat stress and health risks (Gabriel and Endlicher, 2011; Kjellstrom and Weaver, 2009; Kovats and Hajat, 2008; Scherer et al., 2013). Given the large

and steadily growing fraction of urban population, it is imperative to understand future climate conditions in urban environments, allowing for effective mitigation and adaptation strategies tailored to urban areas.

- Despite the high relevance of urban climate, “standard” climate scenarios like the Swiss climate scenarios CH2018 (2018), are often restricted to representative rural sites, where long-term and high-quality observational data series are available. These data series are typically used for statistical post-processing procedures to correct the raw climate model output (Fig. 1, “GCM-RCM simulations”) for systematic biases and to fill the scale gap with local information (CH2018, 2018; Christensen et al., 2008; Rajczak et al., 2016). Results are bias-corrected climate scenarios at the rural site scale, here considered as “standard climate services” (Fig. 1).
- In this study, we introduce a straightforward method to quantify future urban heat based on a large and state-of-the-art ensemble

* Corresponding author.

E-mail address: annkatrin.burgstall@meteoswiss.ch (A. Burgstall).

of global (GCM) and regional climate model (RCM) simulations from the Coordinated Downscaling Experiment for the European domain (EURO-CORDEX; [Jacob et al., 2014](#); [Kotlarski et al., 2014](#)). With empirical quantile mapping (QM), we apply a well-established post-processing approach to generate urban projections at daily resolution for daily minimum (Tmin) and daily maximum temperatures (Tmax) at five large Swiss cities (Basel, Bern, Geneva, Pully/Lausanne, Zurich).

- These results are an important add-on to the existing CH2018 products: they provide customized climate services ([Fig. 1](#), “User-specific climate services”) and help protect those population groups and sectors that have been identified as especially vulnerable towards high temperatures in urban areas (e.g. the elderly; health sector, construction sector/working outside; [Flouris et al., 2018](#); [Kjellstrom and Weaver, 2009](#); [Ragettli et al., 2017](#)). As a “ready-to-use” product (bias-corrected, high-resolved), the generated urban scenarios are directly applicable to climate change impact assessments. Comparing the temperature-based scenarios of the urban sites with the corresponding rural counterparts available through CH2018 enables to account for the UHI effect in future climates and provides users with a first-order estimate of urban–rural temperature differences under various future pathways. In order to make our results more approachable, we communicate our findings in terms of standard climate indices, such as the number of tropical nights (Tmin > 20 °C), summer days (Tmax > 25 °C) and hot days (Tmax > 30 °C; [CH2018, 2018](#); [ETCCDI, 2019](#)). At the same time, we explicitly include major uncertainty sources such as climate model and emission scenario uncertainty.
- Practical implications of this study primarily concern impact modelers, regional and local authorities or climate service centers that provide users with information on urban future climates and the technical information (including limitations and uncertainties of the employed data and method) they might not be aware of or they need support with. The employed method is transferable in both space and scope, i.e. it can be applied to locations outside Switzerland and to further climate services, such as economic or heat-related mortality analyses for urban areas under climate change.

[Oke et al., 1991](#); [Roth, 2013](#)). The UHI effect is especially harmful for human health as it is mainly a nighttime phenomenon, and the body can cope less with daytime heat loads when there is not enough sleep and recovery during the night ([Gabriel and Endlicher, 2011](#); [Grize et al., 2005](#); [Kovats and Hajat, 2008](#); [Ragettli et al., 2017](#); [Scherer et al., 2013](#)). The combined effect of UHI and the suggested increase in mean global air temperatures puts urban heating on the list of key risks due to climate change ([IPCC, 2014](#)) and creates a growing demand for robust projections of future climatic conditions in cities. Projections specifically for urban sites allow for sophisticated impact assessments and help protect the large and continuously increasing urban population ([Arnfield, 2003](#); [Roth, 2013](#); [Stewart and Oke, 2012](#)).

The Coordinated Downscaling Experiment for the European domain (EURO-CORDEX; [Jacob et al., 2014](#); [Kotlarski et al., 2014](#)) initiative provides the largest and state-of-the-art ensemble of climate change projections based on global (GCM) and regional climate model (RCM) simulations, but their spatial resolution of approximately 12 km (EUR-11) and 50 km (EUR-44) is usually too coarse to account for the UHI. By using some statistical adaptations, such as downscaling and bias-correction, this large ensemble has been used to produce climate change projections of heat stress at European rural sites ([Casanueva et al., 2020a](#)) and particularly in Switzerland ([CH2018, 2018](#)). The lack of long and high-quality observations in urban areas, though, hinders a straightforward estimation of future urban climates.

There have been efforts to simulate the evolution of urban climate on the basis of advanced atmospheric models coupled to urban (canopy) models or building energy models at high resolution (e.g. [Chen et al., 2011](#); [Lauwaet et al., 2015](#); [Salamanca et al., 2011](#)). In another study, [Langendijk et al. \(2019\)](#) use multi-model RCM data (at approximately 12 km spatial resolution) to investigate opportunities and limitations of using EURO-CORDEX RCMs in urbanized areas; more precisely, in the research area Berlin (Germany) and its rural surroundings. The RCMs considered in the study of [Langendijk et al. \(2019\)](#) represent urban environments through their land surface parameterization schemes. Results indicate a stronger urban–rural temperature difference based on maximum temperatures compared to minimum temperatures, which contrasts the observational data and a large number of previous studies that propose high gradients especially during nighttime. [Fischer et al. \(2012\)](#) focus on the GCM Community Climate System Model (CCSM4) that explores subgrid-scale urban processes based on an urban canyon model ([Oleson et al., 2010](#)). There, the diurnal temperature cycle of the UHI is captured remarkably well (see also [Oleson et al., 2011](#)). Other studies stress the potential of methods on big data in climate research ([Knüsel et al., 2019](#)). [Oh et al. \(2020\)](#), for instance, use deep-learning (neural network models) to forecast the magnitude and characteristics of the UHI in Seoul (South Korea). Similarly, [Gobakis et al. \(2011\)](#) consider different types of artificial neural networks for UHI prediction in the study area Athens (Greece). Apart from that, some authors propose empirical-statistical methods to generate local climate scenarios for urban sites. [Van der Schriek et al. \(2020\)](#) use a five-member RCM sub-

Introduction

People living in urban environments tend to be more exposed to heat stress and the resulting health risks than people living in non-urban regions ([Gabriel and Endlicher, 2011](#); [Kjellstrom and Weaver, 2009](#); [Kovats and Hajat, 2008](#); [Scherer et al., 2013](#)), because air temperatures at urban sites are often higher than temperatures in nearby rural surroundings. The main causes for the so-called urban heat island (UHI) effect are the larger heat capacity of urban fabrics, the trapping of long-wave radiation in urban canyons, the reduced vertical exchange of air masses, the lower evapotranspiration due to sparser vegetation coverage, and anthropogenic heat emissions (e.g. [Fischer et al., 2012](#);



Fig. 1. Using downscaled and bias-corrected GCM-RCM simulations (left) to generate standard climate services (center; e.g. CH2018 climate scenarios for rural sites) that are further customized to user-specific climate services (right; e.g. climate scenarios for urban sites).

ensemble from EURO-CORDEX to study future temporal UHI trends by contrasting simulated air temperatures of an urban and a rural station in Athens (Greece). They bias correct the simulated data against 20-year observational records using a quantile mapping (QM) technique. Other studies focus on crowdsourced data, i.e. data collected from a large number of people (Muller et al., 2015). Meier et al. (2017), for instance, compare crowdsourced air temperature data from private weather stations (netatmo) with official weather stations of the German Weather Service (DWD) and the urban climate observation network (UCON; Fenner et al., 2014) in the study area Berlin (Germany). They stress the potential of crowdsourced temperature data in terms of cost-efficiency and dense data coverage, especially in urban areas due to high population density. Nevertheless, they conclude that comprehensive quality checks are key to fully benefit from crowdsourced atmospheric data. Burgstall (2019) compares three promising statistical and empirical approaches to generate climate projections for Swiss urban sites under a high emission scenario: (1) a physically-based diagnostic equation designed by Theeuwes et al. (2017) to parameterize the daily maximum UHI, which is then added to existing climate scenarios of rural sites (2) a multiple linear regression, taking various predictor variables of the rural site in order to model temperatures at the urban site, and (3) the post-processing method QM (Rajczak et al., 2016), used to spatially transfer projections of rural sites to urban target sites. Especially for QM, evaluation results demonstrate good skill in creating robust urban climate scenarios at the local scale.

In the present study, we focus on the most promising method presented by Burgstall (2019), QM, whose high performance has also been acknowledged by a large number of previous studies (e.g. CH2018, 2018; Gudmundsson et al., 2012; Gutiérrez et al., 2018; Monhart et al., 2018; Themeßl et al., 2012). We extend the work of Burgstall (2019) by considering a broader range of possible future pathways and validate the QM method more rigorously. We focus on three RCPs (2.6, 4.5 and 8.5

Moss et al., 2010), which range from a mitigation scenario implying fast and substantial reductions in global greenhouse gas emissions (RCP2.6) to continued emission growth and global warming until the end of the century (RCP8.5). Due to the relatively short observational record for urban areas, ranging between 7 and 28 years, we employ QM in a two-step manner: first, bias correcting and downscaling regional climate models (RCMs) to the rural site scale (done within CH2018) and second, spatially transferring scenario data from rural to urban locations (done in this paper), resulting in climate scenarios for urban sites. The

subsequent comparison of rural and urban scenarios (station couples) allows for quantifying the (station couple specific) urban–rural temperature difference, i.e. the UHI, in future climates by assuming a stationary relation between both sites. We analyze the UHI effect in terms of standard climate indices such as the number of tropical nights (TN; $T_{min} > 20\text{ }^{\circ}\text{C}$), summer days (SD; $T_{max} > 25\text{ }^{\circ}\text{C}$) and hot days (HD; $T_{max} > 30\text{ }^{\circ}\text{C}$; CH2018, 2018; ETCCDI, 2019).

The paper is structured as follows. First, we introduce the data and methods including a brief description of the validation framework and skill score. After presenting the results of both the evaluation and the application of the proposed method, the paper is finalized with a discussion on potential limitations and concluding remarks.

Data and methods

Observational data

We consider five station couples in Switzerland, i.e. a rural site with an adjacent urban station (Fig. 2 and Table 1). We use observed and modelled (i.e. quantile-mapped) temperature data at daily resolution, namely: minimum (T_{min}) and maximum (T_{max}) 2 m temperature for both urban and rural stations.

Urban and rural stations are characterized in terms of the local climate zones (LCZ) developed by Stewart and Oke (2012; see Table 2). We use the classification results of Gehrig et al. (2018), whose analyses cover all employed stations except SMA, which is categorized in the framework of this study based on a visual analysis of satellite pictures (Google Maps). Most of the considered stations labeled as *rural* are located in sparsely built areas or in locations with an open low-rise building geometry and low plants. Stations considered as *urban* are surrounded by a compact or open midrise building structure, mostly with paved ground. Note that a strict division of the analyzed stations into the two categories *urban* and *rural* is not always clearly feasible. The sites PUY and BAS, classified as rural, for instance, could also be described as suburban (see Table 2). For the sake of simplicity, though, we continue using the terms *urban* and *rural* station throughout the manuscript and refer to Table 2 for more detailed information on the respective locations.

For rural sites, observations are provided by five automated climate stations, which are operationally run by the Swiss national weather service MeteoSwiss (SwissMetNet stations) and are in accordance with

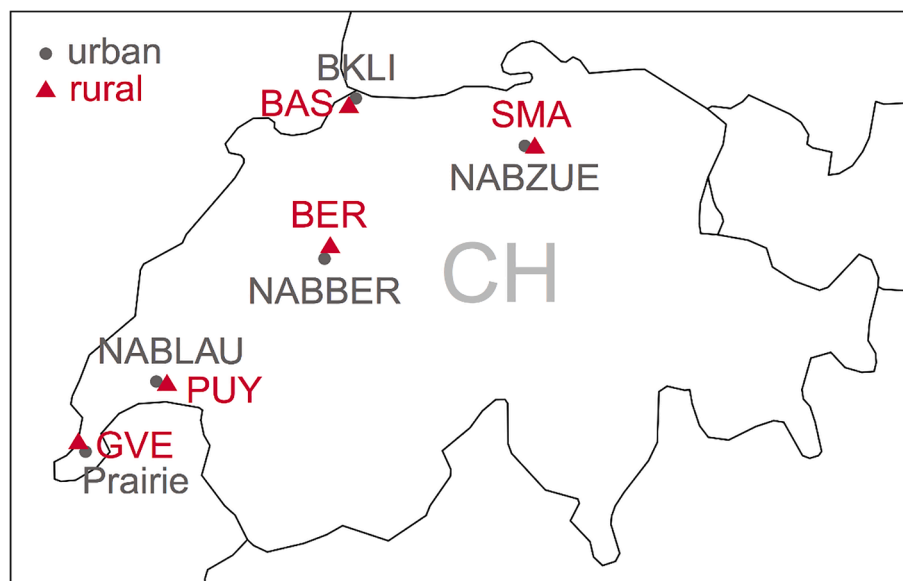


Fig. 2. Station couples (urban and rural site) located in Switzerland (CH) used in this study. Red triangles show rural stations, grey circles indicate urban sites. See Table 1 for full station names. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

The considered station couples (urban, rural) with the respective overlapping time period of available temperature data.

Rural station (code)	Urban station (code)	Time period overlap
Basel Binningen (BAS)	Basel Klingelbergstrasse (BKLI)	2003–2018
Bern Zollikofen (BER)	Bern Bollwerk (NABBER)	2001–2018
Geneva Cointrin (GVE)	Geneva Prairie (PRAIRIE)	2010–2016
Pully (PUY)	Lausanne (NABLAU)	1991–2018
Zurich Fluntern (SMA)	Zurich Kaserne (NABZUE)	1995–2018

Table 2

Classification of the considered stations (rural, urban) in terms of local climate zones (LCZ; Stewart and Oke, 2012) and other local characteristics.

Rural station	LCZ	Definition LCZ	Local characteristics	Altitude (m)
Basel Binningen (BAS)	9D	Sparsely built, low plants	Suburban	316
Bern Zollikofen (BER)	D	Low plants		553
Geneva Cointrin (GVE)	9D	Sparsely built, low plants	Close to airport	412
Pully (PUY)	6D	Open low-rise, low plants	Suburban	456
Zurich Fluntern (SMA)	6B	Open low-rise, scattered trees		556
Urban station				
Basel Klingelbergstrasse (BKLI)	2E	Compact midrise, paved	Rooftop station	285
Bern Bollwerk (NABBER)	5E	Open midrise, paved	Rooftop station	536
Geneva Prairie (PRAIRIE)	2E	Compact midrise, paved	Rooftop station	420
Lausanne (NABLAU)	5E	Open midrise, paved		530
Zurich Kaserne (NABZUE)	5D	Open midrise, low plants	Close to lake	409

Source: Gehrig et al. (2018) and MeteoSwiss (2020).

World Meteorological Organization (WMO) standards. Measurement data for urban sites are taken from various cantonal and university observational networks: the Department of Environmental Sciences of the University of Basel provided data for the station BKLI, the Laboratoire énergie, environnement, architecture of the Haute école du paysage, d'ingénierie et d'architecture de Genève, hepia, for the station PRAIRIE and the National Air Pollution Monitoring Network (NABEL) for the stations NABBER, NABLAU and NABZUE. We consider different time periods depending on the station couple and their extent of overlapping data, i.e. high-quality measurements at both urban and rural sites for the same time period (see Table 1, right column). Independent of the data provider, urban stations tend to have a shorter and (or) sparser observational data coverage compared to rural stations, which are mainly operated by national weather services and guarantee high-quality and long-term data sets. The time period used is thus predominantly determined by the data availability at the respective urban station.

Climate model data

We consider quantile-mapped, i.e. bias-corrected and downscaled, RCM data for T_{min} and T_{max} at rural stations from the CH2018 product DAILY-LOCAL¹ (CH2018, 2018). The original RCM simulations represent a multi-model regional climate projection ensemble (multiple RCMs

driven by multiple GCMs) provided through EURO-CORDEX (Jacob et al., 2014; Kotlarski et al., 2014). RCM output for the period 1981–2099 at daily resolution and at approximately 12 km (EUR-11) and 50 km (EUR-44) spatial resolution were considered. The set of RCP8.5 simulations forms the core of the CH2018 projections and combines 12 different GCMs (including different GCMs or the same GCM with different initial conditions) and 7 different RCM versions (including two different versions of one RCM; see Table A1 in the appendix), offering a decent estimate of climate projection uncertainty. In total, 21 model chains (i.e. GCM-RCM combinations) were used for each of the three RCPs (RCP2.6, RCP4.5 and RCP8.5). Due to an originally smaller number of simulations available for RCP2.6 and RCP4.5, missing simulations for these RCPs were filled using time-shift-based pattern scaling, which guarantees a simulation basis consistent for all considered RCPs (CH2018, 2018). Note that the applied pattern scaling approach (based on 30-year periods) results, by definition, in non-transient projections, that do not support analyses requiring transient model data throughout the century. In this work, we therefore restrict the analysis to 30-year time slices for the reference period 1981–2010 (named “1995”) and for three future scenario periods 2020–2049 (“2035”), 2045–2074 (“2060”) and 2070–2099 (“2085”). Further note that the 30-year historical time slices (1981–2010) include 5 years of scenario data. Time-slice differences were calculated using in each case a RCP scenario period minus the historical period complemented by the corresponding RCP scenario run. In Table A1 in the appendix, we show the final set of all employed simulations for RCPs 2.6, 4.5 and 8.5. The reader is referred to Sørland Lund et al. (2020) for more details on the CH2018 methodology.

Quantile mapping

Despite constant improvements, RCMs are too coarsely resolved for a direct application in climate change impact studies and are prone to systematic biases and uncertainties (Christensen et al., 2008; Rajczak et al., 2016). A common technique to account for these limitations is to consider distribution-based statistical transfer relations that (might) include a downscaling component and correct for potential model biases. Dosio (2016) found that bias adjustment also provides a more robust climate change signal for indices based on absolute thresholds. Within CH2018 (2018), empirical QM has been applied to statistically adjust RCM data to the site-specific climate. The basic principle of QM is to correct a biased simulated distribution towards an observed distribution by calibrating a quantile-based correction function between observed and simulated quantiles (Panofsky and Brier, 1968). QM has been widely used in recent studies (e.g. Gudmundsson et al., 2012; Gutiérrez et al., 2018; Monhart et al., 2018; Themeßl et al., 2012). Gutiérrez et al. (2018), for instance, rate QM among the best performing approaches in an intercomparison of various statistical downscaling and bias correcting methods over Europe. Ivanov and Kotlarski (2017) and Rajczak et al. (2016) confirm the high performance and robust results of QM after validating the approach for a large number of meteorological variables and several official weather stations in Switzerland.

The QM implementation as employed in CH2018 and in the present study is taken from Ivanov and Kotlarski (2017) and Rajczak et al. (2016) and is integrated in an extensive R-package², which we use for the QM application (second QM step, i.e. spatial transfer) in this work. The implementation is based on the correction of the 99 empirical percentiles (1st to 99th percentile) of the modelled distribution towards their observational counterparts. A linear interpolation of the correction is used for values between two percentiles. For values that lie outside the calibration range, i.e. values that are smaller than the first and larger than the last percentile, the correction function of the first and of the last

¹ <https://doi.org/10.18751/Climate/Scenarios/CH2018/1.0>. <https://www.nccs.admin.ch/nccs/en/home/data-and-media-library/data/ch2018—climate-scenarios-for-switzerland.html>.

² qmCH2018 v1.0.1, <https://doi.org/10.5281/zenodo.3275571>, <https://github.com/SvenKotlarski/qmCH2018>.

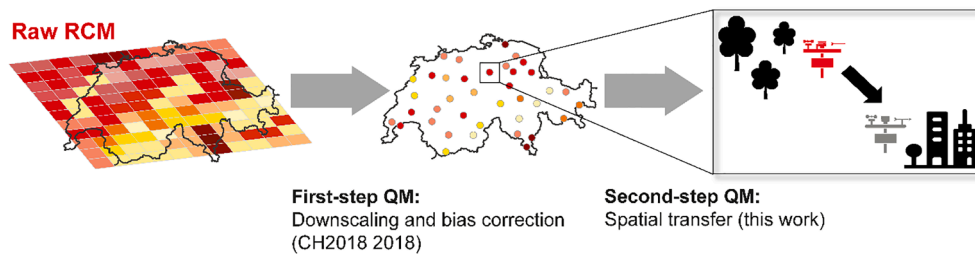


Fig. 3. Two-step quantile mapping (QM) approach. Source: Figure modified from Burgstall (2019).

percentile is applied, respectively (Themeßl et al., 2012). The QM correction function is determined separately for each day of the year (DOY) with a moving window of 91 days. More precisely, the quasi-seasonal transfer function is centered over a certain day and includes 45 days before and 45 days after the respective DOY (Feigenwinter et al., 2018; Rajczak et al., 2016). We apply QM in a two-step manner. The technique was originally developed by Rajczak et al. (2016), who evaluated a number of meteorological variables for a case study about permafrost in Switzerland. Here the method is used and evaluated for the first time in the frame of urban climates.

In the first step (Fig. 3, accomplished within CH2018), simulated and observed distributions are matched by calibrating a correction function in the historical reference period 1981–2010 that translates the simulated quantiles into their observed counterparts. Applying the so-established correction function to the entire simulated period 1981–2099 results in the CH2018 product DAILY-LOCAL, which is available for various (rural) stations in Switzerland and used in the present study for further analyses. The first-step QM comprises both a bias correction and a downscaling component to the rural site scale. In the second step (Fig. 3, carried out in the present work), rural scenarios of the selected sites (see Table 1, left column) available through CH2018 are spatially transferred to the respective urban target site (see Table 1, middle column). Here, the calibrated correction function is calibrated on pairwise daily observations at the urban and the rural location in a common reference period, which depends on the station couple (see Table 1, right column). We apply the calibrated correction functions to the QM data originating from step 1 (see above) for the individual 30-year time slices in the historical period (1995) and in the three scenario periods of the rural data series (2035, 2060 and 2085) to spatially transfer them to the same 30-year time slices at the urban target site. In the second-step, QM does not comprise a bias correction and downscaling component, but a spatial transfer function.

Applying the QM technique two times enables the generation of climate scenarios for urban sites despite their often short and (or) sparse observational data coverage. The resulting climate scenarios of urban stations and their rural counterparts are examined in terms of the heat indices TN, SD and HD (CH2018, 2018; ETCCDI, 2019). Contrasting the frequencies of the projected heat indices within the respective station couples allows to account for some aspects of the UHI in future climates. To reveal the long-term climate change signal, simulated frequencies of heat indices are averaged over each 30-year period. We then analyze the ensemble median values (multi-model medians) of the so-obtained 30-year means and their climate change signals between the future scenario periods and the historical reference period, respectively. To account for model uncertainty, we additionally consider the 5th–95th percentiles of the multi-model ensemble.

Evaluation of quantile mapping

We evaluate the proposed QM method and consider three independent cross-validation techniques, applied to daily summer data (June, July and August; JJA) of Tmin and Tmax of the exemplary station pair SMA-NABZUE in the observational period 1995–2018. The choice of SMA-NABZUE as exemplary station couple is motivated by its high

quality and long data availability of 24 years for both sites. Note that a large number of existing studies has already demonstrated the ability and skill of the first-step QM in correcting systematic model biases at the local scale (e.g. Gudmundsson et al., 2012; Ivanov and Kotlarski, 2017; Themeßl et al., 2012). We thus focus our evaluation on the second-step QM only, i.e. its performance for the spatial transfer of climate data. A brief description of the used cross-validation strategies is given in the following.

- Split sample approach (SSA): The overlapping time period (1995–2018) is split into two chronological data sets of 12 years each. The QM model is trained with the first half of data (1995–2006) and tested on the remaining half (2007–2018), and vice versa. Modeled results of both periods are merged before carrying out the performance analysis, resulting in one skill measure.
- Split sample approach warm/cold (SSA(WC)) and cold/warm (SSA(CW)): Same as SSA, but the data set is split in terms of years with warmer summers and years with colder summers, based on the summer mean temperature within the full period (1995–2018) at the rural site. The 12 years with warmer summers are used for training the model and the 12 years with colder summers for testing the model (WC), and vice versa (CW). The approach results in two independent skill measures.
- Limited data approach (LDA): The data set is split into thirds with the first two-thirds (1995–2010) used for calibrating the QM model and the remaining period (2011–2018) for validating the model. Within the training set, different calibration period lengths are analyzed, starting with 1 year and steadily increasing to 15 years with an increment of 1, resulting in 15 individual skill measures. By randomly combining the years, 16 different calibration samples for each length are validated against observations for the period 2011–2018. Within one calibration sample, there is no repetition of years and the order does not matter (no permutation).

Note that the SSA(CW) technique additionally evaluates the QM skill by accounting for long-term trends; the observed trend in temperature proposes a warmer testing than training period (CH2018, 2018; IPCC, 2013; Rajczak et al., 2016). The LDA has the additional objective to simulate a lack of data availability by employing various sample sizes and to quantify uncertainties by considering random combinations of calibration years (Rajczak et al., 2016). That way, the LDA helps identifying the minimum length of overlapping data necessary to calibrate QM and to still obtain robust results, i.e. a considerably smaller bias compared to shorter calibration lengths.

The validation of the second-step QM focuses on the skill score mean bias (bias). It describes the offset between predicted (\bar{X}_{pred}) and observed data (\bar{X}_{obs}) according to

$$Bias = \bar{X}_{pred} - \bar{X}_{obs} \quad (1)$$

where \bar{X} is the average value over the respective validation period. Note that here \bar{X}_{pred} and \bar{X}_{obs} refer to observed data since the second step of the QM builds upon the differences between the rural and urban observed records. We use the bias to evaluate several variables, namely: Tmin,

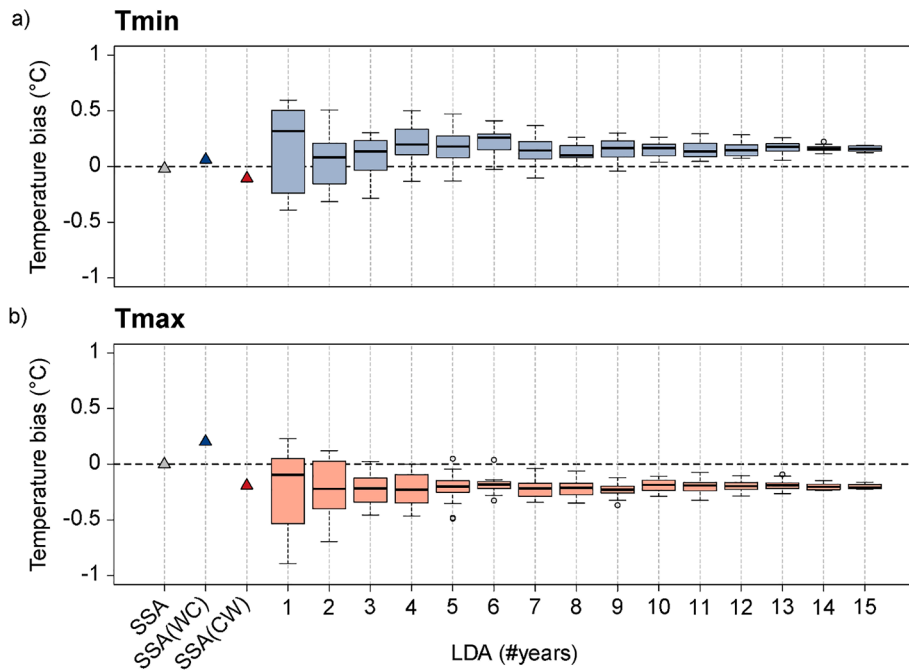


Fig. 4. Validation results for urban daily minimum temperature (Tmin; a) and urban daily maximum temperature (Tmax; b), using the two-step QM approach for the station couple SMA-NABZUE. The x-axis shows the considered validation techniques: SSA (split sample approach), SSA (WC) (SSA warm/cold), SSA (CW) (SSA cold/warm), LDA (limited data approach) with the number of considered years for calibration (#years) increasing from left to right. The y-axis denotes the temperature bias. Each boxplot for the LDA spans the 16 different calibration samples and ranges between the 1st quartile (Q1, lower border) and the 3rd quartile (Q3, upper border), referred to as interquartile range (IQR). The bold black line centered within the IQR shows the median. The lower whisker refers to $Q1 - 1.5 \cdot IQR$ and the upper whisker to $Q3 + 1.5 \cdot IQR$. Circles below/above the whiskers indicate outliers. Source: Figure modified from Burgstall (2019).

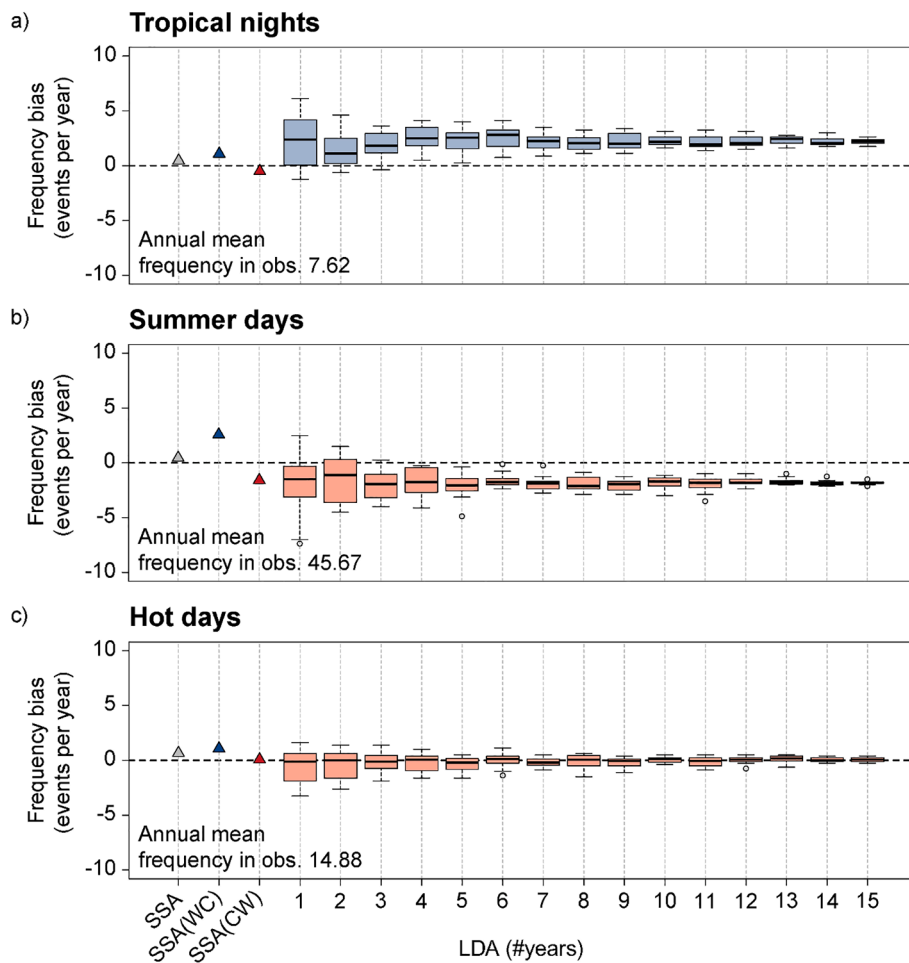


Fig. 5. Same as Fig. 4 but for the annual number of tropical nights (TN; Tmin > 20 °C), summer days (SD; Tmax > 25 °C) and hot days (HD; Tmax > 30 °C), respectively. Averaged over the observational period (obs.) 1995–2018, there are 7.62 TN, 45.67 SD and 14.88 HD per year at the urban site. Source: Figure modified from Burgstall (2019).

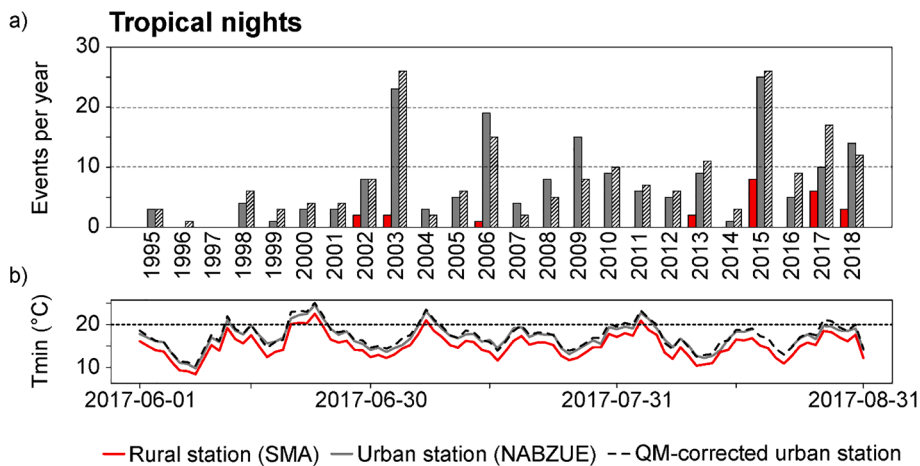


Fig. 6. a) Number of tropical nights (TN; events/year) for the rural (SMA; red), urban (NABZUE; grey) and QM-corrected urban station (dashed and black) based on SSA for the summers of 1995–2018. b) Daily evolution of minimum temperatures (Tmin) for the rural (red), urban (grey) and QM-corrected urban time series (dashed and black) in summer 2017, where the bias in the number of TN (QM vs observations) is especially large. The dotted horizontal line in b) indicates the TN threshold of 20 °C. *Source:* Figure modified from Burgstall (2019). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Tmax and the employed heat indices (TN, SD, HD).

Results

Spatial transfer evaluation

Temperatures

Validation results of the spatial transfer of Tmin and Tmax from SMA to NABZUE, shown in Fig. 4, show remarkably small median biases which amount to approximately 0.25 °C for Tmin and −0.25 °C for Tmax for virtually all validation strategies. The independent cross-validation method SSA with training and testing periods of 12 years, respectively, shows especially good skill with a bias that is nearly zero for both variables. Keep in mind, though, that modeled results of both periods (12 years each) are merged before they are compared to the whole observational time series of 24 years; slightly positive and negative biases of the individual results partly compensate. When focusing on the cross-validation techniques SSA(WC) and SSA(CW), which consider not a single bias for the whole time series but for 12 years, respectively, both variables indicate slightly positive (WC) and negative (CW) biases, depending on the calibration period. Correction functions being calculated in years with warmer summers and applied to years with colder summers, as for SSA(WC), tend to overestimate urban temperatures, whereas the opposite holds true for SSA(CW). This effect is slightly more pronounced for Tmax (Fig. 4b) than for Tmin (Fig. 4a) as the urban–rural temperature difference between the calibration and the validation period based on Tmax is larger (Burgstall, 2019). From the good skill in SSA(CW) for both variables we assume that the spatial transfer of rural temperature data to an urban target site will perform equally well under ongoing climate change. Yet, one should keep in mind that due to slightly underestimated urban temperatures in the validation setting, urban climate projections will be conservative estimates and might be even higher in reality. Results of the LDA reveal a slight but systematic overestimation of urban Tmin values (Fig. 4a) and a slight, systematic underestimation of urban Tmax values (Fig. 4b). Interestingly, median biases of Tmin approach the bias of SSA(WC) with increasing number of years considered for calibration, whereas skills of Tmax behave similar to the results of SSA(CW). The reason is that in the calibration period 1995–2010, where years are randomly selected for LDA, the urban–rural temperature offset based on Tmin is slightly higher than in the validation period 2011–2018, similar to SSA(WC). For Tmax, in turn, the difference between urban and rural temperatures is lower in the training period than in the testing period, similar to SSA(CW; see also Burgstall, 2019). Note that the bias for Tmin and LDA with 15 years of training is larger than for SSA(WC), which is trained with 12 years. The reason is that the urban–rural temperature offset is larger for 1995–2010 with respect to 2011–2018 (LDA) than the training compared to the testing

period in SSA(WC), resulting in a larger overestimation in LDA compared to SSA(WC). For Tmax, the offset in the training versus the testing period in LDA is almost the same compared to the training versus testing period in SSA(CW; see also Burgstall, 2019). The LDA technique also reveals the relation between bias and number of years used for calibrating QM: for both variables, as expected, the skill substantially improves with the number of years employed for calibration (#years in Fig. 4). Median biases as well as uncertainty ranges, i.e. extended whiskers resulting from different combinations of years, indicate better results if at least 7–8 years are considered for the model calibration, which is consistent with the results of Rajczak et al. (2016). Yet, already three years of calibration can offer reasonable results with a median bias close to zero. Relatively large variations are still visible, though.

Climate indices

In terms of the number of TN and SD (Fig. 5a and b), the QM performance reveals a very similar pattern compared to Tmin and Tmax, respectively, and shows good skill throughout all validation strategies. For the number of HD (Fig. 5c), even better skill is achieved. For 15 years of calibration, the bias is almost zero (0.06 HD). As the UHI effect is a nighttime phenomenon and urban–rural temperature differences are mostly visible in terms of Tmin (e.g. Oke, 1982; Oke et al., 1991; Vogt and Parlow, 2011), we focus on the evaluation results of TN.

The SSA technique shows the highest skill among the three validation strategies with a bias of almost zero for TN (Fig. 5a). Its variants WC and CW reveal slightly positive and negative biases of +1 and −0.5 TN. For the LDA, the skill is considerably improved when calibrating the model with seven or more years, especially in terms of the uncertainty range. A systematic overestimation of about two TN per year remains, though, even if considering an extended calibration length of 15 years. Modeled values are overestimated, as the calibrated correction function is based on a larger urban–rural temperature gradient in the training period as it prevails in the testing period (see also Tmin in Fig. 4a).

Fig. 6a indicates the overall good performance of QM for most of the analyzed summers of 1995–2018, cross-validated by the SSA and depicted separately for each year. Particularly remarkable is that QM manages to (approximately) capture the number of TN at the urban site (here: NABZUE) even in years with no TN at the rural site (here: SMA), for instance in 1995, 2005 and 2010. Nevertheless, there are years with poorer skill, for instance in 2017, where the number of TN is considerably overestimated. Despite such strong offsets during individual years, the overall bias shown in Fig. 5a (see SSA) is relatively low, as the bias is calculated from the whole 24-year series. To better understand the varying results of modeled and observed data during individual years like 2017, we focus on the time series of Tmin in that specific summer (Fig. 6b). During the first and the last week of August, observed urban temperatures happen to be often slightly below the TN-threshold of

Tropical nights

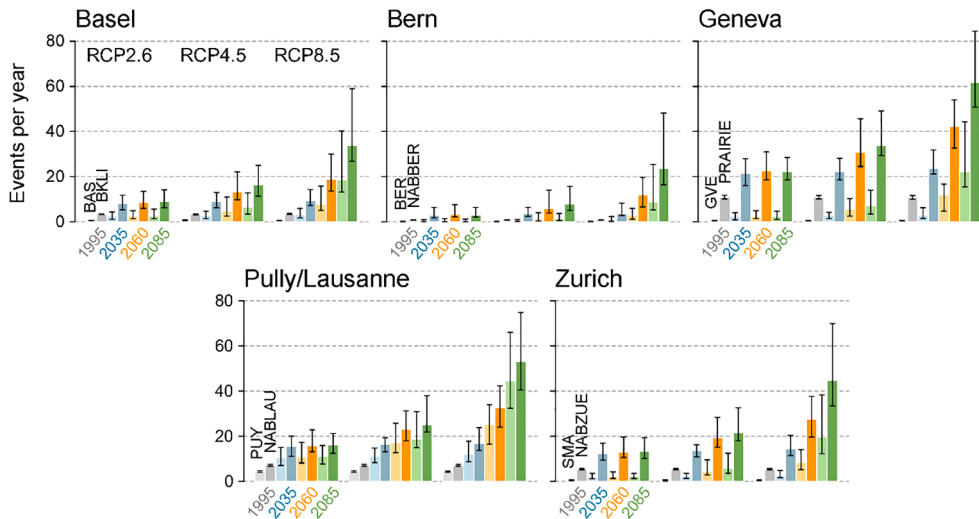


Fig. 7. Number of tropical nights (TN) per year averaged over the 30-year reference period (1995) and the three 30-year scenario periods (2035, 2060, 2085, multi-model combination) for the RCPs 2.6, 4.5 and 8.5 at the considered station couples (see also [Table 1](#)). The respective rural site is shown in lighter colors, the urban site in darker colors. Bars indicate the ensemble median value and whiskers the 5–95% model range. Scenarios of the rural site are based on the first-step QM ([CH2018, 2018](#)); scenarios of the urban site are based on the second-step QM (this study).

20 °C. As the quantile-mapped time series overestimates these instances to temperatures slightly above 20 °C, QM appears to strongly overestimate the number of TN during the considered periods (even though the bias of T_{min} is as small as in other years). Apparently, the absolute nature of the temperature threshold index TN strongly affects the QM results in terms of skill.

Climate scenarios for urban sites

We analyze the projected evolution of the considered heat indices over the course of the 21st century in terms of the absolute number of events per year for four 30-year periods (reference and scenario) and mainly focus on one illustrative station couple (SMA-NABZUE).

For all considered stations, the number of heat indices is projected to increase in the future at both urban and rural sites ([Figs. 7–9](#)), which is a general result of the increasing temperature level. The increases are strongest for RCP8.5 and weakest for RCP2.6. Regardless of the considered period or emission scenario, projections suggest urban areas to be primarily affected by high temperatures in terms of fixed threshold exceedances, especially during nighttime. This is consistent with previous studies about the UHI effect (e.g. [Oke, 1982; Oke et al., 1991; Vogt and Parlow, 2011](#)); we see a clear contrast in the number of TN between rural and urban sites (UHI) across all analyzed station couples ([Fig. 7](#)). For RCP8.5 and the late-century period, the urban stations NABBER (24 TN) and PRAIRIE (62 TN) are projected to experience up to three times as many TN as their rural counterparts (9 TN and 22 TN, respective multi-model medians). For BKLI (34 TN), projections suggest almost twice as many TN compared to BAS (19 TN). The number of TN in NABLAU (53 TN) is expected to be about 18% higher than in the surrounding rural area PUY (45 TN). Besides the higher number of urban tropical nights, also the change in occurrence of these nights is larger in urban environments compared to the rural counterpart. These findings are consistent with [Fischer et al. \(2012\)](#). They found that a larger urban increase in high heat-stress nights stems from statistical non-linearity in the exceedance frequency: based on more frequent present-day exceedance, the same mean shift in the temperature climatology in response to the changing climate leads to a larger change in exceedance frequency.

Also in Zurich ([Fig. 7](#)), urbanites will be exposed to nighttime heat stress more frequently than people living outside the urban area. Already in present-day climate (1995) with TN being basically absent at the rural site SMA, the number of TN (multi-model median) at the urban site NABZUE is around seven times larger. Even for the lower limit of the

uncertainty range and already for the mid-century period (2060), the number of TN in NABZUE exceeds median conditions at the rural site for the late-century period (2085) under all RCPs. At the end of the century, projections for RCP8.5 suggest around 45 TN per year (33–70 TN, 5th–95th percentile range) at the urban site compared to about 20 threshold exceedances per year (12–38 TN, 5th–95th percentile range) at the rural counterpart. Note that the uncertainty ranges under the given period and RCP are especially large in the two stations and overlap ([Fig. 7](#)). The overlap, however, does not imply that a given model chain indicates more TN in the rural than in the urban site; the overlapping range is due to different model chains. The same applies to the other station couples with overlapping uncertainty ranges (Basel, Bern, Pully/Lausanne; [Fig. 7](#)).

The UHI effect is not restricted to nighttime conditions and can also occur during the day ([Tzavali et al., 2015](#)). Various factors such as extra heat release caused by human activities and less evapotranspiration in urban areas due to the lack of vegetation and water contribute to higher temperatures in urban areas during the day (e.g. [Gehrig et al., 2018](#)). For the considered T_{max} -based indices, SD ([Fig. 8](#)) and HD ([Fig. 9](#)), and for most analyzed station couples, we see more frequent threshold exceedances at the urban site. The differences between urban and rural stations, though, are less pronounced than for T_{min} -based indices ([Fig. 7](#)). Projections for the urban sites NABBER and NABLAU, for instance, show 6%–7% more SD ([Fig. 8](#)) and 18%–24% more HD per year ([Fig. 9](#)) compared to the respective rural counterpart. For the station couples BAS-BKLI and GVE-PRAIRIE, in turn, the opposite applies and the number of SD ([Fig. 8](#)) and HD ([Fig. 9](#)) at the rural site is higher. Bear in mind that the considered urban sites are rooftop stations, characterized by lower temperatures during midday compared to the rural counterpart ([Gehrig et al., 2018; Vogt and Parlow, 2011](#)). Also the fact that the rural partner sites are either suburban (BAS) or close to the airport with large asphalt structures (GVE) can contribute to the inverted daytime situation. We acknowledge that rooftop stations are less-ideal settings for UHI analyses. Especially from the perspective of human health, operationally run measurement sites at ground level are the preferred data source, if available. A detailed description of the local characteristics of the measurement sites (see [Table 2](#)) is thus key to make UHI results comparable among different station couples. NABBER is a rooftop station as well. However, here the rural site (BER) is characterized by very natural surroundings, so that temperatures in the rural area stay below urban temperatures at all times ([Figs. 7–9](#)). Focusing on Zurich, SD occur regularly under current conditions at both sites, yet more frequently in the urban area (almost 55 SD; multi-model median)

Summer days

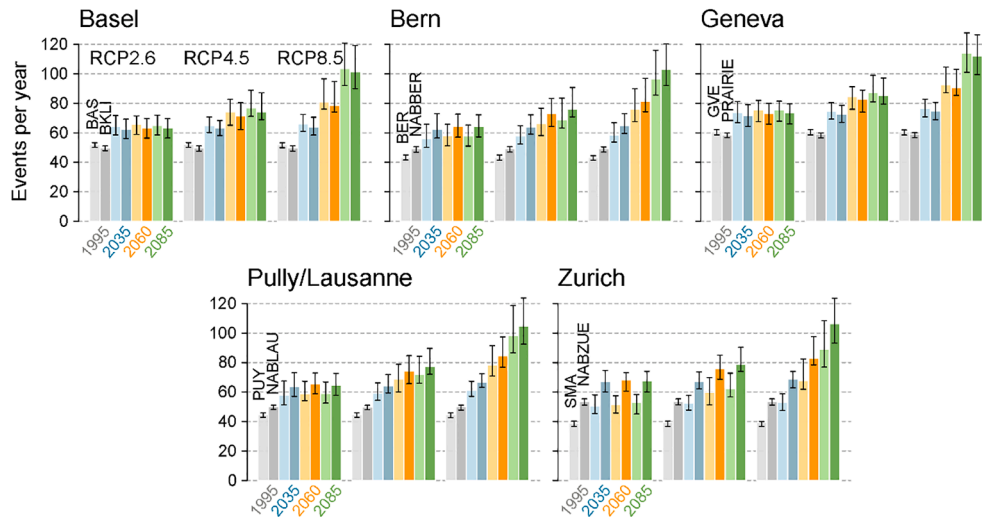


Fig. 8. Same as Fig. 7 but for the number of summer days (SD) per year.

Hot days

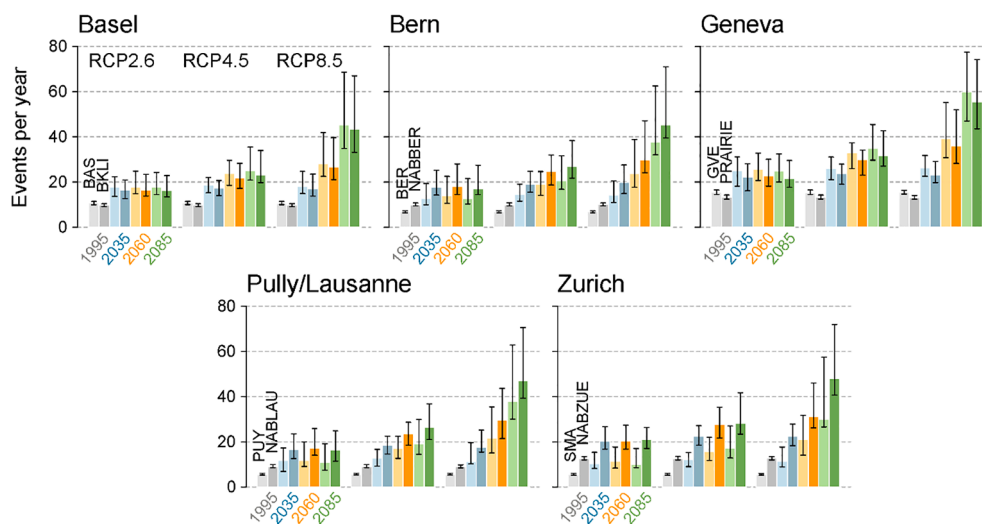


Fig. 9. Same as Fig. 7 but for the number of hot days (HD) per year.

compared to the rural surrounding (almost 40 SD; multi-model median; Fig. 8). These numbers are projected to considerably increase in the future. For RCP8.5 and the late-century period, the rural site (SMA) will be affected by almost 90 SD per year; the urban site (NABZUE) will experience 15% SD more. With almost 105 SD every year (on average), the urban population would experience heat extremes on numerous and continuous days of the year. For RCPs 2.6 and 4.5 and earlier scenario periods, numbers are lower.

Overall, for the strong emission scenario RCP8.5 and the late scenario period, all five analyzed station couples, both urban and rural sites, are projected to experience summer days, on average, for at least a whole summer season (approximately three months).

Compared to the number of SD, HD occur less frequently at both urban and rural site in Zurich. Results in Fig. 9 indicate about 5 days per year at the rural site SMA under present-day climate, which are projected to increase up to 30 days per year at the end of the century for RCP8.5. At the respective urban partner site NABZUE, the temperature threshold for HD is reached more often already in present-day climate (over 12 days per year) and will be exceeded almost 50 times each year

by the end of the 21st century for RCP8.5. For the period 2060, the projected number of HD at the urban site is already higher than the number of HD at the rural site for the period of 2085 (comparing the respective multi-model medians).

Differences in the station altitude between the two stations of a given couple might influence urban–rural temperature offsets as well (see Table 2). This primarily applies to the station couple in Zurich, where the urban site NABZUE is located at a 147 m lower altitude than the rural counterpart SMA, resulting in an especially pronounced urban–rural temperature offset. The opposite is true for the station couple PUY-NABLAU. Here, the urban site is located 74 m higher than the rural site, reducing the urban–rural temperature offset as the higher altitude compensates the higher temperatures at the urban location. For the remaining station couples, differences in station altitude are negligibly small and thus not considered to play a major role.

Limitations and sources of uncertainty

The employed spatial transfer approach applied to climate

projections is associated with a number of assumptions and limitations that need to be taken into account when interpreting our results. Beside uncertainties relevant for climate projections (model uncertainty, emission scenario uncertainty), the application of the presented method QM entails additional sources of uncertainty.

Note that the resulting projections (as shown in Figs. 7–9) cover a certain range of possible future outcomes, which corresponds to the spread of the considered climate simulations (ensemble spread). The spread is different for each emission scenario and arises primarily from model uncertainty. To properly represent the uncertainty associated with the climate projections, we use a multi-model combination set combining 12 different GCMs and 7 different RCM versions (see Table A1 in the appendix and Chapter 3). The uncertainty range in the generated projections can be large and typically increases over the course of the 21st century with the signal strength, in particular for RCPs 4.5 and 8.5. The number of TN at the urban site NABLAU (Fig. 7), for instance, is projected to lie within a range between 41 and 75 TN (5th–95th percentile range) in the late scenario period 2085 under RCP8.5. Despite those uncertainties, model chains consistently show an increase in frequency of all considered indices for the urban and the rural sites until the end of the century. Urban projections, though, show distinctly higher frequencies compared to the rural counterpart for most analyzed station couples and indices.

Further limitations are linked to the employed spatial transfer with QM (second-step QM). A first potential source of uncertainty are non-stationarities of transfer functions under current conditions (Hertig and Jacobeit, 2013) and in a changing climate (Christensen et al., 2008). The approach assumes the calibrated correction function, i.e. in our case the statistical relation between rural and urban site, to be stationary in time (Feigenwinter et al., 2018). This might not be valid until the end of the century for multiple reasons, such as changes of urban areas in terms of building densification, expansion of the urban area, changes in the building material, surface albedo, vegetation cover or anthropogenic heat release (see also Hoffmann et al., 2012). Directly related to this limitation is the fact that QM is based on data of two stations for each analyzed city; we acknowledge that such urban–rural comparisons provide little indication of spatial variations in UHI characteristics and are subject to uncertainties in terms of the station selection. Another source of uncertainty related to QM arises from the treatment of “new extremes”, i.e. values that lie outside the calibration range (Casaneuva et al., 2018; Ivanov et al., 2018). The QM implementation employed in this study uses a constant extrapolation of the correction function for the 1st and the 99th percentiles (Feigenwinter et al., 2018; Themeßl et al., 2012). Thus, it might be that the shape of the correction function in the last percentiles includes statistical artifacts in the future signals (Casaneuva et al., 2018). Yet, the modification of the climate change signals is advantageous in some cases (Gobiet et al., 2015) and constant extrapolation is a more robust approach compared to, for instance, linear extrapolation (Themeßl et al., 2012). A further source of uncertainty is that QM might misrepresent small-scale climate variability on short time scales (e.g. at daily scales). This is due to its deterministic nature during the spatial transfer, meaning a certain temperature value at the rural site always refers to a specific temperature value at the urban site. The urban–rural relation, however, is not constant over time, for instance, due to different weather conditions affecting e.g. urban ventilation through wind advection or surface radiative budget through cloud cover. The distribution of the observed urban–rural temperature difference thus reveals larger variability than the UHI based on quantile-mapped urban data. Lastly, the generation of urban scenarios with QM strongly depends on available observational data of both rural and urban sites for an overlapping time range of at least seven years (see Chapter 4.1). Especially in urban areas, though, long and high-quality measurements are rare. Other bias correction methods, such as trend-preserving approaches like quantile delta mapping (QDM) or the method from the third phase of ISIMIP (ISIMIP3), are sometimes preferred as preserving the climate change signal of the raw models might be an advantage. These methods, though, largely rely on the

quality of the observational reference used for calibration since the simulated signal is transferred to the observations to generate pseudo future observations, to which the quantile mapping is applied, i.e. they show a higher sensitivity to the considered observational dataset (Casaneuva et al., 2020b). In light of the generally short observational datasets available for urban sites, empirical QM is the preferred method for our specific study. Note however, that also other bias correction methods could be employed.

Limitations in terms of the first-step QM (downscaling, bias correction) are detailed in the CH2018 Technical Report (CH2018, 2018) and the literature referenced therein.

Summary and conclusions

This study introduces a method to generate multi-model ensemble scenarios for urban locations, which are often subject to short and (or) sparse data coverage, by means of a straightforward statistical approach. We focus on daily data of minimum (Tmin) and maximum temperature (Tmax) for three different greenhouse gas emission scenarios and multiple climate models. The data set is available for rural sites through the Swiss climate scenarios CH2018 (2018). In CH2018, a first-step quantile mapping (QM) approach has been used to bias correct and downscale simulated data to the rural sites. In this study, we apply a second-step QM procedure, which allows to spatially transfer these CH2018 data series to the urban target site, following the work of Rajczak et al. (2016) for permafrost applications. The resulting products are climate scenarios for five cities in Switzerland for Tmin and Tmax, available at daily resolution for four 30-year time periods (1981–2010, 2020–2049, 2045–2074 and 2070–2099) until the end of the 21st century. Comparing the temperature differences of an urban and an adjacent rural site (station couple) allows for station couple specific information on the respective UHI effect in the future, which we quantify in terms of temperature-based heat indices, namely the number of tropical nights (TN), summer days (SD) and hot days (HD).

Regarding the first-step QM, a large number of studies has already acknowledged its high potential to bias-correct and downscale climate model data (e.g. CH2018, 2018; Gudmundsson et al., 2012; Gutiérrez et al., 2018; Ivanov and Kotlarski, 2017; Monhart et al., 2018; Themeßl et al., 2012). We focus on the second step of the QM technique and validate it in an extensive evaluation framework. Validation results reveal a remarkable performance in the present-day climate with low biases and uncertainties (in terms of data sampling). The method's potential to generate climate projections at sparsely observed locations is helpful for climate impact studies across various research areas, its versatile application is not restricted to specific environments such as urban sites and results are easily transferable to further climate services. Still, our approach is associated with limitations and uncertainties that relate to both climate model projections themselves and the employed postprocessing methods. Model uncertainty can be substantial, especially for the late scenario period and RCP8.5. In terms of the applied method, limitations enclose, for instance, no explicit consideration of values that lie outside the calibration period. Moreover, QM implicitly assumes temporal stationarity of the urban–rural temperature relation, which might not be valid in the future, for instance, due to structural changes in the urban area.

According to the generated projections, climate change will have a major effect on all analyzed indices: results clearly show a strong increase in the number of TN, SD and HD events until the end of the 21st century, especially for the high emission scenario. Even though the mean warming is similar in urban and rural areas, most urban areas will be more strongly affected by rising temperatures than their rural surroundings in terms of fixed threshold exceedances, regardless of the considered period or emission scenario. In a RCP8.5 scenario with over 105 SD (multi-model median) and 45 TN each year (model-model median), the urban population of Zurich, for instance, would suffer from unprecedented heat stress on numerous days of the year by the end of the century. Due to slightly underestimated urban temperatures in the

validation setting, urban climate estimates are conservative and may be even higher in reality. Results for the adjacent rural site reveal distinctly lower frequencies (about 90 SD and 20 TN per year; multi-model median). Urban and rural projections mostly differ in terms of the daily minimum-based index (TN) and less strongly in terms of the indices based on daily maximum temperatures (shown in the number of SD or HD). The urban–rural temperature difference, i.e. the urban heat island (UHI) effect, being generally a nighttime phenomenon is visible throughout all scenario periods, RCPs and analyzed station couples. In addition, the occurrence of TN increases more strongly in the analyzed urban environments than in their rural counterparts due to statistical non-linearity in the exceedance frequency. The suggested increase poses a particularly high risk for human health as the body can cope less with high degrees if nighttime temperatures after a hot day are not falling below a certain level that would allow the body to recover.

The projected strong increases of both nighttime and daytime heat stress, especially in urban areas, reveals the urgent need to focus on the unique aspects of urban climate. This focus, though, should not exclusively lie on projections based on RCP8.5, as the worst-case scenario is not necessarily the most likely one (Hausfather and Peters, 2020). According to Hausfather and Peters (2020), overrating the probability of extreme climate impacts can make mitigation measures appear rather pointless and might lead to defeatism and despair. They thus propose a more realistic range of baseline scenarios, which potentially strengthen the assessment of climate risk. With the findings of the present study, we add important value by offering a wide-ranging quantification of (temperature-based) climatic conditions at selected urban sites in Switzerland until the end of the century, focusing not only on a high emission scenario but considering also a medium as well as a low emission future pathway. Still, as every scenario remains prone to uncertainties, adaptation measures should be robust under a wide range of possible future outcomes.

CRediT authorship contribution statement

Annkatri Burgstall: Formal analysis, Visualization, Writing - original draft. **Sven Kotlarski:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition. **Ana Casanueva:** Conceptualization, Methodology, Writing - review & editing, Supervision. **Elke Hertig:** Writing - review & editing. **Erich**

Fischer: Writing - review & editing. **Reto Knutti:** Writing - review & editing, Project administration.

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 authors acknowledge the fruitful collaboration between the University of Augsburg, ETH Zurich and MeteoSwiss in the frame of the present study. We further like to thank the CH2018 group (<https://www.nccs.admin.ch/nccs/en/home/climate-change-and-impact/swiss-climate-change-scenarios.html>) for technical support and scientific discussions. The authors wish to thank the Swiss National Supercomputing Centre (CSCS) for providing the technical infrastructure. We are grateful for the National Air Pollution Monitoring Network (NABEL), the Department of Environmental Sciences of the University of Basel, the Laboratoire énergie, environnement, architecture of the Haute école du paysage, d'ingénierie et d'architecture de Genève, hepia, and the Federal Office of Meteorology and Climatology MeteoSwiss for making available their observational data for this study. The authors also like to thank Jan Rajczak and Regula Gehrig (both at MeteoSwiss) for their very valuable scientific input. Finally, the authors are also grateful to the editor and two anonymous reviewers who helped to improve the original manuscript.

Financial support

This research has been partly supported by the European Commission (HEAT-SHIELD 668786). EH is supported by the German Research Foundation under project number 408057478.

Appendix A

Table A1

The employed GCM-RCM simulations with the respective initial condition member (init), for the different RCPs and the two horizontal resolutions. The 'x' marks the available simulations and the 'o' indicates the simulations that needed to be substituted by pattern scaling.

GCM	init	RCM	RCP8.5		RCP4.5		RCP2.6	
			0.11°	0.44°	0.11°	0.44°	0.11°	0.44°
ICHEC-EC-EARTH	r1i1p1	KNMI-RACMO22E		x		x		o
	r3i1p1	DMI-HIRHAM5	x		x		x	
	r12i1p1	CLMcom-CCLM4-8-17	x		x		o	
	r12i1p1	CLMcom-CCLM5-0-6		x		o		o
	r12i1p1	SMHI-RCA4	x		x		x	
MOHC-HadGEM2-ES	r1i1p1	CLMcom-CCLM4-8-17	x		x		o	
	r1i1p1	CLMcom-CCLM5-0-6		x		o		o
	r1i1p1	ICTP-RegCM4-3		x		o		o
	r1i1p1	KNMI-RACMO22E		x		x		x
	r1i1p1	SMHI-RCA4	x		x			o
MPI-M–MPI–ESM–LR	r1i1p1	CLMcom-CCLM4-8-17	x		x		o	
	r1i1p1	CLMcom-CCLM5-0-6		x		o		o
	r1i1p1	SMHI-RCA4	x		x			x
	r2i1p1	MPI-CSC-REMO2009	x		x		x	
MIROC-MIROC5	r1i1p1	CLMcom-CCLM5-0-6		x		o		o
	r1i1p1	SMHI-RCA4		x		x		x
CCCma-CanESM2	r1i1p1	SMHI-RCA4		x		x		o
CSIRO-QCCCE-CSIRO-Mk3-6-0	r1i1p1	SMHI-RCA4		x		x		o
IPSL-IPSL-CM5A-MR	r1i1p1	SMHI-RCA4	x		x		o	
NCC-NorESM1-M	r1i1p1	SMHI-RCA4		x		x		x
NOAA-GFDL-GFDL-ESM2M	r1i1p1	SMHI-RCA4		x		x		o

Source: Table modified from CH2018 (2018).

- Panofsky, H.A., Brier, G.W., 1968. Some Applications of Statistics to Meteorology. State University, College of Earth and Mineral Sciences, S.I. University Park, Penn.
- Ragetti, M.S., Vicedo-Cabrera, A.M., Schindler, C., Rösli, M., 2017. Exploring the association between heat and mortality in Switzerland between 1995 and 2013. *Environ. Res.* 158, 703–709. <https://doi.org/10.1016/j.envres.2017.07.021>.
- Rajczak, J., Kotlarski, S., Salzmann, N., Schär, C., 2016. Robust climate scenarios for sites with sparse observations: a two-step bias correction approach. *Int. J. Climatol.* 36 (3), 1226–1243. <https://doi.org/10.1002/joc.4417>.
- Roth, M., 2013. Urban heat islands. In: Fernando, H.J.S. (Ed.), *Handbook of Environmental Fluid Dynamics*, first ed. CRC Press, Boca Raton, FL, pp. 143–159.
- Salamanca, F., Martilli, A., Tewari, M., Chen, F., 2011. A study of the urban boundary layer using different urban parameterizations and high-resolution urban canopy parameters with WRF. *J. Appl. Meteorol. Clim.* 50, 1107–1128. <https://doi.org/10.1175/2010JAMC2538.1>.
- Scherer, D., Fehrenbach, U., Lakes, T., Lauf, S., Meier, F., Schuster, C., 2013. Quantification of heat-stress related mortality hazard, vulnerability and risk in Berlin, Germany. *Die Erde* 144, 238–259. <https://doi.org/10.12854/erde-144-17>.
- Sørland Lund, S., Fischer, A.M., Kotlarski, S., Künsch, H.R., Liniger, M.A., Rajczak, J., Schär, C., Spirig, C., Strassmann, K., Knutti, R., 2020. CH2018 - National climate scenarios for Switzerland: how to construct consistent multi-model projections from ensembles of opportunity. *Clim. Serv.* 20, 1–16. <https://doi.org/10.1016/j.cliser.2020.100196>.
- Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. *B. Am. Meteorol. Soc.* 93, 1879–1900. <https://doi.org/10.1175/BAMS-D-11-00019.1>.
- Theeuwes, N.E., Steeneveld, G.-J., Ronda, R.J., Holtslag, A.A.M., 2017. A diagnostic equation for the daily maximum urban heat island effect for cities in northwestern Europe. *Int. J. Climatol.* 37 (1), 443–454. <https://doi.org/10.1002/joc.2017.37.issue-110.1002/joc.4717>.
- Thiemebl, M.J., Gobiet, A., Heinrich, G., 2012. Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Clim. Change* 112 (2), 449–468. <https://doi.org/10.1007/s10584-011-0224-4>.
- Tzavali, A., Paravantis, J.P., Mihalakakou, G., Fotiadi, A., Stigka, E., 2015. Urban heat island intensity: a literature review. *Fresenius Environ. Bull.* 24, 4537–4555.
- Van der Schriek, T., Varotsos, K.V., Giannakopoulos, C., Founda, D., 2020. Projected future temporal trends of two different urban heat islands in Athens (Greece) under three climate change scenarios: a statistical approach. *Atmosphere* 11, 1–25. <https://doi.org/10.3390/atmos11060637>.
- Vogt, R., Parlow, E., 2011. Die städtische Wärmeinsel von Basel - tages- und jahreszeitliche Charakterisierung. *Regio Basiliensis* 52, 7–15.