Urban Climate 32 (2020) 100606

Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim

Hyper-local, efficient extreme heat projection and analysis using machine learning to augment a hybrid dynamical-statistical downscaling technique

Luke Madaus[a,*], Patrick McDermott[a], Joshua Hacker[a], Julie Pullen[b]

[a] Jupiter Intelligence, 1320 Pearl St., Suite 232, Boulder, CO, USA
[b] Jupiter Intelligence, 251 W. 30th Street, 13th floor, New York, NY, USA

ARTICLE INFO

Keywords:
Climate downscaling
Extreme heat

ABSTRACT

This paper describes a scalable system for quantifying hyper-local heat stress in urban environments and its expected response within the changing climate. A hybrid dynamical-statistical downscaling approach links Global Climate Models (GCMs) with dynamically downscaled extreme heat events using the Weather Research and Forecasting model (WRF). Downscaled historical simulations in WRF incorporate urban canopy physics to better describe localized land surface details in the urban environment relevant to extreme heat. This downscaled library is then used in an analog-based approach. This contribution reports a series of enhancements to existing analog-based methods which can efficiently produce more detailed results. The system here uses advanced statistical methods and simple machine learning (ML) techniques to optimize analog selection, perform spatially-consistent bias correction, and decompose patterns of extreme heat into dynamic components such as the land-sea contrast and inland sea-breeze penetration. Hindcast projections are validated against observational data from in-situ weather observing stations. The results demonstrate the scalability and efficiency of this system as it is deployed in cloud-based architectures with parallelized code. Downscaled predictions are equally applicable to heat stress at weather and climate time scales, supporting infrastructure resilience and adaptation, and emergency response.

1. Introduction

One of the most robust outcomes of the current changing climate is an expected increase in global air temperature (IPCC, 2013). The rate and pattern of change is expected to have considerable local variability, which must be taken into account when planning for the local impacts of climate change. Typically, simulations of the future climate are conducted with Global Climate Models (GCMs), which are better able to capture global and regional variations of $\mathcal{O}(> 100$ km) in climate as opposed to local changes of $\mathcal{O}(1–10$ km). Methods for interpreting the GCM projections in a local context are collectively referred to as *downscaling*, and a number of methods have been developed and applied. The selection of a downscaling method depends on the weather or climate features of interest, the specific geography, and resources available.

The aim of this work is to describe changes in extreme heat in an urban environment in a computationally efficient way. Primary requirements of the downscaling projections include:

* Corresponding author.
E-mail address: luke.madaus@jupiterintel.com (L. Madaus).

https://doi.org/10.1016/j.uclim.2020.100606
Received 30 April 2019; Received in revised form 23 January 2020; Accepted 8 February 2020

- Capturing statistics of extreme temperatures on a daily or sub-daily timescale
- Capturing heat spatial variability at a sub-urban scale, $\mathcal{O}$(1 km)
- Accounting for changes in magnitude and frequency of extreme heat events
- Easily incorporating multiple GCM projections
- Producing these projections in a computationally efficient way

Most downscaling methods can be classified as either statistical or dyanmical methods. Statistical methods rely on statistical relationships between observed temperature distributions and global climate projections to infer spatial variations in temperature and its change. Dynamical methods use a geophysical model to simulate temperature on a spatial grid subject to boundary conditions provided by a GCM or other coarse dataset. Aspects of both statistical and dynamical methods are attractive for heat downscaling. A dynamic geophysical model can capture the variety of daily evolutions of extreme heat in an urban environment in a physically consistent way, however the computational expense of running such models, particularly at high spatial resolution for multiple climate simulations, can quickly become large. Statistically-based methods tend to be more computationally efficient, but their ability to capture variability on spatial and temporal scales smaller and higher than the observations used to train them is less robust.

For these reasons, a hybrid dynamical-statistical approach is developed. A number of methods have been developed that attempt to blend aspects of statistical and dynamical downscaling methods, including, for example, a principal-component-based regression on a limited sample of dynamically downscaled simulations (Walton et al., 2015) and a number of analog-based techniques (e.g., Zorita and von Storch, 1999; Maurer et al., 2010; Pierce et al., 2014). We draw inspiration from these techniques, particularly the analog-based methods, in the development of a downscaling method to support this work.

Analog-based methods work by relating simulated future events in GCM projections to historical events that had a similar atmospheric evolution or pattern. In this way, historical events can be used as a "proxy" for what might occur during such a future event. Often, some sort of statistical scaling of the historical event is also applied to adjust for magnitude changes expected in a changing climate. The analog-based methods are able to meet a number of the requirements proposed above. The timescale used for matching events is flexible, allowing events to be selected based on daily or sub-daily characteristics. The frequency of extreme events is captured through the analog matching, while changes in the magnitude can be accounted for with statistical scaling. The analog finding method is easily applied to multiple GCM projections without significant computational resources.

To satisfy all the requirements here, a number of adjustments were made to existing analog-based techniques. These primarily include:

- While existing analog-based methods often rely on statistically-based reanalyses of fields like temperature as their analog proxies, existing datasets are typically (1) not available on the required spatial or temporal scales and (2) do not incorporate urban processes to fully reflect the evolution of heat in an urban area. Therefore, we use a dynamical weather model to generate dynamically-downscaled simulations of historical extreme heat days and validate these simulations against historical observations. By only simulating the warm season when extreme heat events can occur and focusing on a single urban area, the computational expense is reduced.
- Developing a machine-learning based method to optimally choose predictors for extreme heat days when doing the analog matching
- Incorporating multiple existing GCM projections and other dynamically downscaled simulations to inform both the analog selection and the statistical magnitude scaling for future events. This both builds upon existing work within the climate downscaling community and reduces expense.

This paper details the development of a hybrid statistical-dynamical downscaling method to characterize future urban heat, with an emphasis on the specific characteristics of the method listed above. The geographical application area, specific dynamical model, and datasets used are discussed in Section 2. A step-by-step discussion of how these models and datasets are applied in the method itself is described in Section 3. Outputs from this method focusing on extreme heat are shown in Section 4, with concluding thoughts in Section 5.

2. Application area, dynamic model and datasets

For this work, the New York City (NYC) area was chosen as the urban application area, due to its large population exposed to potential extreme heat consequences and a dynamic spatial variability in heating between urban, rural, and water areas that highlights the capabilities of this method.

2.1. Extreme heat in NYC

While not all cities may necessarily have a clear UHI effect (e.g., Scott et al., 2018), NYC has been the subject of intensive study of the urban heat island, dating back to the original observational evidence (Bornstein, 1968). Since then, numerous historical case studies have been conducted, along with studies of the implications of heatwaves and UHI for the energy sector (Ortiz et al., 2016). Heatwaves in the northeast U.S. are generally defined as a maximum daily temperature above 32.2 °C (90 °F) for three or more continuous days (Rosenzweig et al., 2009), and we use this temperature threshold here to guide our definition of extreme heat. Because heat events in the United States tend to be defined using the Fahrenheit scale, we have chosen to present many of our results

Table 1
Summary of WRF configuration used for downscaling simulations.

Feature	Scheme/source	Reference
Microphysics	WRF Single Moment, 3-class	Hong et al. (2004)
LW Radiation	Rapid Radiative Transfer Model	Mlawer et al. (1997)
SW Radiation	MM5/Duhdia	Dudhia (1989)
Surface Layer	Eta Similarity	Janjić (2001)
Land Surface	Noah Multi-Parameterization	Niu et al. (2011)
Urban Physics	3-category Urban Canopy Model	Kusaka and Kimura (2004)
Boundary Layer	Mellor-Yamada-Janjić scheme	Janjić (2001)
Cumulus Physics	Kain-Fritsch (20 km domain)	Kain (2004)
Land Use Data	National Land Cover Dataset	(Homer et al., 2015)

here in units of Fahrenheit for consistency. Some of the key findings of the literature have been that the UHI is amplified during heatwaves, with the effect being most pronounced in the afternoon (Meir et al., 2013; Ramamurthy et al., 2017) and in the overnight hours (e.g., Ortiz et al., 2018). The sea breeze consists of multiple complex circulation patterns in the NYC area and can offer ventilation from extreme heat (Thompson et al., 2007). Thus capturing the dynamical interaction of heatwaves, UHI and coastal effects necessitates detailed modeling at the local scale.

An increasing number of studies are exploring the impacts of future heatwave conditions on NYC. Some recent studies have employed a pseudo global warming (PGW) methodology (Tewari et al., 2018) or present short duration end of century simulation results (Ortiz et al., 2018). The latest New York City Panel on Climate Change report (NPCC3; González et al., 2019) primarily uses a statistical downscaling method that relies on coarse climate model deltas over the entire New York metropolitan area (at 1 degree resolution), and therefore is not able to well represent the coastal land/sea gradients and associated land-sea breeze circulations.

In contrast to these other studies, the framework described here utilizes full dynamic information from GCMs and is robust across future timescales. Furthermore, this approach is designed to capture the urban heterogeneity at fairly high resolution (1 km) by employing an urban canopy parameterization to represent urban processes in the dynamical downscaling. Other differences include our use of hourly output and actual future year values (compared with long-term 30-yr averages). NPCC3 does consider select time periods (2045—2049 and 2095—2099) of dynamical downscaling at 1 km using CESM forcing in part of their evaluation, but does not adopt an ensemble framework.

2.2. Dynamic downscaling model

To simulate how heat evolves on a local scale, we use the Weather Research and Forecasting model (WRF; Skamarock et al., 2019) version 4.0.1. WRF is the community standard model for regional and local weather simulation in the United States and allows great flexibility in choosing configurations tailored to local geographies and features of interest. Table 1 summarizes the configuration of WRF used in this study. The WRF single-layer urban canopy model is used in conjunction with National Land Cover Dataset (NLCD; Homer et al., 2015) to better represent the effects of the urban heat island in these simulations. Except where noted, the default configurations of these schemes in the WRF model were used.

The WRF allows for nested domains for gradual downscaling to km-scale resolutions. Here, we use a nested 20–4-1 km nesting scheme to downscale to 1 km grid spacing in an area surrounding NYC (Fig. 1; Table 2). To increase the fidelity in reproducing historical events, the outer-most (20 km) domain uses spectral nudging of the mass and temperature fields at wavenumber 3 spatial scales. This nudges features on the spatial scale of 400 km or greater in the 20 km outer-most domain to more closely match the parent reanalysis dataset that provides initial and boundary conditions to each simulation.

We use the WRF model built and deployed in a software container that has been optimized for running on cloud-based compute systems (e.g., Hacker et al., 2017). The model configuration, and particularly the spatial domain configuration described above, were designed to operate within the restrictions of this compute environment. However, using a containerized model allows massive deployment of parallel simulations on the cloud that can take advantage of specific compute resources and market-based pricing. The immediate availability of cloud resources allows large numbers of simulations to be performed within a few days and at reduced expense.

2.3. Datasets

This method is designed to be flexible and efficient in its implementation. Given that dynamical weather and climate simulations are typically computationally expensive, the method leverages existing published reanalyses, global climate model simulations (GCMs), and dynamically downscaled datasets where possible. Below the specific simulation datasets used in this implementation are described, but other similar reanalyses, GCMs and dynamically downscaled simulations can be substituted in as desired.

2.3.1. Reanalysis

To understand the current and historical climate and provide initial and boundary conditions for the WRF downscaled simulations, a reanalysis dataset is required. We use the ERA-Interim reanalysis (Dee et al., 2011) which provides 6-hourly full-atmosphere

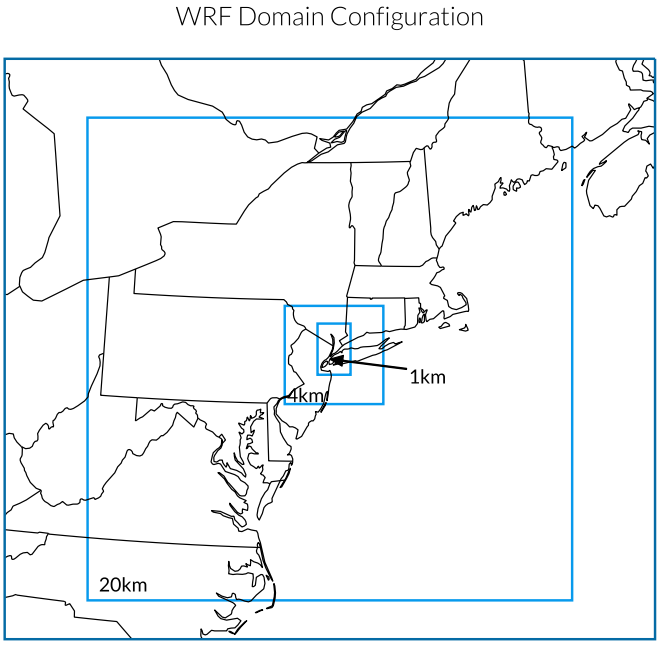

Fig. 1. WRF domain configuration used for downscaling simulations.

Table 2
Specific configuration of the geographical WRF domain.

Parameter	20 km parent	4 km nest	1 km nest
Central latitude	40.47°N	–	–
Central longitude	74.00°W	–	–
West-east gridpoints	60	61	81
South-north gridpoints	60	61	125
Parent *i* start	–	25	21
Parent *j* start	–	25	19
Num. vert. levels	50	50	50
Num. soil layers	4	4	4

analyses from 1979 through the present. The approach uses both the 6-hourly data (for downscaling simulations) and data aggregated to daily timescales (for analog matching purposes).

2.3.2. Global climate model simulations

An ensemble of future climate model simulations to guide the downscaling is one way to account for the uncertainty and natural variability in the climate system. The Community Earth System Model – Large Ensemble (CESM-LENS; Kay et al., 2015) is used here, as it includes 40 ensemble members that provide daily projections over historical (1920–2005) and future (2006–2080) time periods. The CESM-LENS uses the RCP 8.5 scenario as forcing for the future projections. To sample uncertainty due to differences in possible future climate forcings, the CESM-Medium Ensemble (Sanderson et al., 2017) is also considered. The medium ensemble branches from 15 members of the CESM-LENS in 2006, but uses the RCP 4.5 scenario as its climate forcing. Collectively, we refer to these datasets as the “CESM ensemble”.

2.3.3. Regionally downscaled climate simulations

GCM simulations like the CESM ensemble provide projected atmospheric changes at global to regional scales, $\mathcal{O}$(100 km); further downscaling is needed to understand local responses to climate change at $\mathcal{O}$(1–10 km). The computational expense to dynamically downscale all days for all years in all 55 members of the CESM ensembles would be rather large; it is telling that most dynamically downscaled simulations to $\mathcal{O}$(1–10 km) cover limited spatial areas, selected future years, few scenarios and/or use a limited number of GCMs as boundary conditions. But limited downscaled simulations exist, and we leverage those datasets in this method.

A 4-km grid spacing dynamically downscaled dataset over the continental United States (Liu et al., 2017) is available from the National Center for Atmospheric Research (Rasmussen and Liu, 2017). This dataset downscales 2000–2013 of ERA-Interim re-analysis, and then applies a “delta” calculated as the ensemble mean change of the Coupled Model Intercomparison Project v5 (CMIP5; Taylor et al., 2012) simulations for the end of century (2080–2100; RCP 8.5 scenario) and repeats these downscaling simulations with that delta added to the initial and boundary conditions for the downscaling model (also WRF). The result is a sample

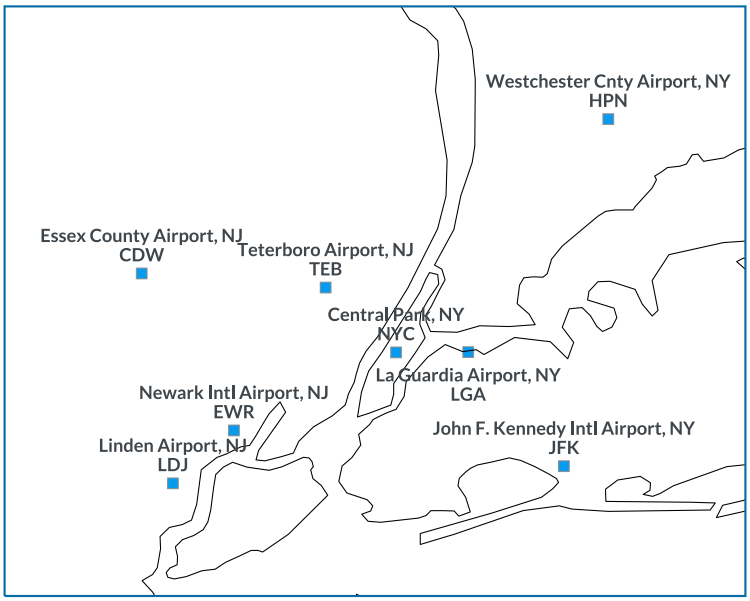

Fig. 2. Locations of GHCN observations used for verification.

of end-of-century climate change patterns on a local scale.

2.3.4. In-situ observations

Daily 2-m temperature observations from the Global Historical Climate Network (GHCN; Menne et al., 2012) support bias correction and hindcast verification. GHCN includes many observations from co-op observers which only report for a limited period of time and with uncertain data quality. Here we limit ourselves to GHCN observations with a long period of observation (> 10 year) within the NYC area or from known airport stations that also provide data through the Automated Surface Observing System (ASOS) network (Fig. 2).

3. Method

The downscaling method can be summarized in five steps:

1. A bias-corrected library of dynamically-downscaled warm-season simulations is produced
2. An analog finder is trained to match climate projection days to historical days
3. Downscaled output from matched historical extreme heat days is assigned to the climate projections
4. Heat-related variables in each matched downscaled simulation are scaled in a manner consistent with expected future warming
5. The output from these selected and scaled simulations is used to compute heat-relevant statistics on a local scale.

Below we describe each of these steps in more detail.

3.1. Generation of downscaled historical catalog and bias correction

Ideally, only days where extreme heat was observed to have occurred anywhere within the NYC area would need to be simulated to construct the historical catalog. However it is difficult to a priori identify such days with limited observations that do not observe the domain on the kilometer-scale we are simulating. Instead, a blanket simulation approach is chosen where all days each year within a four-month period (May through August) are simulated with WRF for all years in the ERA-Interim archive (1979–2017). In practice, each simulation is eight days long with one day of overlap between different simulations to account for model spinup. This simulation strategy was chosen for the following reasons:

- To complete these simulations efficiently in a reduced amount of time, simulations for different eight-day periods could be run in parallel
- As noted above, these simulations were conducted in a commercial cloud-computing system with pre-emptable compute nodes. This environment was chosen to reduce overall simulation costs, but comes with the risk of having simulations terminated prematurely. While a workflow using different large restart files could also have been used, having the simulations broken into smaller chunks reduced the need to restart potentially long-running simulations

- Though synoptic-scale nudging is applied to the outer-most domain, restarting the WRF simulations with new initial conditions taken from reanalysis reduces the chance of model drift in the interior domains with long-running simulations

Since the dynamical simulations represent the bulk of the cost in this method, we expand on the costs of our implementation here as an example of the required resources. Each of our overlapping 8-day WRF simulations can be completed in a matter of hours using the cloud-based WRF container. The simulation scheme described above requires 624 simulations, which can all be completed in parallel with scalable cloud resources in a single afternoon. The remainder of this method leverages inexpensive statistical methods and other existing datasets, which are freely available except the cost of their storage. By comparison, using the same configuration of WRF to dynamically downscale warm season historical and future days (1979 through 2100) for all 40 members of the CESM RCP 8.5 ensemble alone would require over 75,000 simulations, increasing the computational cost by a factor of 125×. While the relative benefits of pure dynamical downscaling versus using this hybrid method are beyond the scope of this work, the relative efficiency and cost-effectiveness of this approach is well-illustrated by this comparison.

Only three variables are required from these WRF simulations to construct the historical catalog for extreme heat: hourly 2-m temperature, daily maximum temperature, and daily minimum temperature. For simplicity, daily maximum and minimum temperatures are computed over 0000 UTC to 0000 UTC (following day) — a 24-h period. In the NYC warm season, this corresponds to 2000 through 2000 (following day) local daylight time, and we expect the daily maximum and minimum temperatures to be typically captured in this time range. The final catalog has about 5000 historical days.

We bias correct each of these variables separately using bias estimated from the GHCN observations. Though the GHCN observations used have a long period of record and are well-vetted, studies that have used supplemental observations over shorter periods have observed localized variations in the UHI throughout NYC that may not be well-captured by this observing network (e.g., Gedzelman et al., 2003; Gaffin et al., 2008; Ramamurthy et al., 2017; Ortiz et al., 2018). We attempt to compensate for the sparseness of GHCN observations in the domain by borrowing from the spatial patterns in heat from our large library of WRF simulations.

Bias maps are generated by using weighted regressions built from covariances estimated with the WRF library. Specifically, for daily maximum, minimum and average temperature, the WRF simulation outputs are used to build a covariance between each of those variables at the locations of the GHCN/ASOS observations and all other spatial points in the WRF simulation domain. Then, the WRF output is interpolated to the locations of these observations, and the WRF predicted values are compared with the observed values. A bias is calculated as the mean difference between the WRF prediction and the observed values at each station location. This bias is then "broadcast" to all other spatial points by using the covariance maps to build a regression from the location of the observation to all other gridpoints. The influence of each bias estimate is spatially localized to 50 km. We broadcast these biases estimated at all observation locations iteratively, repeating the the iterations 50 times while changing the order of the observations in each iteration. We take the mean of the resulting bias maps produced as our bias-correction map for that variable.

An example of the bias-correction map for maximum temperature computed in this way is shown in Fig. 3. Closer to the urban core (e.g., near Central Park (NYC) or LaGuardia (LGA)), the WRF has very little bias, while negative biases in maximum temperature are apparent at locations farther away. The final bias map blends and extends these bias estimates to surrounding gridpoints.

The impact of applying this bias correction is shown through changes in root-mean-squared-error (RMSE) in Fig. 4. For locations with relatively high bias and high RMSE (such as HPN), the impact of the bias correction can be substantial, reducing RMSE by over 50%. After bias correction, the RMSE in maximum temperature across all sites is fairly consistent at about 1.5 °F.

3.2. Analog downscaling procedure

A robust *analog finder* can match days in future climate simulations (here, the CESM ensemble) to similar historical days (here, in ERA-Interim). Determining the closeness of analog pairs can depend on any number of variables and parameters. Different choices for these parameters can lead to drastically different results. Previous data-driven approaches to selecting these parameters include both cross-validation (CV) techniques (Zhao and Giannakis, 2016) and Bayesian modeling (McDermott and Wikle, 2016; McDermott et al., 2018). We use an approach similar to CV by splitting the data (temporally) into a test and training set, and then select the parameters

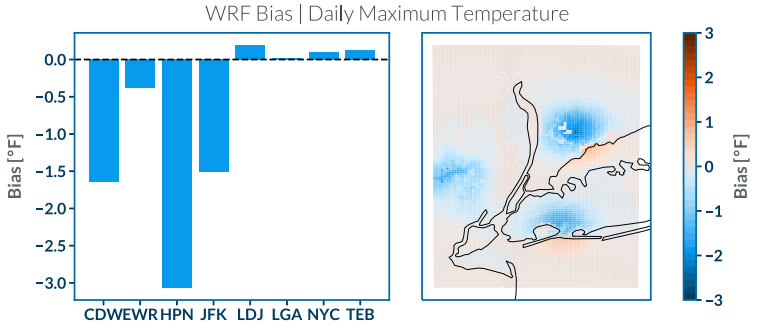

Fig. 3. WRF bias in maximum temperature for all simulated days (left). These bias estimates are then spread throughout the domain using a covariance-based statistical technique to produce bias maps that can be used to correct the WRF output at each grid point (right).

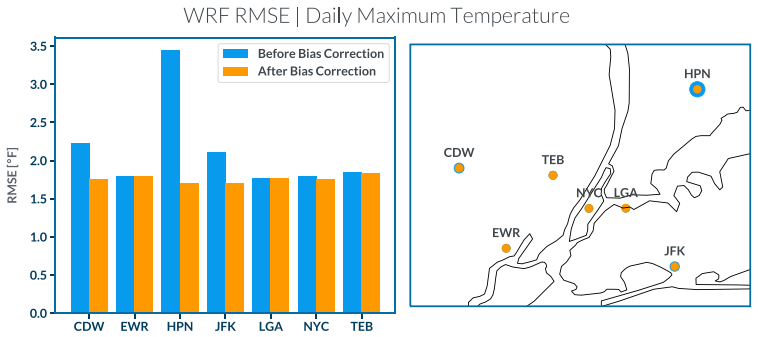

Fig. 4. WRF root-mean-squared-error (RMSE) in maximum temperature for all simulated days (left). RMSE is shown both before (blue) and after (orange) bias correction is applied. The map (right) shows the before and after RMSE estimates at each station location, with the size of the circle proportional to the magnitude of RMSE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that minimize the test data MSE. Four primary parameters are tested in the minimization: the extent of the spatial region used to find analogs, the center point of this spatial region, the distance metric used to rank analogs, and the climate variable(s) for identifying analogs.

The analog procedure utilized here is split into three steps. First, we use a CV approach to select parameters using a training and test data set. Next, we identify candidate analog days for both the ERA-Interim and the CESM data; this produces a library of CESM analog days. Finally, using the parameters selected in the first step and the candidate analog days from the second step, all of the candidate CESM ensemble analog days are matched to the closest ERA-Interim days to construct the analog catalog.

3.2.1. Analog parameter selection

The goal of the parameter selection is to find the parameter set that leads to the most accurate spatial downscaling. To evaluate a particular parameter set, we create a proxy high-resolution spatial temperature field to act as the "truth" using a bilinear interpolation scheme. The CESM data from 2000 to 2006 is interpolated to approximately the same resolution as the WRF grid, and this procedure is also repeated for all of the ERA-Interim training data (from 1979 to 1999). We will refer to these two datasets as the high-resolution ERA-Interim (HR-ERA) and CESM (HR-CESM) data, respectively. That is, the high-resolution CESM data is treated as the test data "truth" by creating an idealized scenario where one has access to the downscaled variable of interest (i.e., future CESM data), as shown in Fig. 5.

The analog parameter selection procedure is as follows: given a particular coarse-resolution CESM day from the test period (2000–2006), the most similar ERA-Interim training day (regridded to the CESM resolution) is selected as its analog. The corresponding interpolated version of this ERA-Interim day (HR-ERA) is treated as the "forecast." Finally, this forecasted HR-ERA analog day is compared with the HR-CESM test data for the test day of interest, and an error metric (MSE) is computed. This procedure is repeated for all of the CESM daily data in the test set, over a discrete grid for the four analog parameters listed above. Finally, the set of parameters that provide the lowest test MSE are retained.

The analog procedure implementation for NYC considered two possible variables to identify analogs, daily maximum temperature and geopotential height at 850 hPa. Moreover, we also considered two distance metrics, Euclidean distance and the Pearson correlation coefficient. The spatial extent was varied over a discrete grid ranging from 3 square grid cells (3 degrees in latitude and longitude) to 15 square grid cells (15 degrees in latitude and longitude). As discussed above, the parameter selection was conducted by treating data from 1979 to 1999 as training data and data held out from 2000 to 2006 as the test dataset. The final parameters selected during the analog selection optimization process were: a spatial region of 7×7 degrees centered approximately 140 km southeast of NYC, with daily maximum temperature being the best predictor variable and Euclidean distance the best closeness metric.

3.2.2. Identification of candidate analog days

Ideally the analog finder would be applied to all days in the CESM ensemble projections and all historical ERA-Interim days would be candidate matches, but finding all of these analog matches would be computationally expensive. Because the main concern here is extreme heat events, we are able to rule out many days in both datasets for inclusion in the analog finder. For the historical ERA-Interim days, the limitation is simple—we only chose to dynamically downscale historical warm-season days (May–August), so only days in those months may be potential matches in ERA-Interim.

The season for extreme heat events in the future is expected to expand beyond the traditional warm season months, and we choose a different thresholding approach to determine candidate analog days in the CESM ensemble. Fig. 6 illustrates how this threshold is chosen from ERA-Interim and the downscaled WRF maximum temperature catalog. For each historical day in the catalog, the maximum temperature for the ERA-Interim grid cell over Central Park is obtained. The days are binned by ERA-Interim maximum temperature (Fig. 6, blue histogram), and we record the fraction of the downscaled 1-km WRF domain where the (bias-corrected) maximum temperature exceeds a 90 °F (32.2 °C) threshold on each of these days (Fig. 6, black line). This threshold (90 °F / 32.2 °C) is

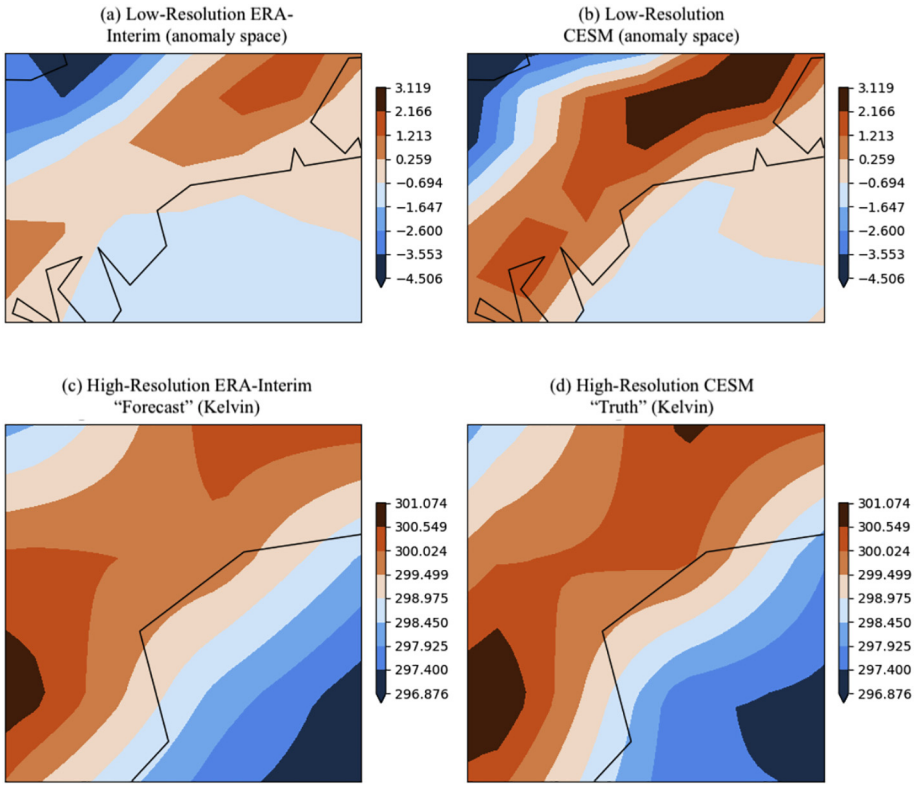

Fig. 5. Representative example of the analog parameter selection procedure using both low-resolution and high-resolution ERA-Interim and CESM data. (a) A low-resolution analog from the CESM data that was found to be the most similar to the CESM data spatial field of interest (shown in (b)); (b) The CESM data spatial field of interest. (c) The high-resolution (interpolated version) of the CESM analog identified in (a); (d) The "true" downscaled CESM spatial field for the spatial field of interest.

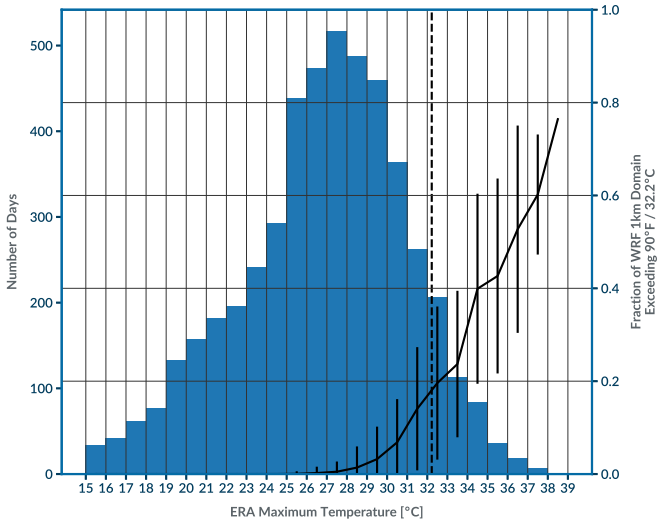

Fig. 6. Histogram of ERA-Interim daily maximum temperatures for the grid cell over Central Park for all historical days downscaled with WRF (blue histogram; day counts on left axis, total of 4914 days). The fraction of the bias-corrected WRF daily maximum temperature grid cells on each day that exceed 90 °F (32.2 °C), binned by the ERA-Interim maximum temperature values are also shown (black line). The maximum and minimum WRF domain fraction exceeding this threshold for each bin are shown with vertical black bars. The vertical dashed line highlights the 90 °F / 32.2 °C threshold. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

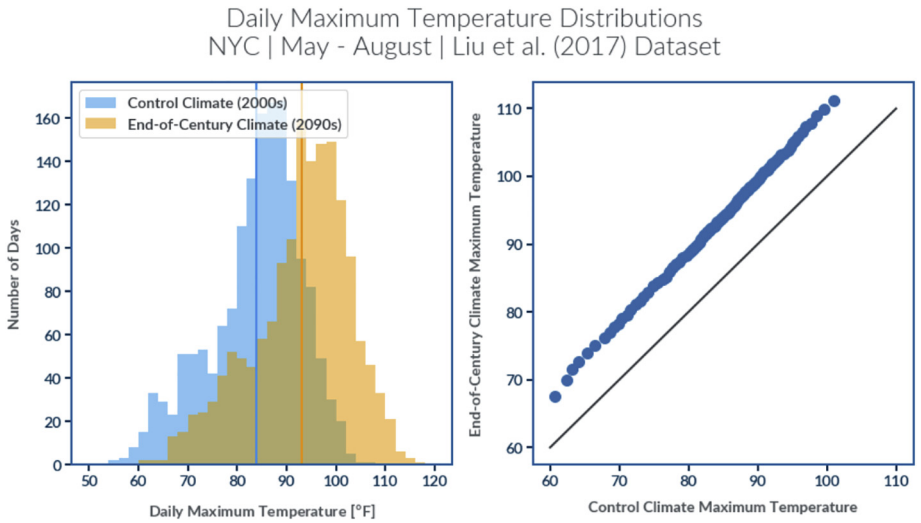

Fig. 7. Distributions of maximum temperature at Central Park (NYC) over both the current (orange) and end-of-century (blue) periods from the Liu et al. (2017) dataset. Shown as histograms (left) and as a quantile-quantile plot (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

highlighted as a vertical dashed line. There are many days when the ERA maximum temperature is below this threshold but a portion of the WRF maximum temperature grid cells exceeds this threshold. To better choose days where extreme heat is a possibility anywhere within the WRF simulation domain, we use a threshold of 23 °C (73.4 °F) applied to the maximum temperature for the CESM grid cell over New York City to identify candidate future extreme heat dates.

3.2.3. *Matching analog days*

Prior to identifying candidate analog days and doing the analog matching, the CESM data mean and standard deviation are bias-corrected with respect to the ERA-Interim data over the overlapping historical period.

The analog selection based on the selected parameters and candidate analogs is conducted for all candidate days in the CESM ensemble. As is common in the literature, analogs are identified in anomaly space (e.g., McDermott and Wikle, 2016; Zhao and Giannakis, 2016). Anomalies for the CESM data are calculated across ensemble members for a particular year. Calculating anomalies with respect to a given year allows for proper comparisons of CESM data far into the future, when temperatures have drastically increased, with the ERA-Interim data.

3.3. *Climate scaling*

Once appropriate analogs have been identified from the historical catalog, the assumption is that the spatial and temporal patterns of temperature on those days are a good proxy for what would be observed on those particular days in the climate simulations. However, the underlying distribution of temperature is expected to shift in future years. Even if a similar spatial and temporal pattern were to occur, the magnitudes of temperature will be different. To account for this, we employ a statistical *scaling* to the chosen analogs that is computed with both the CESM ensemble predictions and the local pattern of warming given by the Liu et al. (2017) downscaled dataset. This is similar to the "amplitude matching" applied in methods like LOCA (Pierce et al., 2014).

First, the end-of-century and current-climate distributions of temperatures are computed for the New York City area from the Liu et al. (2017) dataset. Daily average, maximum, and minimum temperatures are computed for May–August. Fig. 7 shows an example of the maximum temperature distribution for these months calculated for the Liu et al. (2017) gridpoint over Central Park (NYC). While more sophisticated methods like quantile mapping were explored, a simple shift in the mean of the distribution (a "delta") accounted for the majority of the change.

These Liu et al. (2017) dataset values were regridded from their 4-km grid to our 1-km WRF grid using bilinear interpolation. Then, this mean delta (from "current" to "end-of-century" climate) was computed at all gridpoints. Fig. 8 shows a map of the mean delta for maximum temperature. The difference in land-sea temperature delta is clear, but the spatial variability in warming over land is smaller.

While this provides maps of the rate of warming over the NYC area with kilometer-scale spatial resolution, the Liu et al. (2017) dataset only provides two points in time—"current" and "end-of-century". To extrapolate these changes to individual years, we draw inspiration from the work of Walton et al. (2015). In their downscaling experiments, they perform dynamical downscaling in southern California for only a limited number of future years, then linearly relate the spatial patterns of temperature change in the downscaled fields to large-scale predictors in the parent global climate models. This allows them to infer spatial patterns of temperature change for years and global climate models that were not downscaled.

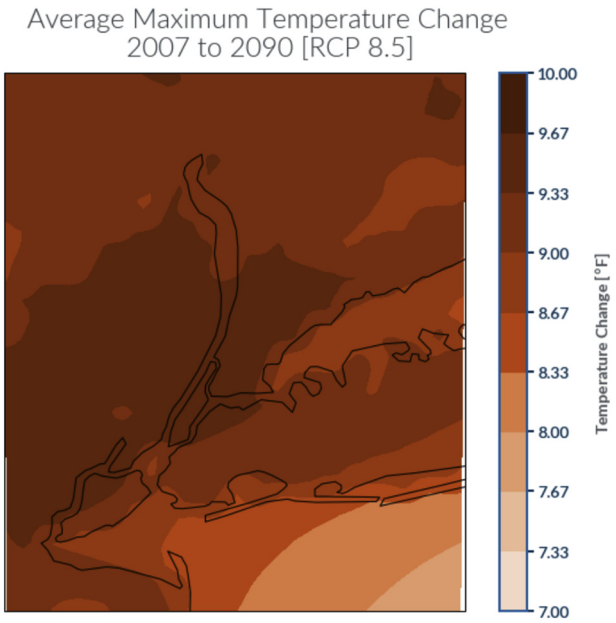

Fig. 8. Mean maximum temperature deltas from current to end-of-century areas in the Liu et al. dataset over the New York City WRF domain.

For this work, we once again consider the CESM ensemble, and relate the coarse CESM projections of warming over the NYC area to the higher-resolution spatial pattern of warming indicated by the Liu et al. (2017) dataset. For a large-scale predictor, we calculate maximum, minimum and average temperatures for the May–August period for each ensemble member and each year in the projections (1920 through 2090). The means of each of these distributions are recorded ($\overline{T}_{mem,yr}$). We then choose values to correspond to the two time periods downscaled by the Liu et al. (2017) dataset as "reference" values: the distribution means from 2006 to represent the "current" time period and those in 2090 to represent the "end-of-century" period. Taking the difference in the ensemble means of these values in 2090 ($\overline{T}_{ens,2090}$) relative to 2006 ($\overline{T}_{ens,2006}$) gives us a "reference scaling factor" for the CESM ensemble. Then the difference in the distribution mean from each ensemble member and year relative to 2006 is computed relative to this "reference scaling factor" to produce a scaling factor. Eq. 1 describes this computation of the scaling factor:

$$\Delta_{mem,yr} = \frac{\overline{T}_{mem,yr} - \overline{T}_{mem,2006}}{\overline{T}_{ens,2090} - \overline{T}_{ens,2006}}. \tag{1}$$

Fig. 9 shows these scaling factors for each ensemble member as a function of year from 2006 to 2080. These scaling factors could be thought of a "fraction of the way" from the 2006 values to the 2080 values relative to the ensemble mean change.

When a WRF simulation is chosen as an analog, this bias-corrected field from that simulation ($\mathbf{T}_{hist}$) is adjusted by adding to it the spatial warming field from the Liu et al. (2017) dataset (Δ) scaled by the appropriate scaling factor the particular ensemble member

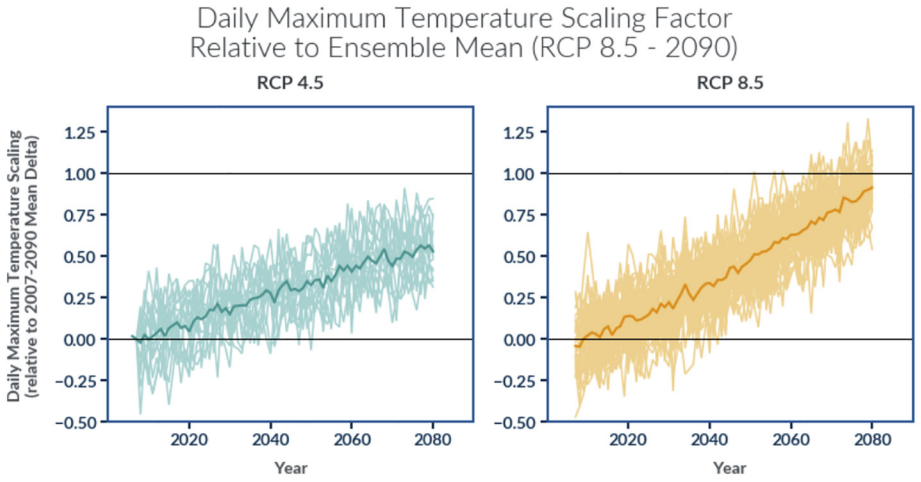

Fig. 9. Scaling factors computed for CESM ensemble members in the medium ensemble (RCP 4.5, left) and large ensemble (RCP 8.5, right).

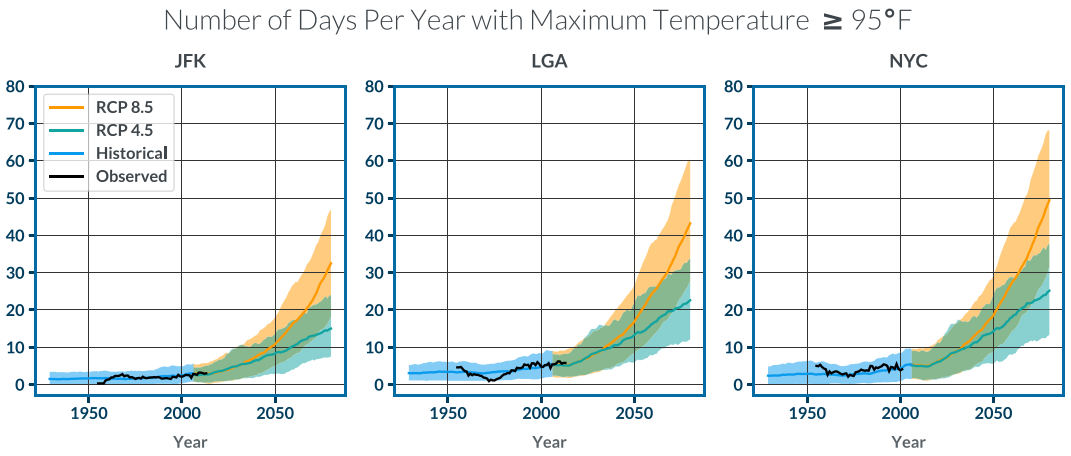

Fig. 10. Annual number of days where the maximum temperature is at or above 95 °F. Blue represents the historical simulation period in the CESM-LENS, and after 2006 both the CESM-LENS RCP 8.5 (orange) and CESM medium ensemble RCP 4.5 (green) projections are shown. Solid lines indicate the ensemble mean with observed counts in black. A 10-year running mean has been applied to the observed and ensemble mean values. The shaded ranges represent 10th to 90th percentile confidence intervals as estimated from the ensemble member projections. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and year for which this analog is drawn ($\delta_{mem,\ yr}$). Eq. 2 illustrates this process:

$$\mathbf{T}_{mem,yr} = \mathbf{T}_{hist} + \delta_{mem,yr} \times \Delta. \tag{2}$$

This is repeated for all analogs chosen. For the CESM Medium Ensemble (RCP 4.5), the reference scaling factor is still taken as the reference factor calculated from the Large Ensemble members (RCP 8.5), as the goal is to compute the scaling relative to the Liu et al. (2017) dataset, which also uses RCP 8.5.

3.4. Aggregation of heat statistics

When all scaled analog WRF simulation fields have been collected for all years and ensemble members, summary statistics are calculated on the WRF grid to evaluate the changing distributions of extreme heat. Uncertainties are estimated by computing distributions across the CESM ensemble members. We also separate the RCP 4.5 and RCP 8.5 members, to illustrate the differences based on the emissions scenario considered. Further analysis using machine-learning-based decomposition methods such as Principle Component Analysis (PCA) is also performed and described in Section 4.5.

4. Results

4.1. Verification

We leverage the long historical simulations of the CESM ensemble to validate this technique against in-situ observations. To do this, we compare the estimated historical climatological distribution of values produced using the CESM ensemble and this downscaling method with actual observed values at several locations. Fig. 10 shows the number of days at or above 95 °F at John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA) and Central Park (NYC) in the historical, RCP 8.5 and RCP 4.5 projections. Observed values are also shown in black. A 10-year running mean has been applied to the ensemble mean projections and observed values to facilitate easier comparison. The historical projections show good agreement with the observed counts at this location. Even for years prior to 1979 (the beginning of the historical downscaled WRF library), good agreement between the back-projected values and observations is evident. Spatial variability in these counts is captured; in particular the coastal location and influence of the sea breeze on JFK is reflected in the fewer number of days exceeding the 95 °F threshold in both observations and the projections. We note that the WRF simulations include a fixed representation of urbanized land in all simulations (using 2011 NLCD values), and this should be considered when looking at more historical time periods when urban development in the vicinity of these locations was not the same. In addition, we cannot validate future projections since there are no observations available from these future times, but this is a limitation of all future projection methods.

4.2. Number of days per year exceeding threshold temperatures

A wide variety of variables can be derived from this data, and here we show a sample of the possible output projections. Maps of the ensemble mean number of days expected to exceed 95 °F are shown in Fig. 11. The structure of the urban heat island is apparent, with greater numbers of days exceeding these thresholds evident in central NYC and adjoining areas of New Jersey. These areas

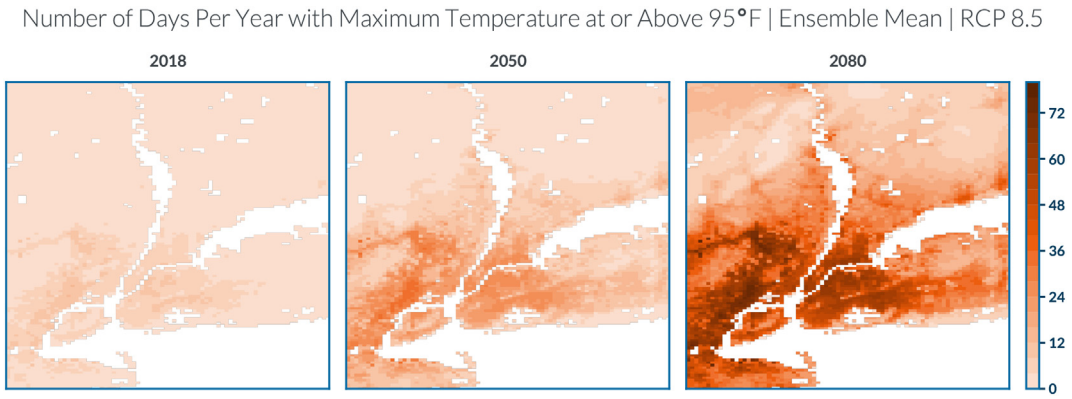

Fig. 11. Projected annual number of days where the maximum temperature is at or above 95 °F. These maps represent the ensemble mean values from the RCP 8.5 scenario for 2018, 2050 and 2080.

match the pattern of warming expected under the influence of the combined effects of UHI and extreme heat (e.g., Meir et al., 2013; Ramamurthy et al., 2017). The WRF resolution is sufficient to resolve large park and greenland belts throughout the NYC area, including Central Park on Manhattan and stretching through Brooklyn and Queens on western Long Island. More rural areas north of the New York City core urban area in Westchester County and along the Hudson River show fewer days exceeding this threshold.

4.3. Multi-day heat events

By considering each CESM ensemble member as a different possible trajectory through the future climate, each member time-series and its associated scaled analogs can be evaluated for day-to-day variability in extreme heat. This allows projections of multi-day heat events. Multi-day heat events can amplify the impact of extreme heat, particularly in areas such as human health impact and energy demand. Fig. 12 shows projections of the average annual number of events with two or more consecutive days at or above 95 °F using this method. While currently these events only happen three or fewer times per year, this number is projected to increase by 2050 to approximately 8–10 events per year within the urbanized areas, and further to 15–20 events per year by the end of the century in those areas.

4.4. Sub-daily heat statistics

WRF simulations with hourly output enable future projections that include the diurnal cycle. Understanding the duration of time over which certain temperature thresholds are being exceeded can be critical for projecting wear and maintenance needs on infrastructure. Fig. 13 illustrates the number of hours expected to exceed 95 °F on days where any hour is above that threshold. The variability of this statistic is more even across the area, though there are decreases in the immediate vicinity of the coasts. In 2018, most days where 95 °F is reached only exceed that temperature for 2–3 h. This is projected to increase to an average of 5–6 h by the end of the century.

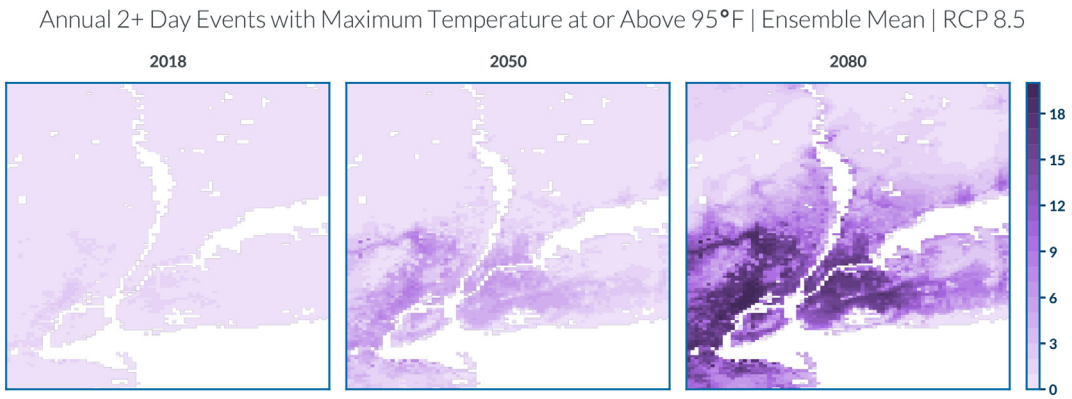

Fig. 12. Projected annual number of days where the maximum temperature is at or above 95 °F for at least 2 consecutive days. These maps represent the ensemble mean values from the RCP 8.5 scenario for 2018, 2050 and 2080.

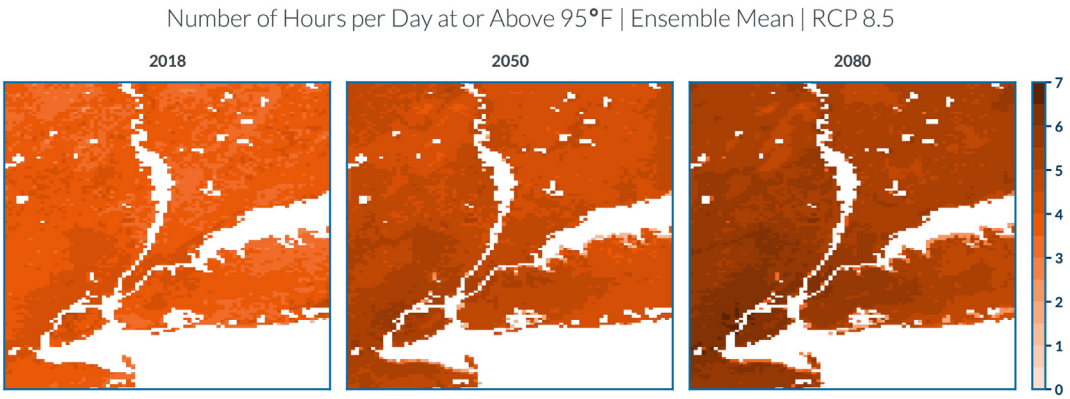

Fig. 13. Projected average number of hours that the temperature is above 95 °F on days where the maximum temperature is at or above 95 °F. These maps represent the ensemble mean values from the RCP 8.5 scenario for 2018, 2050 and 2080.

4.5. Preferred patterns of extreme heat events

This downscaled dataset provides a unique opportunity to understand not only how the raw counts of extreme days might change, but also how the evolution of extreme heat events may also differ in the future. To explore this, we perform a principle component analysis (PCA) on the daily maximum temperature fields for the entire historical WRF downscaled catalog. Fig. 14 shows the spatial patterns of the leading three modes identified by the PCA. Fig. 15 shows the fraction of the total variance in maximum temperature patterns explained by each mode. The dominant mode represents an overall land-sea contrast, but modes 2 and 3 physically resemble the influence of the sea-breeze front and its penetration onto Long Island and across the mainland (e.g., see Thompson et al., 2007).

We then examine how these historical heat days are resampled by the analog finder to explore how these different patterns change. In Fig. 16, the average magnitude of each of the three leading modes is evaluated for each future year and CESM-LENS (RCP 8.5) ensemble member based on which analog days are sampled and the magnitude of each PC mode on each of those days. Additionally, a series of reference timeseries are bootstrapped from the original WRF catalog, with the assumption that future analog days are randomly being selected.

Since the analog selector only functions on candidate extreme heat days while the WRF catalog spans the entire warm season, patterns conducive to warmer-than-normal days are preferentially being selected (i.e., the average PC magnitudes on the selected days are not necessarily centered on zero). PC1 and PC2 tend to have positive magnitudes throughout, which reflects a large number of days selected with (1) amplified land ocean contrast (PC1) and (2) mild sea-breeze penetration (PC2). The average magnitude of PC3 remains near zero for much of the historic period, indicating no strong preference on these days for deeper sea-breeze penetration.

For years after 2050, a shift in the types of analog days being selected is evident. The average magnitudes of PC1 and PC2 become even larger, and exceed the estimated range of magnitudes expected if the same types of days chosen historically continued to be randomly selected (black lines). This indicates a statistically significant shift to more days with larger land-ocean contrasts and sea-breeze events. The magnitude of PC3 becomes more positive, suggesting that deep sea-breeze penetration becomes more likely as the climate warms.

5. Conclusions

This contribution outlined a novel method for performing local downscaling of extreme heat projections in possible future climates. This method draws elements from other established downscaling methods in a way that capitalizes on the benefits of both

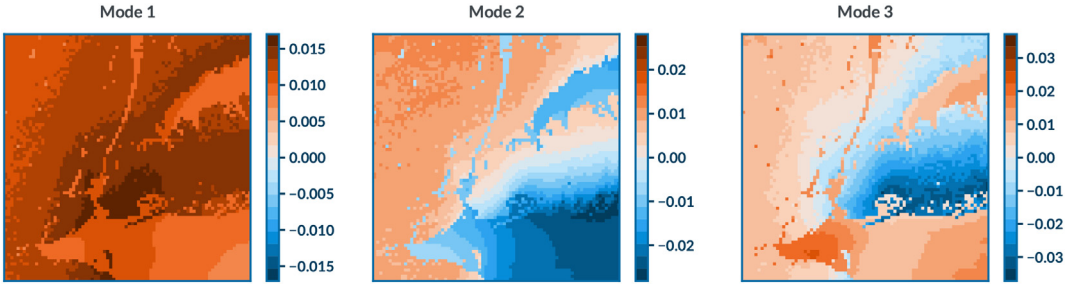

Fig. 14. The spatial patterns of the three leading principle components that explain the most variance in the maximum temperature fields. Units have not been normalized.

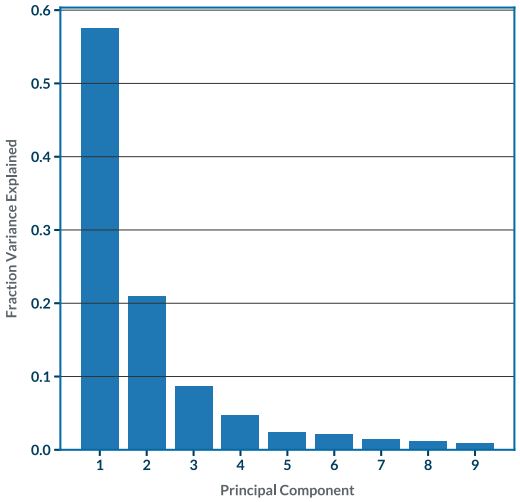

Fig. 15. The fraction of the total variance in maximum temperature explained by the first 9 principle components of the maximum temperature field decomposition.

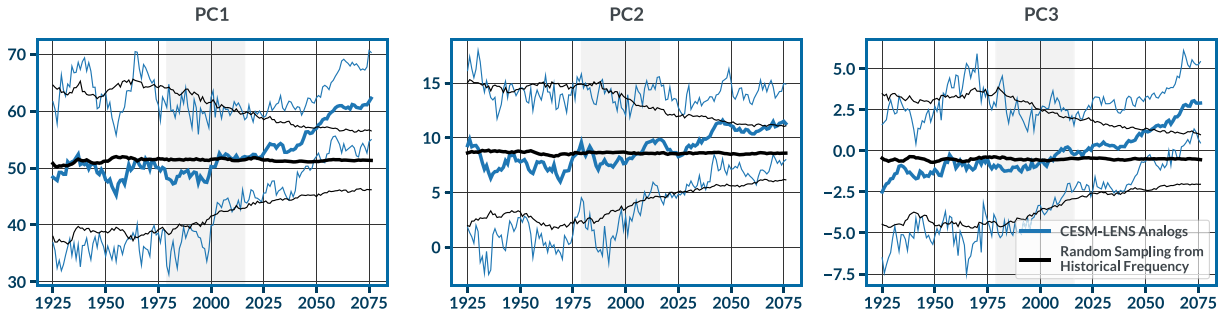

Fig. 16. The average magnitude of PCs from each analog event selected in a given year. The ensemble mean and 10th to 90th percentiles from the CESM-LENS member analogs are shown in blue with a 10-year running mean applied. In black, a bootstrap resampling of the PC magnitudes on analog days selected during the historical period (1979–2017, grey shading) is shown. The number of resampled days drawn each year is drawn from the distribution of the number of candidate extreme heat days identified in CESM-LENS for each year. No scaling was applied to compute the magnitudes of PCs on selected days. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

statistical and dynamical downscaling techniques, while balancing efficiency and computational cost. By leveraging existing global climate and regional dynamical downscaling simulations, computational resources can be targeted at capturing the evolution of extreme heat events on very local scales. A series of historical simulations, focused on the particular area of interest, allows calibration against historical observations and forms a core analog library that can be efficiently reused for climate projections. Statistical and machine learning techniques are used to blend the existing global and regional climate downscaling simulations with this localized analog library.

The results show the benefits of having a historical library to validate against, as good agreement is achieved with observations without any posterior bias correction applied to the localized projections. The calibrated results for the NYC area highlight how this method captures the urban heat island and can resolve many localized features of the landscape, including waterways and parklands. Projections into the future with this technique show a significant increase in the frequency and duration of extreme heat both on a daily basis and in the context of multi-day events. The impacts are magnified in the urban heat island. Machine learning techniques like principle component analysis may also be applied to this dataset to understand how the dynamics of extreme heat days may change in the future. By applying PCA here, we find evidence for an increased number of deep sea-breeze penetration events by the end of the 21st century in the New York City area, which may help to offset some of the possible extreme heat events.

Recently, Ortiz et al. (2018) conducted 1-km dynamical downscaling simulations over the NYC area where they used expected climate conditions from brief, five-year periods in the middle (2045–2049) and end (2095–2099) of the 21st century. Their dynamical simulations also suggest the importance of sea-breeze as the largest dynamical contributor to changes in the spatial variability of the urban heat pattern. The analog selection and scaling method we employ can be directly applied to a wide variety of future climate projections without the added cost of additional dynamical simulations for every climate projection.

One important component in simulating the spatial variability of extreme heat is the accurate portrayal of land surface

characteristics. While we note that we use NLCD (2011) land use data, this is a static snapshot of land use in time. To increase accuracy of future projections, a method for incorporating the effects of possible land use changes could be adopted. Recent work has attempted to explore this effect, including in the context of WRF simulations (e.g., Adachi et al., 2012; Georgescu et al., 2012; Krayenhoff et al., 2018; Argüeso et al., 2015; Lemonsu et al., 2015; Tewari et al., 2018). We plan to consider methods for including land use changes in future expansion of this work.

This downscaling technique may be further enhanced with additional postprocessing to map heat variability to even finer scales in the urban environment. Recent studies have employed satellite and in-situ observations at high spatial resolution to better map localized variations in heat across urban areas (Voelkel and Shandas, 2017; Shandas et al., 2019). A subsequent paper will describe an extension of the method described here that employs machine learning techniques and satellite-derived quantities to remap extreme heat projections to even finer scales.

Declaration of Competing Interest

None.

References

Adachi, S.A., Kimura, F., Kusaka, H., Inoue, T., Ueda, H., 2012. Comparison of the impact of global climate changes and urbanization on summertime future climate in the Tokyo metropolitan area. J. Appl. Meteorol. Climatol. 51, 1441–1454.

Argüeso, D., Evans, J.P., Pitman, A.J., DiLuca, A., 2015. Effects of city expansion on heat stress under climate change conditions. PLoS One 10.

Bornstein, R.D., 1968. Observations of the urban Heat Island effect in New York City. J. Appl. Meteorol. 7 (4), 575–582.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.G., Haimberger, L., Healy, S.B., Hersback, H., Holm, E.V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, N., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, G.K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.N., Vitart, F., 2011. The erainterim reanalysis: configuration and performance of the data assimilation system. Quart. J. Roy. Meteor. Soc. 137, 553–597.

Dudhia, J., 1989. Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. J. Atmos. Sci. 46, 3077–3107.

Gaffin, S.R., Rosenzweig, C., Kahnbilvardi, R., Parshall, L., Mahani, S., Glockman, H., Goldberg, R., Blake, R., Slosberg, R.B., Hillel, D., 2008. Variations in New York City's urban heat island strength over time and space. Theor. Appl. Climatol. 94, 1–11.

Gedzelman, S.D., Austin, S., Cermak, R., Stefano, N., Partridge, S., Quesenberry, S., Robinson, D.A., 2003. Mesoscale aspects of the urban heat island around New York City. Theor. Appl. Climatol. 75, 29–42.

Georgescu, M., Moustaoui, M., Mahalov, A., Dudhia, J., 2012. Summertime climate impacts of projected megapolitan expansion in Arizona. Nat. Clim. Chang. 3, 37–41.

González, J.E., Ortiz, L., Smith, B.K., Devineni, N., Colle, B., Booth, J.F., Ravindranath, A., Horton, R., Towey, K., Kushnir, Y., Manley, D., Bader, D., Rosenzweig, C., 2019. New York City panel on climate change 2019 report chapter 2: new methods for assessing extreme temperatures, heavy downpours and drought. Ann. N. Y. Acad. Sci. 1439, 30–70. https://doi.org/10.1111/nyas.14007.

Hacker, J.P., Exby, J., Gill, D., Jimenez, I., Maltzahn, C., See, T., Mullendore, G., Fossell, K., 2017. A containerized mesoscale model and analysis toolkit to accelerate classroom learning, collaborative research and uncertainty quantification. Bull. Amer. Meteor. Soc. 98, 1129–1138.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States – representing a decade of land cover change information. Photogramm. Eng. Remote. Sens. 81, 345–354.

Hong, S.Y., Dudhia, J., Chen, S.H., 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon. Weather Rev. 132, 103–120.

IPCC, 2013. Climate change 2013: the physical science basis. In: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.

Janjić, Z.I., 2001. Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model. In: Technical Report 437. National Centers for Environmental Prediction Office Note.

Kain, J.S., 2004. The Kain-Fritsch convective parameterization: an update. J. Appl. Meteorol. 43, 170–181.

Kay, J.E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S., Danabasoglu, G., Edwards, J., Holland, M., Jushner, P., Lamarque, J.F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., Vertenstein, M., 2015. The community earth system model CESM large ensemble project: a community resource for studying climate change in the presence of internal climate variability. Bull. Amer. Meteor. Soc. 96, 1333–1349.

Krayenhoff, E.S., Moustaoui, M., Broadbend, A.M., Gupta, V., Georgescu, M., 2018. Diurnal interaction between urban expansion, climate change and adaptation in US cities. Nat. Clim. Chang. 8, 1097–1103.

Kusaka, H., Kimura, F., 2004. Coupling a single-layer urban canopy model with a simple atmospheric model: impact on urban heat island simulation for an idealized case. J. Meteor. Soc. Japan 82, 67–80.

Lemonsu, A., Viguie, V., Daniel, M., Masson, V., 2015. Vulnerability to heat waves: impact of urban expansion scenarios on urban heat island and heat stress in Paris (France). Urban Clim. 14, 586–605.

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A.J., Prein, A.R., Chen, F., Chen, L., Clark, M., Dai, A., Dudia, J., Eidhammer, T., Gochis, D., Gutmann, E., Kurkute, S., Li, Y., Thompson, G., Yates, D., 2017. Continental-scale convection-permitting modeling of the current and future climate of north america. Clim. Dyn. 49, 71–95.

Maurer, E.P., Hidalgo, H.G., Das, T., Dettinger, M.D., Cayan, D.R., 2010. The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamow in California. Hydrol. Earth Syst. Sci. 14, 1125–1138.

McDermott, P.L., Wikle, C.K., 2016. A model-based approach for analog spatio-temporal dynamic forecasting. Environmetrics 27.

McDermott, P.L., Wikle, C.K., Millspaugh, J., 2018. A hierarchical spatiotemporal analog forecasting model for count data. Ecol. Evol. 8, 790–800.

Meir, T., Orton, P., Pullen, J., Holt, T., Thompson, W., Arend, M., 2013. Forecasting the New York City urban heat island and sea breeze during extreme heat events. Wea. Forecast. 28, 1460–1477.

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. An overview of the global historical climatology network-daily database. J. Atmos. Ocean. Technol. 29, 897–910. https://doi.org/10.1175/HTECH-D-11-00103.1.

Mlawer, E.J., Toubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the long-wave. J. Geophys. Res. 102D, 16663–16682.

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res. 116D.

Ortiz, L.E., González, J.E., Gutierrez, E., Arend, M., Legbandt, T., Neufeld, S., 2016. Forecasting building energy demands for New York City with a coupled weather-building energy model. In: Proceedings of ASME 2016 10th International Conference on Energy Sustainability collocated with the ASME 2016 Power Conference and the ASME 2016 14th International Conference on Fuel Cell Science, Engineering and Technology, pp. V001T11A004–V001T11A004.

Ortiz, L.E., González, J.E., Lin, W., 2018. Climate change impacts on peak building cooling energy demand in a coastal megacity. Environ. Res. Lett. 13 (9) 094008.
Pierce, D.W., Cayan, D.T., Thrasher, B.L., 2014. Statistical downscaling using localized constructed Analogs (LOCA). J. Hydrometeorol. 15, 2558–2585.
Ramamurthy, P., Li, D., Bou-Zeid, E., 2017. High-resolution simulation of heatwave events in New York City. Theor. Appl. Climatol. 128 (1–2), 89–102.
Rasmussen, R., Liu, C., 2017. High resolution WRF simulations of the current and future climate of North America. In: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, https://doi.org/10.5065/D6V40SXP. accessed: 2018-05-30.
Rosenzweig, C., Solecki, W.D., Parshall, L., Lynn, B., Cox, J., Goldberg, R., Hodges, S., Gaffin, S., Slosberg, R.B., Savio, P., Dunstan, F., Watson, M., 2009. Mitigating New York City's heat island: integrating stakeholder perspectives and scientific evaluation. Bull. Amer. Meteor. Soc. 90, 1297–1311.
Sanderson, B.M., Xu, Y., Trebaldi, C., Wwhner, M., O'Neill, B., Jahn, A., Pendergrass, A.G., Lehner, F., Strand, W.G., Lin, L., Knutti, R., Lamarque, J.F., 2017. Community climate simulations to assess avoided impacts in 1.5 and 2°c futures. Earth Syst. Dynam. 8, 827–847.
Scott, A.A., Waugh, D.W., Zaitchik, B.F., 2018. Reduced Urban Heat Island intensity under warmer conditions. Environ. Res. Lett. 13.
Shandas, V., Voelkel, J., Williams, J., Hoffman, J., 2019. Integrating satellite and ground measurements for predicting locations of extreme urban heat. Climate 7, 5. https://doi.org/10.3390/cli7010005.
Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Huang, D.M.B.X.Y., 2019. A description of the advanced research WRF version 4. In: Technical Report TN-556. National Center for Atmospheric Research. https://doi.org/10.5065/1dfh-6p97.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc. 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1.
Tewari, M., Yang, J., Kusaka, H., Palou, F.S., Watson, C., Treinish, L., 2018. Interaction of urban heat islands and heat waves under current and future climate conditions and their mitigation using green and cool roofs in New York City and Phoenix. Arizona. Env. Res. Lett. 14 (3). https://doi.org/10.1088/1748-9326/aaf431.
Thompson, W., Holt, T., Pullen, J., 2007. Investigation of a sea breeze front in an urban environment. Quart. J. Roy. Meteor. Soc. 133, 579–594.
Voelkel, J., Shandas, V., 2017. Towards systematic prediction of urban heat islands: grounding measurements, assessing modeling techniques. Climate 5, 41. https://doi.org/10.3390/cli5020041.
Walton, D., Sun, F., Hall, A., Capps, S.C., 2015. A hybrid dynamical–statistical downscaling technique, part I: development and validation of the technique. J. Clim. 28, 4598–4617.
Zhao, Z., Giannakis, D., 2016. Analog forecasting with dynamics-adapted kernels. Nonlinearity 29.
Zorita, E., von Storch, H., 1999. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J. Clim. 12, 2474–2489.