Spatio-Temporal Analysis of Urban Heatwaves Using Tukey g-and-h Random Field Models

DAISUKE MURAKAMI, GARETH W. PETERS, TOMOKO MATSUI, AND YOSHIKI YAMAGATA

ABSTRACT
Real-time heatwave risk management with fine-grained spatial resolution is important for analysis of urban heat island (UHI) effects and local heatwaves. This study analyzed the spatio-temporal behavior of ground temperatures and developed methods for modeling them. The developed models consider two higher-order stochastic spatial properties (skewness and kurtosis), which are key to understanding and describing local temperature fluctuations and UHI effects. Application of the developed models to the greater Tokyo metropolitan area demonstrated the feasibility of statistically incorporating a variety of real datasets. Remotely sensed imagery and data from a variety of ground-based monitoring sites were used to build models linking urban covariates to air temperature. Air temperature models were used to capture high-resolution spatial emulator outputs for modeling ground surface temperatures. The main processes studied were the Tukey g-and-h processes for capturing spatial and temporal aspects of heat processes in urban environments. The main finding is that consideration of not only the mean temperature but also the variance, skewness, and kurtosis parameters can reveal hidden heatwave structures.

INDEX TERMS
Heatwave, local approximate Gaussian process, ground temperature modeling, remote sensing, spatial statistics.

I. MOTIVATION
Since the first papers on urban heat islands (UHIs) were published (see [3]), the need for quantifying and modeling spatially fine granular temperature processes in urban environments has grown in prominence. It is becoming increasingly possible to improve both the resolution of spatial-temporal temperature models as well as to distinguish between air temperature models and ground surface temperature models. Using a newly created framework, we developed methods for combining different data sources to model both air and surface temperatures in local urban environments in a consistent manner.

We see two main directions for the use of such air and ground temperature models. The first is related to climate initiatives and smart city design. The second is related to policies on the reduction of risk and the mitigation of population mortality effects arising from urban heatwaves (see for instance [2]). We briefly provide context for these two application domains below.

From the perspective of climate initiatives and smart city design, global initiatives are being established to study local temperature processes through such channels as “green finance” by international organizations such as the Climate Bond Initiative. These organizations are actively developing guidance and frameworks for the quantification of pollution reduction in smart city environments. In particular, they are interested in the inter-relationships between CO2 emissions in urban environments and local temperature profiles, especially how such temperature profiles are related to emissions in urban environments (see discussions in [5], [7], [8]). This interest is driven by a certification and regulation perspective in which funding for low-carbon buildings and efficient transportation systems is increasingly linked to verifiable
emissions targets. In such applications, there is a strong demand for accurate modeling of the spatial-temporal dynamics of both ground and air temperatures.

Local urban temperature modeling is important because it can also inform policy for emission reduction standards. The components of such models are increasingly being used as benchmark references to set standards to be achieved by green infrastructure projects for transportation and low-carbon building initiatives in smart city environments. They are even increasingly being used to justify access to the proceeds from, for instance, green bond debt instruments. The eligibility criteria for such instruments are increasingly reliant on scientific certification of the reductions to be achieved. Quantification of such reductions can be obtained from statistical modeling of local temperature profiles using models like the ones presented here. Different attributes of a model can be applied to different financial products such as green bonds and climate insurance. For instance, model outputs can be used to assess the trigger for green bond cancellation. For climate insurance, the outputs from the local models can be summarized and used to construct indexes for use in determining the size of insurance payouts for claims based on climate loss events. The amounts of financing for green project funding through debt instruments, such as green bonds and their associated securitization markets, are projected to reach 1–2 trillion USD globally in the next couple of years [1]. This places an even stronger spotlight on the methods used to quantify spatial temperature processes and thereby better understand UHI effects for use in assessing the suitability of environmental projects funded by such debt capital (please see discussions in [6]).

In the context of developing local temperature models for health-related policies, the outputs of local UHI and heatwave models for both air and ground temperatures can help inform policy for mitigation and severity/exposure reduction for at-risk populations. Such models can be used to develop strategies for reducing high mortality rates due to extreme temperature distress, especially in elderly urban populations. The effects that extreme temperature events can have on populations are exemplified by the large number of deaths from heat-related distress during the great European heatwave of 2003 (estimated to be >52,000) and the Russian heatwave of 2010 (estimated to be around 15,000). Numerous other heatwave events in India, Pakistan, the Middle East, and Australia have also resulted in extremely high mortality levels. Strengthening resilience against and improving adaptability to climate-related hazards, such as these severe heatwave events, from a political perspective was thus identified as one of the 17 Social Development Goals adopted by the United Nations in 2015. This has made heatwave monitoring and management an important issue in smart city design and policy agendas for aging urban populations around the world.

In this study, we targeted the Tokyo metropolitan area, which is the largest metropolitan area in terms of population (35 million people) and which contains one of the largest urban-dwelling elderly populations on the planet. We analyzed the spatio-temporal behavior of the ground temperatures, focusing on not only mean and variance, which have been considered in many spatial and temporal temperature studies, such as [9], [10], but also skewness and kurtosis, which are key parameters describing extreme heat. We also developed methods for modeling the behavior.

II. STATISTICAL MODELING CONTEXT

We created a statistical modeling framework for producing emulator output in the form of a spatially resolved gridded dataset of ground temperatures. This emulator output is constructed by combining temperature observations in space and time from a combination of satellite remote sensing imagery and a variety of ground-based monitoring site data. The resulting emulator output data is modeled using Tukey g-and-h processes for capturing the spatial and temporal aspects of heat processes in urban environments. These processes are the natural extension of Gaussian process (GP) regression models and are used to incorporate transformations and/or warping properties that explicitly enable skewness and kurtosis to be parameterized. Spatial modeling is achieved by applying the Tukey g-and-h random fields model to the combined dataset (see [11]–[13] and [14]).

The dataset is very large due to the use of fine-scale spatial and temporal resolutions, resulting in a significant computational challenge for the model estimation component of the application. The amount of computation is reduced by applying spectral covariance rank reduction when the model is applied to the emulator modeled ground temperature dataset.

Our analysis of the data for Tokyo suggests that the mean, variance, and spatial skewness are important in characterizing ground temperatures and subsequent UHI and local heatwave propensity. It also revealed that the model parameters varied considerably among urban, suburban, and mountain areas and that one can cluster such areas to form adaptive local climate policy responses to UHI effects in different parts of the cityscape.

A. SPATIAL-TEMPORAL TEMPERATURE MODELING FOR URBAN ENVIRONMENTS

Spatial and temporal data related to temperature, noise, and pollution are increasingly being collected as part of smart city environmental monitoring initiatives (see [15], [16]). This has been facilitated by the development of remote sensor technology and global positioning systems (GPS), along with other remote sensing modes (see [17]).

Our focus here is on the modeling of such spatio-temporal data. Models for such data are commonly developed using GP techniques. However, the number of observation points or design points n for fine-resolution models applicable to the study of local urban environments can be very high dimensional for both time resolution and spatial resolution. In such cases, the classical estimation of a spatio-temporal GP, or kriging, requires the inversion of an \( nT \times nT \) covariance matrix, where n is the number of sample sites and T is the
number of observation times. The computational cost for estimation and prediction is of the order $O((nT)^3)$, so computation is feasible only when $n$ and $T$ are reasonably small. However, when modeling a local climate region, $n$ is typically very large. Since the time resolution is often on the scale of minutes, $T$ can be very large when ground-based sensor data are combined with daily satellite data over many months or years. These computational issues can be addressed using a variety of GP approximation approaches for large sample situations. They include low-rank approaches (e.g., fixed rank kriging [18], predictive process modeling [19], and multiresolution GP [20]), sparse approaches (e.g., covariance tapering [21], the local approximate GP [22], the nearest-neighbor GP [23]), and spatial partitioning approaches [24] (please see [25] for a review of approximation approaches). Many of the processes in these approaches have been extended for spatio-temporal modeling. For example, low-rank spatio-temporal models have been studied [19], [26], [39] and sparse spatio-temporal models have been studied [27], [28]. The widespread use of such GP and kriging models is largely due to the relative simplicity they provide for model formulation and the statistical characteristics they reveal related to the sufficiency of the analysis of second-order process information on the spatial-temporal mean and covariance structure of the underlying process. We argue that, in the context of local resolution models, one may wish to develop models to account for higher-order stochastic structures that may vary in space and time. In the local climate modeling setting we studied, we had large samples in space and time that we believe could potentially have rich information on not only second-order process characteristics of mean and variance. Increased sample resolution can potentially enable the measurement and therefore modeling of higher-order aspects of such spatial-temporal processes, such as spatial skewness and kurtosis describing extreme tail behavior. Such higher-order information should be included in spatial-temporal models to ensure that the resulting estimations are accurate, especially when seeking to explain extreme events such as extreme heat and UHIs.

Modeling spatial and temporal extremes has been attracting attention in geostatistics owing to the increasing availability of data with improved sampling resolutions. The max-stable process [34], which is a natural infinite-dimensional extension of the univariate generalized extreme value distribution, has been extended in a spatial setting, e.g., [29]–[31]. The proposed spatial max-stable processes have frequently been applied to model extreme climate events, e.g., [32], [33]. Unfortunately, as the likelihoods associated with the spatial max-stable processes have no closed-form expressions ([31]), they are not easily developed, especially for spatial-temporal contexts with high-resolution sampling, such as when $n$ or $T$ is large. Moreover, there has been little to no work on the temporal time series characterization of such processes.

Furthermore, max-stable processes are based on the assumption that the index of regular variation characterizing the extreme spatial behavior is homogeneous in space and time, an assumption we did not wish to make with the fine-resolution process modeling undertaken in this work. Furthermore, such approaches are more applicable to the characterization of only extreme behaviors as opposed to our objective, which is to extend the modeling from second-order characteristics to higher-order stochastic structures while also allowing for the regular variation properties of models to capture potential heavy-tailed extremes should they appear locally in some regions of space over time.

Therefore, we aimed to develop new classes of processes, building on the recent studies of [13], [14], and [58]. In those studies, different models of the Tukey g-and-h (TGH) random field (TGH-RF) were proposed. Such models are typically developed as transforms or warpings of a GP via TGH transformation. This family of transformations enables explicit modeling and estimation of higher-order features such as skewness and kurtosis of an underlying process (please see explicit derivations of such population process characteristics in [58]). Regarding temporal components, such models were recently extended by Yan and Genton [35] to capture non-Gaussian autoregressive processes.

Warped GP models in the TGH family are now more suitable for modeling in such big data contexts as evaluating likelihood, first outlined in [56] and [57]. These models have been resolved in efficient computational frameworks, such as the numerically efficient approach reported in [36] for maximizing an approximate likelihood, at least for small-$n$ sample settings. We have developed methods that reduce the numerical cost incurred in the inversion of the $n \times n$ covariance matrix, which normally makes computation intractable when $n$ is large. In other words, we have developed a TGH-RF modeling approach that is fast and applicable to large spatial datasets.

We can use such models and observation data extracted from remotely sensed images for temperature data analysis, which produces a very large sample size in space. The data used to train a TGH spatial process model typically is of the order of gigabytes per minute. To facilitate the use of such data in local temperature modeling contexts, we developed two approaches, a low-rank TGH-RF modeling approach and a sparse TGH-RF modeling approach. We compared them with the original TGH-RF approaches introduced above in terms of computational time and estimation accuracy. The developed approaches are natural extensions of the ideas of the laGP framework ([22]) extended to the TGH process context.

Another contribution of this study is the development of a high-resolution dataset of emulated ground temperatures in the Tokyo metropolitan area; the developed TGH-RF models were applied to this dataset for heatwave analysis. There are at least two monitoring systems available to provide temperature data: station monitoring and remote sensing. The former monitors temperature, humidity, and other climate variables.
by the minute and by the hour at spatially distributed monitoring stations. The latter monitors ground temperatures with ultra-high spatial resolution for a particular observation window per day; however, values are often missing in certain spatial regions due to cloudy/rainy conditions (see Section III). We combine these data for temperature emulation. For discussions on combining multiple spatial datasets, see [37]. In short, we developed low-rank and sparse TGH-RF modeling approaches and applied them to the emulated temperature dataset. This study is organized as follows. In Section III, we describe the emulator we developed for generating spatially fine ground temperature maps daily by combining surface temperature data from satellite remote sensing with ground-based air temperature data observed at monitoring stations in the study region, the greater Tokyo metropolitan area. Section IV describes the development and application of our approximated fast TGH-RF models for analyzing temperature distribution properties, including spatial co-skewness and spatial co-kurtosis, and compares this novel solution to classical approaches. Section V further explores temperature distribution properties. Finally, Section VI summarizes the key points and mentions potential future work.

III. COMBINING SATELLITE REMOTE SENSING DATA WITH GROUND MONITORING STATION DATA

In this section, we first outline the properties of the different data sources we used. Next, we explain the first stage of our analysis in which the framework we created was used to combine satellite remote sensing ground temperature data with ground-based weather monitoring air temperature data as well as spatial covariate information and thereby produce emulator model output data.

A. AIR AND GROUND TEMPERATURE DATA PROCESSING AND EMULATOR MODELING: OVERVIEW

The ground temperature was our primary focus as it reflects the absorption and radiation of heat from roads, buildings, and other urban materials, which determine the intensities of heatwaves and UHIs. To emulate ground temperatures accurately, we combined the air temperature data, which had low spatial resolution and high accuracy, with remotely sensed ground temperature data, which had higher spatial resolution and lower accuracy.

Air temperature datasets for the target area were obtained from two sources. The Japan Meteorological Agency (JMA) provided hourly temperatures monitored at 78 stations (as of 2013) since 1976. NTT DOCOMO Inc. provided air temperatures observed every minute at 206 stations since 2013. The locations of the 78 + 206 monitoring stations are plotted in Fig. 1.

A ground temperature dataset was obtained from remotely sensed images captured by the Moderate Resolution Imaging Spectroradiometers (MODIS) onboard the Terra and Aqua satellites (http://modis.gsfc.nasa.gov/). They generate worldwide ground temperature maps with a spatial resolution of 1 km at 01:30, 10:30, 13:30, and 22:30 every day. There are 31,235 observation points in the target area. We cross-checked the data against the other two data sources to ensure accuracy. Example plots of ground temperatures on six days in August 2013 are shown in Fig. 2. As you can see, some values are missing due to cloudy/rainy conditions. The missing values in August had a daily mean ratio of 0.70 while the ratio for September was 0.65. Nevertheless, we were able to utilize 9,284 and 10,813 MODIS data values in August and September, respectively.

As described above, both the air and ground temperature datasets were incomplete. The next section explains how we combined the data from these different sources with local spatial covariate information, which is detailed in Table 1, to develop a regression-based emulator model. This model was used to estimate the missing values, enabling us to obtain a complete ground temperature model output profile in space and time.

Use of the additional location-specific explanatory variables in Table 1 was found to result in statistically significant improvements in the emulator model used to identify discontinuous changes in air/ground temperatures along the border separating urban and non-urban areas.

Fig 3 summarizes the stages in the emulator regression model used to produce an emulator spatial-temporal map for ground temperatures. The input data are (A1) the air temperature data (JMA and NTT DOCOMO) and (A2) the ground temperature data (MODIS). We first modeled the air temperatures using (B1) a regression Spatial Best Linear Unbiased Estimator (S-BLUE) ( [38]) and the explanatory variables in Table 1, as described in Section III-B2). This produced (C1) the emulated air temperature data. The emulated air temperatures and variables in Table 1 were used as explanatory variables in (B2) a rank reduced S-BLUE approach (see Section III-C) to emulate the (C2) ground temperature data in 1 km grids.

B. REGRESSION S-BLUE FOR AIR TEMPERATURE EMULATION (B1)

Section III-B1 explains the model, and Section III-B2 explains the regression S-BLUE approach.
**FIGURE 2.** Illustrative example images of MODIS ground temperatures in greater Tokyo area at 13:30 (August 12–17, 2013). Temperatures could not be observed for some areas due to cloud obstruction.

**TABLE 1.** Explanatory variables.

| Variable   | Description                                                                 | Source | Year |
|------------|-----------------------------------------------------------------------------|--------|------|
| Station    | Exponential of negative distance to nearest station (accessibility to nearest railway station) [m] | NLNI¹  | 2005 |
| Population | Nighttime population density [people/km²]                                   | Census² | 2005 |
| Cropland   | Proportion of cropland in 1 km grids                                       |        |      |
| Forest     | Proportion of forest land in 1 km grids                                     | NLNI   | 1997 |
| Urban      | Proportion of urban land in 1 km grids                                      |        |      |
| Water      | Proportion of water bodies in 1 km grids                                    | NLNI   | 2006 |
| Ocean      | Proportion of ocean in 1 km grids                                           |        |      |
| Elevation  | Elevation [m]                                                               | NLNI   | 2006 |
| Latitude   | Latitude [degree]                                                          | JMA³   |      |

¹ NLNI: National Land Numerical Information download service (https://nlftp.mlit.go.jp/ksj/index.html)
² Census: Population of Japan 2005 (https://www.stat.go.jp/english/data/kokusei/2005/poj/mokuji.html)
³ JMA: Japan Meteorological Agency (https://www.jma.go.jp/jma/indexe.html)

**FIGURE 3.** Overview of data processing (blue: raw input data (A1–3); green: model fitting (B1–2); red: emulated data produced from fitted model (C1–2)).
1) SPATIAL REGRESSION MODEL
We assumed the following regression structure to model the observed air temperatures:
\[
Y = Xz + \epsilon E[\epsilon] = 0Cov[\epsilon] = C,
\]
where \( Y = [Y(s_1), \ldots, Y(s_i), \ldots, Y(s_n)]^T \) is a vector of air temperatures monitored at locations \( \{s_1, \ldots, s_i, \ldots, s_n\} \in D \subseteq \mathbb{R}^2 \), \( X \) is an \((n \times K)\) matrix of the explanatory variables in Table 1, \( z \) is a vector of regression coefficients, and \( \epsilon \) represents the noise process.

The \((n \times n)\) covariance matrix of the noise process, denoted \( C \), was parametrized by a distance-decay exponential kernel capturing spatial dependence:
\[
e[d(s_i, s_j)] = \begin{cases} \tau^2 \exp(-d(s_i, s_j)/r) & \text{if } d(s_i, s_j) > 0 \\ \tau^2 + \sigma^2 & \text{otherwise}, \end{cases}
\]
where \( d(s_i, s_j) \) is the Euclidean distance between sites \( s_i \) and \( s_j \) measured in units of meters \([m]\), \( r \) is a range parameter, and \( \tau^2 \) and \( \sigma^2 \) are variance parameters for spatial and non-spatial variations, respectively. The kernel can be replaced with Gaussian, spherical, or other positive definite kernels (see [39]). Our model describes discontinuous changes in temperature, e.g., at borders separating urban and non-urban areas, using the regression term and describes continuous changes using a spatially dependent process.

The same structure is assumed to be behind the unobserved temperature at location \( s_a \in D \subseteq \mathbb{R}^2 \):
\[
Y(s_a) = X(s_a)^Tz + \epsilon(s_a)Cov[\epsilon(s_a), \epsilon] = \epsilon(s_a),
\]
where \( X(s_a) \) is a \((K \times 1)\) vector of the explanatory variables observed at location \( s_a \), and \( \epsilon(s_a) \) is a \((n \times 1)\) vector in which the \( i\)-th element equals \( e[d(s_i, s_a)] \).

2) LINEAR APPROXIMATION USING S-BLUE
A non-linear spatial smoothing function behind air temperatures was linearly approximated using a spatial best linear unbiased estimator, S-BLUE, as studied in [38]. S-BLUE, denoted here as \( \hat{f}(s_a) \), is a linear approximation of a non-linear spatial smoothing function used to estimate missing observation \( Y(s_a) \). It does this by minimizing the mean-squared error \( E\left[(y(s_a) - \hat{f}(s_a))^2\right] \) under the constraints of linearity and unbiasedness. Solving the minimization problem on the basis of (1 and 3) produces the regression S-BLUE estimator:
\[
\hat{f}(s_a) = X(s_a)^T\hat{z} + \hat{c}(s_a)^T\hat{C}^{-1}[Y - X\hat{z}],
\]
where \( \hat{z} = (X^T\hat{C}^{-1}X)^{-1}X^T\hat{C}^{-1}Y \). We used this regression S-BLUE to emulate air temperatures in 1 km grids. The variance parameters \([\tau^2, \sigma^2, r]\) were estimated using the robust weighted least squares method [40]. We used the gstat R package (https://cran.r-project.org/web/packages/gstat/) for the estimation.

C. RANK REDUCED REGRESSION S-BLUE FOR GROUND TEMPERATURE EMULATION (B2)
The variables in Table 1 and (C1) are the emulated air temperatures, which are used as explanatory variables for (B2) the ground temperature emulation. Unfortunately, the original S-BLUE, which involves matrix inversion \( \hat{C}^{-1} \), is computationally infeasible because of the large sample size of the MODIS data. To reduce the computational cost, we approximated \( \hat{C} \) by \( \hat{E}_L\hat{\Lambda}_L\hat{E}_L^T \), where \( \hat{\Lambda}_L \) is an \( L \times L \) diagonal matrix consisting of the \( L \) largest eigenvalues approximated by the Nyström extension ([49]), and \( \hat{E}_L \) is an \( n \times L \) matrix consisting of the corresponding \( L \) approximated eigenvectors. Thus, we approximate \( \hat{C} \) using the \( L \) principal eigenpairs. Given the eigen-decomposition, the S-BLUE estimator yields
\[
\hat{f}(s_a) = X(s_a)^T\hat{z} + \hat{c}(s_a)^T\hat{E}_L \hat{\Lambda}_L^{-1}\hat{E}_L^T[Y - X\hat{z}].
\]

The complexity for the rank reduced regression S-BLUE is \( O(n) \), making it feasible for the MODIS data.
D. EMULATION RESULTS

1) AIR TEMPERATURES (C1)
We used the regression S-BLUE approach explained in Section III-B2 to emulate air temperatures in 1 km grids at 13:00 JST each day for 61 days. The mean R-squared value of the model for each day was 0.818, which verifies the accuracy of our emulation model. We illustrate a randomly selected, but representative, example of the fitting results below.

Fig. 4 shows example boxplots of the estimated t-values of the regression coefficients for the air temperature model. Note that the variance inflation factor, which exceeds 10 if there is severe multicollinearity, verifies that the explanatory variables are not collinear with each other. It can be seen that the air temperatures were significantly lower in cropland areas and significantly higher in populated areas. These results are plausible considering the cooling effects of natural environments and the trapping of heat due to UHI effects in populated areas. Furthermore, we see that elevation is negatively significant. The positive sign for latitude might reflect the cooling effect of the ocean in the south combined with the thermal storage effect of the inland areas in the north.

Emulated air temperatures at 11:00, 12:00, and 13:00 on August 17 are plotted in Fig. 5 (top). A land cover map is shown in (6) for reference. The plots show a rapid increase in air temperatures, especially near and in the urban areas.

2) GROUND TEMPERATURES (C2)
Ground temperatures in 1 km grids at 13:00 JST each day for 61 days were emulated using the rank reduced S-BLUE approach (see Section III-C). In accordance with the results of preliminary analysis, \(L\) was set to 200. For days with less than 5,000 observations, we used data for that day plus the data for one day before and one day after to stabilize the estimation. To linearize the air/ground temperature relationship, we applied the Box-Cox transformation to air temperatures, using a multiplier estimated a priori by maximizing the likelihood of the linear model between the two temperatures. The mean R-squared value across the days was 0.7303, suggesting
reasonable emulation accuracy. Box plots of the estimated t-values of the regression coefficients are shown in Fig. 7. As expected, the air temperatures increased the ground temperatures. The populated and urban areas had a positively significant effect on the ground temperature. These results demonstrate the existence of UHIs in these areas. Furthermore, latitude had a positively significant effect. We attribute this to the existence of a basin in the target area that typically has higher temperatures in the north.

Emulated ground temperatures at 11:00, 12:00, and 13:00 on August 17 are also plotted in Fig. 5 (bottom). Their spatial patterns differ from those of the air temperatures; ground temperatures are clearly a better indicator of heat in urban areas (see Fig. 6). This is because ground temperature reflects thermal storage and radiation due to urban building materials, which are the principal sources of UHIs (e.g., [43]). This is also the reason ground temperatures are typically higher than air temperatures. Although air temperatures are commonly used for heatwave risk assessment (e.g., [41], [42]), our results suggest that ground temperature is a better indicator of UHI effects. Although the results are only for one of the 61 target days, examination of the data for the other days showed that ground temperatures in urban areas were high on most of the target days. The temperature increases after 11:00 were relatively small in the suburban areas, such as those indicated by the ellipses in Fig. 5 (bottom), because the temperature had typically peaked by 11:00.

Emulated ground temperatures at 13:30 on August 12–17 are plotted in Fig. 8. Comparison with the original
The Tukey g-and-H (TGH) distribution is formulated as

\[ \tilde{Y}_t(s_i) = a + b \tau_{g,h}[Z_t(s_i)] \]

where \( Z_t(s_i) \) is a Gaussian random variable satisfying \( E[Z_t(s_i)] = 0, \text{Var}[Z_t(s_i)] = 1, \text{and Cov}[Z_t(s_i), Z_t(s_j)] = c[d(s_i, s_j)] \). The TGH-RF model is an extension of the GP model and is used to estimate skewness \((g)\) and kurtosis \((h)\). The log-likelihood of the model at time \( t \) is given as

\[
L(\theta_1, \theta_2) \propto - \frac{1}{2} Z_t^T(C^{-1} + hI)Z_t - \frac{1}{2}|C|\text{log}(b) \\
- \sum^n_{i=1} \log[\exp(gZ(s_i))] \\
+ \frac{1}{g}[\exp(gZ(s_i) - 1)hZ(s_i)]]
\]

C. TUKEY G-AND-H RANDOM FIELDS (TGH-RF) MODELS

1) ORIGINAL MODEL

The original TGH-RF model [46] is an extension of the GP model incorporating skewness and kurtosis parameters. The TGH-RF model is formulated as

\[ \tilde{Y}_t(s_i) = a + b \tau_{g,h}[Z_t(s_i)] \]

\[ \tau_{g,h}[Z_t(s_i)] = \frac{1}{g}\exp(gZ_t(s_i) - 1)\exp\left(h\frac{Z_t^2(s_i)}{2}\right) \]

Unfortunately, this algorithm is not suitable for our large dataset due to the iterative calculation. To cope with this problem, the subsequent two subsections extend the TGH-RF model by introducing low-rank and sparse approximations.

D. LOW-RANK TGH-RF MODELING

1) MODEL AND ESTIMATION

To accelerate the computation, we approximated \( Z_t \) with \( E_L \Gamma_L \), where \( \Gamma_L \sim N(0, p \Lambda_L^T \Lambda_L) \). Thus, \( \Lambda_L \) is an \( n \times L \) matrix of the first \( L \) eigenvector of a prespecified spatial correlation matrix, \( \Gamma_L \) is an \( L \times L \) diagonal matrix in which the elements are the first \( L \) eigenvalues of the matrix, and \( p = \frac{1}{m m^T} \) (which is required to ensure \( \text{Var}[E_L \Gamma_L] = 1 \), and \( m \) is a parameter determining the scale of the spatial process. Eigenvectors corresponding to greater eigenvalues are emphasized when \( m \) is large, and the resulting GP has a global map pattern; the reverse is true for small \( m \). Thus, \( Z_t \) satisfying \( E[Z_t] = 0, \text{Var}[Z_t] = I, \) and \( \text{Cov}[Z_t] = C \) is approximated by \( E_L \Gamma_L \) satisfying \( E[E_L \Gamma_L] = 0, \text{Var}[E_L \Gamma_L] = I, \) and...
eigenpairs accelerates the computation, as we will demonstrate below. The distance between sites $i$ and $j$ is the Euclidean parameter (48). More importantly, the fixed $r$ covering the sample sites. Although the fixed $r$ by the maximum distance of the minimum spanning tree $m$ and another eigen-approximation technique (see 51).

The $i,j$-th elements of the $C$ matrix are given by the exponential kernel $\exp(-d_{ij}/r)$, where $d_{ij}$ is the Euclidean distance between sites $i$ and $j$, and $r$ is a given range parameter. In accordance with the method in [47], $r$ is given a priori by the maximum distance of the minimum spanning tree covering the sample sites. Although the fixed $r$ is somewhat restrictive, the scale parameter $m$ assumes the role of the range parameter (48). More importantly, the fixed $r$ drastically accelerates the computation, as we will demonstrate below.

The log-likelihood (10) can be rewritten using the eigenpairs:

$$L(\theta_1, \theta_2) \propto -\frac{1}{2}Z_t^T E(p\Lambda_L^{-m} + hI)E_L^TZ_t - \frac{1}{2}p\prod_{l=1}^{L} \lambda_l^m - n\log(b) - \sum_{i} \log\left[\exp(gZ_t(s_i))\right] + \frac{1}{8}\exp(gZ_t(s_i)) - 1)hZ_t(s_i).$$

$Z_t(C^{-1} + hI)Z_t$ and $|C|$, which have complexity $O(n^3)$, are replaced with $Z_t^T E(p\Lambda_L^{-m} + hI)E_L^TZ_t$ and $\prod_{l=1}^{L} \lambda_l^m$, which have complexities $O(Ln)$ and $O(L)$, respectively. The computationally efficient log-likelihood specification is not available without fixing the $r$ parameter (because the eigenpair changes depend on the $r$ value). Equation (11) can be maximized exactly like Eq. (10), in which $\theta_2 = m$.

Monte Carlo simulation of the low-rank approach revealed that it accurately approximated the $a, b, g, h$ parameters relative to the original TGH-RF model in a computationally efficient manner as long as the number of basis functions was not too small. Details of the validation study are in Supporting Material B.

2) APPLICATION TO EMULATOR MODEL GROUND TEMPERATURES

The low-rank approach was used to estimate the moment parameters (location, scale, skew, kurtosis) for 7 clusters for the 61 target days. As shown in Fig. A2, the clusters were numbered in accordance with their distance from the center. The estimated parameter values are plotted in Fig. 9. The top panels show them for the three urban clusters, i.e., the three clusters with geometric centers closest to Tokyo Station, and the bottom ones show them for the other four, non-urban, clusters (see Fig. 14). In each panel, the solid lines represent parameter estimates, and the dashed lines represent their 95% confidential intervals.

For the non-urban clusters, location parameter $a$ gradually decreased over the period, whereas for the urban clusters, it remained large across the period. This may be due to the fact that urban materials store heat. Heatwaves are thus estimated to last longer in urban areas than in non-urban areas. The gradual increase in scale parameter $b$ for the urban clusters indicates that heat is more variable at the end of the summer than at the beginning.

Skew parameter $g$ was positively significant for many clusters. The temperatures are estimated to have a right-skewed tail. In other words, the temperature can become extremely high, resulting in hazardous heat in rare cases. This suggests the importance of considering skewness, which is ignored in standard second-order spatial models (e.g., the GP model), in heat risk estimation. By contrast, kurtosis parameter $h$ had large standard errors and thus was statistically insignificant in most cases. This indicates that kurtosis is less important in heatwave modeling.
The estimated probability densities for each day for each cluster are plotted in Fig. 10. The distributions for clusters 1, 3, 4, and 5, which are near the center of the Tokyo area, are right-skewed while those for clusters 6 and 7, which are in mountain areas, are nearly Gaussian. This means that if we assume a Gaussian distribution for temperature, the right skew is ignored, and risk is underestimated.

E. SPARSE TGH-RF MODELING

1) OUTLINE

The low-rank approach has two limitations. First, it tends to underestimate small-scale spatial variations (see [52]). Second, it is based on the assumption that the a, b, g, h values are the same across space although they might actually vary across geographical space. To overcome these limitations, we developed a sparse TGH-RF approach by extending the approached used by Gramacy and Apley [22].

Their approach, which is called local approximate GP (laGP), estimates a local GP for each prediction location independently, using n(s_j) local subsamples, which we denote D(s_j). Although the n(s_j) nearest-neighbors (NN) method works well in many situations, the NN-based design is not optimal because accurate GP estimation requires a minimum amount of subsample spread ([54]). Gramacy and Apley thus proposed a greedy search algorithm to find D(s_j) [22], minimizing the Bayesian mean-squared predictive error. We reduced the computational burden of the TGH-RF model by applying their approach.

2) LOCAL APPROXIMATE TGH-RF

We assume the following local TGH process around a location s_0:

\[
\begin{pmatrix}
    y(s_{j+1}) \\
    y(s_0)
\end{pmatrix}
\sim TGH[\mu, \tilde{C}, \tilde{g}, \tilde{h}]
\]

\[
\tilde{\mu} = \begin{pmatrix} 
    \mu(s_{j+1}) \\
    \mu(s_0)
\end{pmatrix},
\]

\[
\tilde{C} = \begin{pmatrix}
    C(s_{j+1}, s_{j+1}) & c(s_{j+1}, s_0) \\
    c^T(s_{j+1}, s_0) & c(s_0, s_0)
\end{pmatrix},
\]

\[
\tilde{g} = \begin{pmatrix}
    g(s_{j+1}) \\
    g(s_0)
\end{pmatrix},
\]

\[
\tilde{h} = \begin{pmatrix}
    h(s_{j+1}) \\
    h(s_0)
\end{pmatrix},
\]

where s_{j+1} = \{s_1, s_2, ... s_{j+1}\} is a set of the j + 1 nearest locations from location s_0, y(s_{j+1}) is a vector of observations at the j + 1 sites, \mu(s_{j+1}) and \mu(s_0) are mean functions, \tilde{C} is a kernel matrix used to model residual spatial dependence, and g(s_0) and h(s_0) are the g-and-h parameters around location s_0. Given (12), we estimate the conditional mean a(s_0) and variance b(s_0) of y(s_0):

\[
y(s_0) | y(s_{j+1}) \sim TGH[ a(s_0), b(s_0), g(s_0), h(s_0) ].
\]

The expectation of a(s_0) is given by

\[
a(s_0) y(s_{j+1}) = \mu(s_0) + c^T(s_0, s_{j+1}) C^{-1}(s_{j+1}, s_{j+1}) \epsilon_j + 1
\]

\[
= \mu(s_0) + \begin{pmatrix}
    c^T(s_0, s_j) & c(s_j, s_{j+1})
\end{pmatrix}^{-1} \begin{pmatrix}
    \epsilon_j \\
    \epsilon_{j+1}
\end{pmatrix},
\]

where \epsilon_j = T_{gh}^{-1}(y(s_j) - \mu(s_j)) and \epsilon_{j+1} = T_{gh}^{-1}(y(s_{j+1}) - \mu(s_{j+1})). C^{-1}(s_{j+1}, s_{j+1}) \epsilon_{j+1}, which is an \( j + 1 \) \times 1 vector that appears in (14), has the following recursive expression ([22], Eq. 9):

\[
C^{-1}(s_{j+1}, s_{j+1}) \epsilon_{j+1}
\]

\[
= \begin{pmatrix}
    C^{-1}(s_j, s_j) \epsilon_j + g(s_j, s_{j+1}) h(s_{j+1}) \beta(s_{j+1}) + \epsilon_{j+1} \\
    h(s_{j+1}) + \epsilon_{j+1}/\beta(s_{j+1})
\end{pmatrix},
\]

where g(s_{j+1}) = C^{-1}(s_j, s_j) c(s_j, s_{j+1})/\beta(s_{j+1}), and h(s_{j+1}) = \epsilon_j^T g(s_j, s_{j+1})/\beta(s_{j+1}) is defined later. Using (15), we can sequentially update a(s_0)\( y(s_{j+1}) )

\[
a(s_0)| y(s_{j+1}) = a(s_0) | y(s_j) + c(s_j, s_{j+1}) h(s_{j+1})
\]

\[
+ \epsilon_{j+1} / \beta(s_{j+1})
\]

where s_{j+1} = \{s_1, s_2, ... s_{j+1}\} is a set of the j + 1 nearest locations from location s_0, y(s_{j+1}) is a vector of observations at the j + 1 sites, \mu(s_{j+1}) and \mu(s_0) are mean functions, \tilde{C} is a kernel matrix used to model residual spatial dependence, and g(s_0) and h(s_0) are the g-and-h parameters around location s_0. Given (12), we estimate the conditional mean a(s_0) and variance b(s_0) of y(s_0):

\[
y(s_0) | y(s_{j+1}) \sim TGH[ a(s_0), b(s_0), g(s_0), h(s_0) ].
\]

The expectation of a(s_0) is given by

\[
a(s_0) y(s_{j+1}) = \mu(s_0) + c^T(s_0, s_{j+1}) C^{-1}(s_{j+1}, s_{j+1}) \epsilon_j + 1
\]

\[
= \mu(s_0) + \begin{pmatrix}
    c^T(s_0, s_j) & c(s_j, s_{j+1})
\end{pmatrix}^{-1} \begin{pmatrix}
    \epsilon_j \\
    \epsilon_{j+1}
\end{pmatrix},
\]

where \epsilon_j = T_{gh}^{-1}(y(s_j) - \mu(s_j)) and \epsilon_{j+1} = T_{gh}^{-1}(y(s_{j+1}) - \mu(s_{j+1})). C^{-1}(s_{j+1}, s_{j+1}) \epsilon_{j+1}, which is an \( j + 1 \) \times 1 vector that appears in (14), has the following recursive expression ([22], Eq. 9):

\[
C^{-1}(s_{j+1}, s_{j+1}) \epsilon_{j+1}
\]

\[
= \begin{pmatrix}
    C^{-1}(s_j, s_j) \epsilon_j + g(s_j, s_{j+1}) h(s_{j+1}) \beta(s_{j+1}) + \epsilon_{j+1} \\
    h(s_{j+1}) + \epsilon_{j+1}/\beta(s_{j+1})
\end{pmatrix},
\]

where g(s_{j+1}) = C^{-1}(s_j, s_j) c(s_j, s_{j+1})/\beta(s_{j+1}), and h(s_{j+1}) = \epsilon_j^T g(s_j, s_{j+1})/\beta(s_{j+1}) is defined later. Using (15), we can sequentially update a(s_0)\( y(s_{j+1}) )

\[
a(s_0)| y(s_{j+1}) = a(s_0) | y(s_j) + c(s_j, s_{j+1}) h(s_{j+1})
\]

\[
+ \epsilon_{j+1} / \beta(s_{j+1})
\]
Since the sequential calculation does not involve the heavy inversion of $C^{-1}$, it is computationally efficient.

Regarding $b^2(s_0)$, Xu and Genton showed that the variance in the TGH process is identical to the variance in GP (see Lemma 3 in [14]). Therefore, for a local TGH process, the conditional variance is given exactly like laGP (see Eq. 5 in [22]):

$$b^2(s_0)|y(s_{j+1}) = \frac{\psi_j}{f - 2} \beta_{j+1}(s_0),$$  \hspace{1cm} (17)

where $\psi_j = \epsilon_j^T C^{-1}(s_j, s_j) \epsilon_j$, and

$$\beta_{j+1}(s_0) = c(s_0, s_0) - c^T(s_0, s_{j+1}) C^{-1}(s_{j+1}, s_{j+1}) c(s_0, s_{j+1}).$$  \hspace{1cm} (18)

The following recursive equation is based on the one in [22]:

$$\beta_{j}(s_0) - \beta_{j+1}(s_0) = \beta_{j}(s_{j+1})^T c(s_{j+1}, s_{j+1}) G_j(s_{j+1}) c(s_0, s_{j+1})$$
$$+ 2 c(s_0, s_{j+1}) c^T(s_0, s_{j+1}) g_j(s_{j+1})$$
$$+ \frac{c(s_0, s_{j+1})^2}{\beta_j(s_{j+1})}$$
$$\psi_{j+1} = \psi_j + h_j(s_{j+1})^2 \beta_j(s_{j+1}) + 2 \epsilon_j h_j(s_{j+1}) +$$
$$+ \frac{\epsilon_j^2}{\beta_j(s_{j+1})},$$  \hspace{1cm} (19)

where $G_j(s_{j+1}) = g_j(s_{j+1}) g_j(s_{j+1})^T$. Thus, $b^2(s_0)|y(s_{j+1})$ is updated sequentially by substituting (19) and (20) into (17).

3) LOCAL SAMPLING DESIGN
Since the variance updating equations for the local TGH process are identical to those for the laGP, the expression used for the laGP variance can be used here:

$$b^2_{j+1}(s_0)|y(s_{j+1}) = b^2(s_0)|y(s_{j+1}) - \frac{\psi_j}{f - 2} R_{j+1}(s_0).$$  \hspace{1cm} (21)

$$R_{j+1}(s_0) =$$
$$\left[ c(s_0, s_{j+1}) - c^T(s_{j+1}, s_{j+1}) C(s_{j+1}, s_{j+1})^{-1} c(s_0, s_{j+1}) \right]^2$$
$$\left[ c(s_0, s_0) - c^T(s_0, s_{j+1}) C(s_{j+1}, s_{j+1})^{-1} c(s_0, s_{j+1}) \right]^2.$$  \hspace{1cm} (22)

Therefore, the following local design algorithm, which was proposed in Algorithm 1 of [53], can be used here.

1) Let $s_{k(-j)}$ denote the $k$ nearest neighbors to $s_0$ among the locations not currently in sub-design $s_j$. Set $\delta_{j+1}$ as the maximum variance reduction from $N_{jk}(s_0)$. That is,

$$\delta_{j+1} = \max_{u \in s_{k(-j)}} R_u(s_0).$$  \hspace{1cm} (23)

2) Set

$$\gamma = \Phi^{-1} \left( \frac{\delta_{j+1}}{\sqrt{1 + \sqrt{1 + \epsilon_j^T C(s_j, s_j)^{-1} c(s_j, s_0) \|\epsilon_j\|^2 + \beta_{j+1} / \lambda_{\min}}} \right),$$

where $\lambda_{\min}$ is the minimum eigenvalue of $C(s_j, s_j)$. Let

$$T(s_j) = \{ u \in s_{k(-j)} : \| u - v \|_2 \leq \gamma \text{ for some } v \in \{ s_0, s_j \} \}.$$  \hspace{1cm} (24)

Then,

$$s_{j+1} = \arg \max_{u \in T(s_j)} R_u(s_0).$$  \hspace{1cm} (25)

3) Set $j = j+1$ and repeat 2 and 3 until either the reduction in variance $R_u(s_0)$ falls below a prespecified threshold or the local design budget is met.

4) LOCAL TGH PROCESS ESTIMATION PROCEDURE
The following procedure ([22]) is applicable if the laGP parameter estimation part is replaced by the laGH estimation part.

1) Choose a reasonable starting global lengthscale parameter $\theta_0$ for all sites $s_0$.

2) Calculate local design $s_j$ for each $s_0$ independently by sequential application of the procedure above.

3) Independently calculate the MLE lengthscale $\hat{\theta}(s_0)|s_j$, thereby explicitly obtaining a global nonstationary predictive surface.

4) Set $\theta(s_0) = \hat{\theta}(s_0)$ possibly after spatial smoothing over all $s_0$ locations.

5) Repeat steps 2–4 as desired. Then independently output each prediction $s_0$ based on $s_j$ and possibly smoothed $\hat{\theta}(s_0)$.

Unlike the low-rank approach, the sparse approach estimates local variations of skewness and kurtosis. Furthermore, the independent computation for each site enables parallelizing local model estimations.

We performed a Monte Carlo simulation experiment to examine the parameter estimation accuracy of the local TGH modeling approach. The results suggest that it properly estimates a positive $g$ value if the true process has a positive $g$. The same is true for negative $g$ values. However, the estimates have a relatively large variation because only a small number of samples are used for the estimation. The local averaging, which we assume in the fourth step of the local TGH-RF estimation procedure, is valuable for reducing the variation so that credible estimates can be obtained.

Regarding kurtosis parameter $h$, the estimate properly takes values near zero when the true process has a zero $h$ value while the estimate tends to take larger values when the true process has a positive $h$ value. However, the estimates are smaller than the true value. Estimation of the $h$ parameter using small local samples would thus be useful. See Supporting Material B for further details.

5) APPLICATION TO ENUMERATED GROUND TEMPERATURES
The sparse TGH-RE approach was used to estimate the $(a, b, g, h)$ parameters at each site. To analyze the typical spatial patterns of these parameters and also to stabilize the estimates, the sparse TGH approach was fitted to 5-day intervals from July 1, and the resulting parameter estimates were averaged over the target days. For fast computation, the sizes of the local samples, which were optimized using the procedure explained in the previous section, were constrained to be equal to or less than 200.

Fig. 11 illustrates the estimated parameters averaged across 11 days (except for July 19, when the mean maximum temperature was the highest). The estimated $\hat{a}(s_0)$ parameters
indicate that the temperatures in the urban areas were higher than those in the mountain areas. The estimated $b(s_0)$ values were larger in the mountain areas, where the up- and downdrafts over the mountains make the weather more variable. These results show that this approach captures such local variability in the temperature. In contrast, $\hat{g}(s_0)$ and $\hat{h}(s_0)$ were nearly zero across the region. These values were almost zero across the 11 days, meaning that skewness and kurtosis are not so influential, at least on typical summer days.

Fig. 12 plots the estimated parameters on July 19, when the mean maximum temperatures were the highest across the target days. Interestingly, their spatial patterns are quite different from the average patterns shown in Fig. 11. The $\hat{a}(s_0)$ parameter shows a similar spatial pattern in the mountain areas but much higher values near the center. This reflects the stronger presence of UHI effects near the center urban area. The large values of $b(s_0)$ in the Tokyo Bay area suggest that sea breezes, humidity, and other sea-related factors make the temperature more variable on hot days. The $\hat{g}(s_0)$ parameter had large positive values for the center of Tokyo, indicated by the ellipse. This suggests that extreme heat means not only an increase in temperature but also changes in the temperature distribution shape, which determines risk (e.g., exceedance probability). The $\hat{g}(s_0)$ values tended to take positive values in the mountain areas as well, meaning that skewness is especially important on hot days.

The $\hat{h}(s_0)$ values tended to be high in the basin area indicated by the white ellipse due to the flow of hot air into the basin. The large values indicating a fat tail might be caused by dramatic changes in temperature depending on the existence of a hot air flow. A fat tail (or a large $h(s_0)$ value) appeared only for the intensively hot day, an important finding for evaluating probabilistic risk appropriately.

V. EXPLORATORY TEMPORAL ANALYSIS OF EMULATED GROUND TEMPERATURES

The previous section described the analysis of the spatial variation in the skewness and kurtosis in the ground temperatures. This section further describes the distribution properties, focusing on the temporal aspects while considering the spatial heterogeneity of the distributions.

A. OUTLINE

Given that skewness and kurtosis were spatially inhomogeneous, we divided the study area into seven sub-regions\(^1\). The clustering was performed as follows:

(i) The TGH distribution was fitted to $\{\hat{Y}_1(s_i), \ldots, \hat{Y}_{61}(s_i)\}$, where 1–61 represents the index of the 61 days, to estimate the moment parameters $\{a(s_i), b(s_i), g(s_i), h(s_i)\}$ for each grid $s_i$. We used the $l$-moment matching method of Peters et al. [45]. They demonstrated that their method estimates the moment parameters more robustly and accurately than classical moment matching, maximum likelihood, and quantile matching methods.

(ii) $k$-means clustering was applied to $\{a(s_i), b(s_i), g(s_i), h(s_i)\}$, and the 30,572 grids were grouped into seven sub-regions. We used a standard $k$-means method that defines the $k$-mean centers on the basis of the arithmetic means of the subsamples.

The next subsection explains the $l$-moment matching method. The subsequent subsection explains the estimation results for

\(^1\) We first attempted to optimize the number of sub-regions using BIC minimization, but the resulting number of clusters was too large and the clusters were difficult to interpret. We thus used seven clusters, which were interpretable (see Supporting Material B), and analyzed the distributional properties for each sub-region.
the \(l\)-moment parameters. The next subsequent subsection explains the results of spatial clustering based on the estimated moment parameters.

### B. L-MOMENT MATCHING

\(l\)-moments [44] are defined by linear combinations of order statistics. The \(m\)-th \(l\)-moment for a sample of ordered observations \(\{Y_1 \leq \ldots \leq Y_j \leq \ldots \leq Y_m\}\) is defined as

\[
l_m = \frac{1}{m} \sum_{j=0}^{m-1} (-1)^j \binom{m-1}{j} E[Y_{m-j}].
\]  

(26)

An \(l\)-moment exists if and only if the distribution has a mean; thus, \(l\)-moments can characterize a wider range of distributions than classical moments ([45]). The first four \(l\)-moments are expressed as

\[
l_1 = \int_0^1 F_y^{-1}(u)du,
\]

(27)

\[
l_2 = \int_0^1 F_y^{-1}(u)(2u - 1)du,
\]

(28)

\[
l_3 = \int_0^1 F_y^{-1}(u)(6u^2 - 6u + 1)du,
\]

(29)

\[
l_4 = \int_0^1 F_y^{-1}(u)(20u^3 - 30u^2 + 12u - 1)du,
\]

(30)

where \(F_y(u)\) is the distribution function for the random variable \(y\) generated at quantile level \(u\). Because \(l_3\) and \(l_4\) depend on scale \(l_2\), it was suggested in [44] that population \(l\)-skewness \(\tau_3 = l_3/l_2\) and population \(l\)-kurtosis \(l_4/l_2\) be used to measure skewness and kurtosis.

Unbiased estimators for the \(l\)-moments, which are also called sample \(l\)-moments, are given by

\[
\hat{l}_1 = q_0,
\]

(31)

\[
\hat{l}_2 = 2q_1 - q_0,
\]

(32)

\[
\hat{l}_3 = 6q_2 - 6q_1 + q_0,
\]

(33)

\[
\hat{l}_4 = 20q_3 - 30q_2 + 12q_1 - q_0.
\]

(34)

where \(q_m\) equals

\[
q_m = \begin{cases} 
\frac{1}{n} \sum_{j=1}^{n} Y_j & \text{if} \ m = 0 \\
\frac{1}{n} \sum_{j=1}^{n} \frac{(j-1)(j-2)\cdots(j-m)}{(n-1)(n-2)\cdots(n-j)} Y_j & \text{if} \ m > 0. 
\end{cases}
\]  

(35)

An \(l\)-moment matching method was proposed in [45] to estimate the moment parameters in the TGH model. Here we assume a general setting, and \(\{a(s), b(s), g(s), h(s)\}\) are simplified as \(\{a, b, g, h\}\). The proposed estimation has two steps:

(i) Estimate \(g\) and \(h\) by solving

\[
\{\hat{g}, \hat{h}\} = \arg \min \{\tau_3 - \hat{\tau}_3)^2 + (\tau_4 - \hat{\tau}_4)^2, \}
\]

(36)

where \(\hat{\tau}_3 = \hat{\tau}_3/\hat{l}_2\) and \(\hat{\tau}_4 = \hat{l}_4/\hat{l}_2\).
(ii) Estimate \(a\) and \(b\) using their unbiased estimators, resulting in
\[
\hat{b} = \frac{\hat{l}_2}{l_2} \quad \text{(37)}
\]
\[
\hat{a} = \hat{l}_1 - \hat{b}l_1. \quad \text{(38)}
\]

For further details of these estimations, see [45].

1) ESTIMATION RESULTS

The \(l\)-moment matching was applied to the emulated ground temperatures in each grid at 13:00 on the 61 target days \(\{Y_1(s_i), \ldots, Y_{61}(s_i)\}\) to estimate the moment parameters \(\{\hat{a}(s_i), \hat{b}(s_i), \hat{g}(s_i), \hat{h}(s_i)\}\). As shown in Fig. 13, these parameters changed across regions. Both \(\hat{a}(s_i)\) and \(\hat{b}(s_i)\) were higher in the highly urbanized areas, i.e., at the center and toward the west. Urbanization is believed to not only increase the mean temperatures but also to expand the temperature variations, which can cause extreme heat. Unlike the ground temperatures plotted in Fig. 13, \(\hat{a}(s_i)\) was highest to the west of the center. This is reasonable because this area includes an inland area noted for experiencing high heat values. Note that the estimated \(\hat{g}(s_i)\) had positive values, i.e., right-skewness, in most areas. This indicates that temperature distributions have positive skewness, which can introduce extreme temperature. The \(\hat{h}(s_i)\) values were very small in many areas, which is significant as it indicates the probability of temperatures being higher than normally expected.

2) CLUSTERING RESULTS

As described in Section V-A, to analyze the temperature behavior in each region, we divided the target area into seven clusters (see Supporting Material A for a study on the effects of the number of clusters considered). We applied \(k\)-means clustering to the estimated location, scale, skew, and kurtosis related Tukey g-and-h parameters, \(\{\hat{a}(s_i), \hat{b}(s_i), \hat{g}(s_i), \hat{h}(s_i)\}\). The results are plotted in the left panel of Fig. 14. The four parameters were standardized before clustering to eliminate scale dependency. For comparison, the \(k\)-means method considering only mean temperature \(\hat{a}(s_i)\) was also applied. The results (right panel of Fig. 14) simply indicate the tendency of the central area to be the hottest and for the temperature to drop as the distance from the center increases. By contrast, when \(\{\hat{a}(s_i), \hat{b}(s_i), \hat{g}(s_i), \hat{h}(s_i)\}\) were considered, the central area and the north were in the same cluster. The result is intuitively reasonable: these two areas are noted hazardous heat areas due to hot air being carried from central Tokyo by southerly winds. This means that consideration of not only the mean but also the variance, skewness, and kurtosis can reveal the hidden heatwave structure.

The goodness of two clustering results were compared using
\[
D = \frac{1}{n^2} \sum_{C=1}^{7} \sum_{C' = 1}^{7} \frac{\sum_{s_i \in C} \sum_{s_i' \in C'} |\frac{1}{|C||C'|} \|m(s_i) - m(s_i')\|_2^2|}{\sum_{s_i \in C} \sum_{s_i' \in C'} |\frac{1}{|C||C'|} \|m(s_i) - m(s_i')\|_2^2|}, \quad \text{(39)}
\]
where \(m(s_i) \in \{\tilde{a}(s_i), \tilde{b}(s_i), \tilde{g}(s_i), \tilde{h}(s_i)\}\), the four parameters equal \(\{\hat{a}(s_i), \hat{b}(s_i), \hat{g}(s_i), \hat{h}(s_i)\}\) after standardization, \(C\) denotes a cluster, and \(|C|\) is the number of grids in the cluster. Distance \(D\) is large if the seven clusters are well separated: it was 1.05 for (a) and 0.86 for (b). Consideration of not only \(\hat{a}(s_i)\) but also \(\{\hat{b}(s_i), \hat{g}(s_i), \hat{h}(s_i)\}\) was found to be necessary to cluster ground temperatures accurately while consideration of skewness and kurtosis was found to be necessary to determine heatwave risk (see Section I).

Fig. 15 displays boxplots of the parameters for each cluster. It shows that the seven clusters are distinctive; i.e., there is little overlap of the value ranges for the mean, variance, and skewness parameters. These three parameters tended to have higher values near the center and lower values in the peripheral areas, indicating that the temperature distribution property greatly depends on the degree of urbanization.

![Figure 15. Boxplots of estimated parameters for each cluster. Cluster 1 is nearest the center, and cluster 7 is farthest from the center.](image-url)
VI. CONCLUSION

This study analyzed the spatial and temporal skew and kurtosis behavior of ground temperatures, which were emulated by combining ground and remotely sensed observations. An $l$-moment matching approach was used for temporal analysis, and low-rank and sparse TGH-RF models were used for spatial analysis. The results demonstrated the importance of considering skewness and kurtosis in heatwave modeling. Specifically, strong skewness values were estimated at the center of Tokyo, which are known to experience a considerable amount of heat. The low-rank and sparse TGH-RF models were found to be useful for revealing tail structures in large spatial data.

Several issues remain to be addressed. First, we need to extend our low-rank and sparse TGH models to dynamic spatio-temporal models. It is also important to quantify the heatwave “risk” that emerges when hazards (e.g., high temperature), exposure (e.g., many people are outside), and vulnerability (e.g., many of them are elderly) are all set in the analysis. Fortunately, data relating these factors are increasingly available. For example, MODIS ground temperatures are useful for evaluating heatwave hazards (i.e., higher temperatures correspond to greater probabilities of heatwave hazards). Moreover, mobile GPS data can be used to determine how many people are exposed to a hazard. National census data including the elderly ratio and household income are valuable for estimating the ratio of people who are vulnerable to heat. District-level heatwave risk estimation using a wide variety of micro spatial and temporal information would be an important step towards data-driven heat risk management, which is increasingly important as global warming advances.

VII. SUPPORTING MATERIAL

A. CLUSTER ANALYSIS WITH DIFFERENT NUMBER OF CLUSTERS

Cluster analysis of the parameters $\{a(s_i), b(s_i), g(s_i), h(s_i)\}$ as in Section V-B2 demonstrated that $n_C = 7$ is reasonable, where $n_C$ is the number of clusters.

Fig. 13 plots the clustering results when $n_C$ was optimized by minimizing the Bayesian information criterion (BIC). In this case, $n_C = 49$ was found to be optimal. However, the resulting clusters were extremely difficult to interpret. To obtain an interpretable clustering result, we restricted the number of clusters to at most 10.

Fig. 14 plots the Akaike information criterion (AIC) and BIC of the clusters when $n_C$, was 3, 4, ..., 10. Both decreased as $n_C$ increased. However, the decrease became relatively slow after around $n_C = 7$. Fig. 15 displays the clustering results when $n_C$ was 3, 7, and 10. When $n_C = 3$, the center and suburban areas merged into one cluster. However, the central area is likely to have different heat behavior. Thus, the results for $n_C = 3$ are too simple. The results for $n_C = 7$ have a relatively clear spatial cluster pattern whereas those for $n_C = 10$ have a more mosaic-like pattern.

B. SIMULATION EXPERIMENT USING LOW-RANK TGH-RF MODEL

We compared our developed low-rank TGH-RF model to the original TGH-RF model of [14] in terms of computational time and estimation accuracy of $a, b, g, h$.

Samples for July 1 for the cluster including the central area were used for the comparison. In the low-rank approach, the number of eigenpairs in our approximation was increased from 200 to 1200 by 200. The calculations were performed on a Windows 10 64-bit system with 48 GB of memory and coded using R (version 3.4.2).

As shown in Table 2, the computational time required for the low-rank TGH-RF model estimation was much less than that for the original model, especially when $L$ was small. Interestingly, the low-rank approach took only 51 s even when $L = 1200$ while the original full rank model, which implies $L = n = 1500$, took 3,494 s. This demonstrates...
the computational efficiency of our approach, i.e., estimating scale parameter \( m \) instead of range parameter \( r \).

Fig. 23 plots the estimated parameters. Those estimated using the low-rank approach were reasonably similar to those for the original approach when the number of eigenpairs was equal to or greater than 600.

In summary, the low-rank approach accurately estimated the \( a, b, g, h \) parameters in a computationally efficient manner.

C. SIMULATION EXPERIMENT USING SPARSE TGH-RF MODEL

We examined the estimation accuracy of the sparse TGH-RF approach by simulation. TGH-RF was generated 100 times on 39 by 39 grids, and the estimation accuracy of the \( g \) and \( h \) parameters at the center (20, 20) of the grid space was evaluated by fitting the sparse TGH-RF model. The center is denoted by \( s_0 \). Following the process described in Section IV-E5, we set the sample size used to estimate the
FIGURE 22. Boxplots of estimated $h(s_f)$ (kurtosis) parameter for each cluster.

FIGURE 23. Comparison of estimated parameters for July 1 for cluster including central area (black: eigen-approximation; red: exact).

FIGURE 24. Boxplots of $h(s_0)$ and $g(s_0)$ parameters. Red horizontal lines indicate true parameter values.
model for $s_0$ to or less than 200. An exponential kernel $\exp(-d(s_0, s)/r(s_0))$ was used to model spatial dependence, where $r(s_0)$ is an unknown parameter. The true values for $a(s_0)$, $b(s_0)$, and $r(s_0)$ were 0, 1, and 1, respectively. The true values for the $g(s_0)$ and $h(s_0)$ parameters were specified as $g(s_0) \in \{0.5, 0.0, 0.5\}$ and $h(s_0) \in \{0.0, 0.25, 0.5\}$; for each case, the sparse TGH-RF model was fitted 100 times.

Boxplots of the estimated $g(s_0)$ parameters are displayed at the top of Fig. 24. Our approach successfully detected a positive $g(s_0)$ value when the true value was positive. The same is true for negative $g(s_0)$. Although the error tended to increase as $h(s_0)$ was increased, the estimates remained unbiased. Our approach was found to detect skewness accurately. Boxplots of the estimated $h(s_0)$ parameters are displayed at the bottom of Fig. 24. The estimates are nearly unbiased in all cases, and the variances of the estimates are similar across cases. Our approach is thus accurate for kurtosis as well.

### References

1. C. B. Initiative. (2019). *Green Bonds Reach Record 255bn for CY 2019–New Milestone 350-400bn Climate Bonds Initial Forecast for 2020*: 1Trillion in Annual Green Investment in Sight for Early 2020. Available: https://www.climatebonds.net/resources/press-releases/202001/green-bonds-reach-record-255bn-cy-2019-new-milestone-350-400bn

2. J. C. Semenza, C. H. Rubin, K. H. Falter, J. D. Selanikio, W. D. Flanders, H. L. Howe, and J. L. Wilhelm, “Heat-related deaths during the July 1995 heat wave in Chicago,” *New England J. Med.*, vol. 3, no. 2, pp. 84–90, 1996.

3. T. R. Oke, “The energetic basis of the urban heat island,” *Quart. J. Roy. Meteorol. Soc.*, vol. 108, no. 455, pp. 1–24, Jan. 1982.

4. Q. Weng, D. Lu, and J. Schubring, “Estimation of Land Surface Temperature–vegetation abundance relationships for urban heat island studies,” *Remote Sens. Environ.*, vol. 89, no. 4, pp. 467–483, Feb. 2004.

5. E. L. Glaeser and M. E. Kahn, “The greenness of cities: Carbon dioxide emissions and urban development,” *J. Urban Econ.*, vol. 67, no. 3, pp. 404–418, May 2010.

6. S. Kidney, B. Sonerud, and P. Oliver, “Growing a green bonds market in China,” in *Proc. Climate Bonds Initiative, Int. Inst. Sustain. Develop. (IBSD)*, 2015.

7. J. A. Matross, S. Kidney, M. Mallon, and M. Hughes, “Mobilizing private finance to drive an energy industrial revolution,” *Energy Policy*, vol. 38, no. 7, pp. 3263–3265, Jul. 2010.

8. B. Buchner, A. Falcoro, M. Hervé-Mignucci, C. Trabacchi, and M. Brinkman, “The landscape of climate finance,” *Climate Policy Initiative, Venice*, vol. 27, 2011.

9. D. Murakami, G. W. Peters, Y. Yamagata, and T. Matsui, “Participatory sensing data tweets for micro-urban real-time resiliency monitoring and risk management,” *IEEE Access*, 2016, pp. 347–372.

10. I. D. Stewart and T. R. Oke, “Local climate zones for urban temperature studies,” *Bull. Amer. Meteorological Soc.*, vol. 93, no. 12, pp. 1879–1900, Dec. 2012.

11. M. Jorge and I. Boris, “Some properties of the tukey g and h family of distributions,” *Commun. Statist.-Theory Methods*, vol. 13, no. 3, pp. 353–369, Jan. 1984.
GARETH W. PETERS received the B.Sc. and B.Eng. (Hons.) degrees in mathematics and electrical engineering from The University of Melbourne, VIC, Australia, the M.Sc. (by research) degree from the Statistical Signal Processing Laboratory, University of Cambridge, Cambridge, U.K., and the Ph.D. degree in statistics from the University of NSW, Sydney, Australia. He is currently a Lecturer with the Department of Statistical Science, University College London, and with the Department of Mathematics and Statistics, University of NSW. He is also a Visiting Scientist with the Commonwealth Scientific and Industrial Research Organisation, Australia. His main research interests include Bayesian statistical modeling and sampling theory (Markov chain Monte Carlo and sequential Monte Carlo).

DAISUKE MURAKAMI received the Ph.D. degree in engineering from the University of Tsukuba, Tsukuba, Japan, in 2014. From 2014 to 2017, he was a Research Associate at the National Institute for Environmental Studies, Japan. Since August 2017, he has been an Assistant Professor at the Department of Statistical Data Science, The Institute of Statistical Mathematics, Japan. His research interests include spatial statistics, quantitative geography, and urban analysis.

TOMOKO MATSUI (Senior Member, IEEE) received the Ph.D. degree in computer science from the Tokyo Institute of Technology, Tokyo, Japan, in 1997. From 1988 to 2002, she was a Researcher at several NTT laboratories, focusing on speaker and speech recognition. From 1998 to 2002, she was a Senior Researcher at the Spoken Language Translation Research Laboratory, ATR, Kyoto, focusing on speech recognition. In 2001, she was an Invited Researcher at the Acoustic and Speech Research Department, Bell Laboratories, Murray Hill, NJ, USA, working on finding effective confidence measures for verifying speech recognition results. She is currently a Professor at The Institute of Statistical Mathematics, Tokyo, working on statistical spatial-temporal modeling for various kinds of applications, including speech and image recognition. She received the Best Paper Award from the Institute of Electronics, Information, and Communication Engineers of Japan, in 1993.

YOSHIKI YAMAGATA received the Ph.D. degree in system science from The University of Tokyo, Tokyo, Japan, in 1985. Since 1991, he has been with the National Institute for Environmental Studies, where he is currently studying climate change risk management with the Center for Global Environmental Research. He is also associated with the International Institute for Applied Systems Analysis, Vienna, and The Institute of Statistical Mathematics, Tokyo. He has lectured at The University of Tokyo, the University of Tsukuba, and Hokkaido University. His recent research interests include land-use models, resilient urban planning, and social networks. Internationally, he has served as a Lead Author of a report by The Intergovernmental Panel on Climate Change, as a member of the Scientific Steering Committee of the Future Earth Global Carbon Project, and as a member of the Editorial Board of Applied Energy and Climate Policy.

* * *