Methodology for Evaluating Influences of Anthropogenic Factors on Temperature Formation of Summer Seasons Considering their Randomness -Verification Through the Hierarchical Bayesian Model-

Takashi Kyakuno*

1Professor, School of Policy Studies, Kwansei Gakuin University, Japan

Abstract

A new regression model that considers random nature is proposed for describing summer air temperatures in urban areas through the concept of an empirical hierarchical Bayesian model. In this model, the regression coefficients of explanatory variables are regarded as having random effects because of their locations, whereas ordinal regression models do not allow such randomness. Data of air temperature, the climatic situation, and geographically related factors in the Kinki area were collected and compiled by GIS, and then applied to both the new model and the standard regression models. The performances of the models were compared and their differences were discussed in relation to the geographic or climatic situations. It was established that the proposed model succeeded in accounting for the random nature and it predicted air temperature more accurately than the standard models based on the MCMC simulation. Furthermore, the confidence intervals of the values predicted by the new model became more rigid than for the ordinal regression models. This means that the new model is capable of presenting a more accurate picture without affecting the convenience of the ordinary regression models.

Keywords: Summer temperatures; Random effects; Empirical Bayesian model; GIS; MCMC

1. Introduction

In response to the heat island problem of recent years, arguments have been repeated regarding appropriate methods for city planning. In particular, appropriate estimates of the influences of land use on temperature have important consequences for land use plans. To appropriately estimate such influences, the direct approach is to build regression models that make temperature the objective variable. By evaluating the regression coefficients, it is possible to not only estimate the influence of each land use type but also forecast climatic change in urban areas. There have been various methodologies proposed for such evaluations: the ordinal regression model, extended regression model, models that consider statistical correlations, and neural network models (Honjo et al., 2000; Jeong et al., 2015; Kotani et al., 1996; Li et al., 2003; Paez et al., 1998; Shudo et al., 1996; Tanimoto et al., 1996). Considerable knowledge has been acquired regarding the relation between land use and temperature, especially the reduction of temperature in green areas.

Most models implicitly treat the influence of land use on temperature as fixed. However, the degree of influence should be supposed changeable and dependent on the geographical situation, as has been discussed in a previous research (Irie, 2003). Thus, building a model that can accommodate the changeability of the parameters and coefficients in relation to the geographic situation is necessary. However, this could lead to problems of model stability because some parameters do not converge in the simulation process. Therefore, it becomes more important to build a model that allows changeability of the regression coefficients with a certain degree of control, without spoiling the merits of the regression model. Various methodologies have been proposed for such simulations, e.g., models that treat the random nature directly, such as the generalized linear mixed model and geographically weighted regression model, or models that consider randomness in an indirect way, such as the multilevel model or spatial autoregressive model. These models are used in many academic fields not only for estimating urban climate but also for forecasting areas inhabited by specific animals and plants, estimating land price distributions, or estimating domestic energy demands (Ando et al., 2014; Furutani, 2004; Itagawa et al., 2012; Kobayashi, 2013; Onishi et al., 2010). These approaches have the advantages that the results are comparatively simple and they can be understood easily and intuitively. Most importantly, the results can be used directly in planning theory. Nevertheless, there are limitations; simply
considering the complex mutual relations among many different factors and expressing them in mathematical logic is not sufficient. Proposing a new model that can account for the more complex relations by enhancing the changeability and extendibility is necessary. For this purpose, the hierarchical Bayesian method can be adopted because within this framework, complex relations among many factors can be flexibly expressed and the intensities of the relations between specified factors can be set as desired (Clark and Gelfand, 2006; Congdon, 2003; Congdon, 2006).

In this research, a new model is proposed to explain the summer temperature of urban areas. This is achieved through the extension of the regression model by adopting the concept of the hierarchical Bayesian method to consider random effects. The model is verified, and its performance is assessed both quantitatively and statistically. The trends of the fluctuations of the regression coefficients are also verified and discussed in relation to the surrounding land use or geographical conditions. The novelty of this study exists in discussing the overestimation or underestimation of regression coefficients caused by random effects of atmospheric phenomena in regression models and proposing a methodology for reducing them by using the frame of the Bayesian inference method.

2. Methodology and Research Materials

Initially, three temperature indices on the basis of observations during two periods (July 21 to September 8, 2011 and July 21 to September 8, 2012) were determined. These indices represented the five-day arithmetic average of the daily maximum temperature, five-day arithmetic average of the daily minimum temperature, and five-day arithmetic average of the daily mean temperature. The proposed hierarchical Bayesian model was used to simulate these phenomena, as were two traditional models for the purpose of assessment of the performance of the new model. The first traditional model was an ordinal regression model (1). The ordinal regression model and the second was an ordinal regression model that incorporated an item indicating time intervals of the observations (2. The ordinal regression model with time-interval item). The second model can express fluctuations in the time series of temperature. In this study, both models partially adopted the concept of the hierarchical Bayesian method. The newly proposed model incorporated the random nature of the coefficients, which allows the consideration of the fluctuations of the coefficients by observation points, as well as by the time-interval item, using the hierarchical Bayesian method (3. The random effect model with time-interval item). The details of the models are described in chapter 3.

Table 1. Factors of Temperature Formation (the explanatory variables)

| Name | Detail | Data source |
|------|--------|-------------|
| Daily wind velocity (Wind velocity) | The mean of daily average wind velocity in each observation interval. | AMeDAS by Japan Meteorological Agency |
| Daily number of sunshine hours (Sunshine hours) | The mean of daily average number of sunshine hours in each observation interval. | AMeDAS by the Japan Meteorological Agency |
| Daily precipitation (Precipitation) | The mean of daily average precipitation in each observation interval. | AMeDAS by the Japan Meteorological Agency |
| Distance from the central line of the river (Distance from river) | The distance from the nearest central line of the river. It is computed by GIS with the 10m spatial resolution. | National land numerical information (Rivers) Created in 2009 |
| Distance from the coastline (Distance from coastline) | The distance from the coastline. It is computed by GIS with the 10m spatial resolution. | National land numerical information (Coastal lines) Created in 2006 |
| Altitude | The altitude of the observation point. | Digital map 25000 (Spatial data framework) |
| Normalized Differntial Vegetation Index (NDVI) | The mean of NDVI values in the circular region of a 100 m radius. The values were computed from the radiance of the near infrared band and visible band. | ALOS AVNIR-2 (2009.4.19) Scene ID:YKP000 ALA2VA172272900_0 |
| Land use (Building areas) | Logical value. It presents TRUE when the land use of the observation point corresponds to building areas. | National land numerical information (Land utilization tertiary mesh data) Created in 2009 |
| Land use (Forest) | Logical value. It presents TRUE when the land use of the observation point corresponds to forests. | National land numerical information (Land utilization tertiary mesh data) Created in 2009 |
| Land use (Agricultural areas) | Logical value. It presents TRUE when the land use of the observation point corresponds to agricultural areas. | National land numerical information (Land utilization tertiary mesh data) Created in 2009 |
| Designated commercial land use zone (Commercial zone) | The area of the commercial district or neighborhood commercial district designated by the city planning law in the circular region of a 100 m radius. | National land numerical information (Land use zoning) Created in 2011 |
| Mesh population (Population) | The population of an approximately 1km × 1 km quadrat in which the observation point is located 1). | The population census of 2010 |
| Mesh road density (Road density) | The total length of roads in an approximately 1km × 1 km quadrat in which the observation point is located 2). | National land numerical information (Road Density / Road Length Mesh Data) Created in 2010 |

1) Because Kumatori is located on the border of two quadrats, their average value was adopted as its representative.
2) Same as 1)
To assess the performance of the new model, the following four points were investigated. (1) The correlation coefficients between the actual measurements and simulated predicted values. (2) The probability that the actual observed values fall into the range of the 95% confidence intervals of the simulated values. (3) The width of the 95% confidence interval of the simulated values. (4) The DIC (Deviance) of the results of each model. Furthermore, the differences of each regression coefficient for the observation points were verified, and the influences of geographic and land use factors on such differences were analyzed.

The study area of the research was in the Kinki area: Osaka prefecture and the southern part of Hyogo prefecture in Japan (Fig.1.). Temperature data from 42 observation points were collected from the following sources: "AMeDAS," issued by the Japan Meteorological Agency; "The homepage of the constant monitoring of air pollutants," issued by Osaka prefecture; "The situation of the atmospheric environment of Hyogo," issued by the Environmental Bureau of the Hyogo prefectural government; and "The system of constant monitoring of the environment of Kobe," issued by Kobe city. The temperature data used in this study covered 100 days; the first 50 were from July 21 to September 8, 2011, and the second 50 were from July 21 to September 8, 2012. From these, the daily maximum temperature, daily minimum temperature, and daily mean temperature were calculated for every observation point. Then, the total observation period was divided into 20 intervals of 5 days, and the average values of each temperature index were calculated for each interval. These were set as the representative values for each of the 42 locations and 20 intervals, providing 840 data elements for the analysis.

Thirteen factors (explanatory variables) that influence temperature were determined based on previous researches, and they are listed in Table 1. However, the daily wind velocity, daily number of sunshine hours, and daily precipitation data were not available for all observation points. Therefore, missing values were estimated based on AMeDaS point data using the spline spatial interpolation method. For the Normalized Difference Vegetation Index (NDVI), because reliable data for the summer seasons of recent years were unavailable, data from late April 3 were used instead. For the purpose of complements, the land use data specified for the classes of forests and farmland were prepared and incorporated in the model.

3. Model Details

In this study, the regression coefficient vector and regression coefficient matrix were computed in each model based on the hierarchical Bayesian method. The details of each model are described below.

Model 1: The ordinal regression model

Presuming that temperature is expressed by the regression model and that the errors follow the normal distribution of the mean 0 and precision (reciprocal of variance) \( \beta_i \), the regression coefficient vector including the intercept is expressed as \( \beta \), presuming that \( w \) and \( \beta_i \) follow the distributions mentioned below. The function of the posterior probability can be obtained as the product of the likelihood functions and the function of the prior probability. The likelihood function is also given in the form of the products of the normal distribution function \( g_i \). The prior distribution of each element \( w_i \) of \( w \) is also assumed to follow the normal distribution of the mean 0 and precision \( 1.0 \times 10^{-3} \), which is expressed as \( h_i \). Furthermore, the prior distribution of \( \beta_i \) is presupposed as a uniform distribution. Consequently, the distribution of the posterior probability is given as shown in Eq. (1).

\[
f_i(w, \beta_i) \propto k(\beta_i) \times \prod_k h_i(w_k) \times \prod_{ij} g_i(y_{ij}|w, \beta_i)
\]

(1)

The details of functions and the symbols are as follows:

\[
g_i(y_{ij}|w, \beta_i) = \text{Norm}(w_0 + w_1 WD_{ij} + w_2 S_{ij} + w_3 PR_{ij} + \sum_{k=1}^{645} w_k x_{ik}, \beta_i)
\]

\[
h_i(w_k) = \text{Norm}(0, 1.0 \times 10^{-3})
\]

\[
k(\beta_i) = \text{Unif}(0, 100)
\]

where \( i \) is the number of observation points, and \( j \) is the number of observation time intervals. \( y_{ij} \) refers to the temperature (the average) at observation point \( i \) and time interval \( j \). \( WD_{ij} \) refers to the five-day average of daily wind velocity at point \( i \) and time interval \( j \). \( S_{ij} \) refers to the daily number of sunshine hours, and \( PR_{ij} \) refers to the daily precipitation at point \( i \) and time interval \( j \). Moreover, coefficient \( w_i \) refers to the regression coefficient of the k-th explanatory variable (the intersect is included), and \( x_{ik} \) expresses the k-th explanatory variable (geographic factors other than the climatological factors shown in Table 2.) at point \( i \). Norm (-) means the normal distribution function and it possesses two parameters: the mean and precision. Unif (-) means the uniform probability distribution of an interval indicated by the parameters.

Model 2: The ordinal regression model with the time-interval item

Based on the same idea as model 1, the distribution function of the posterior probability is given in Eq. (2). In this model, the errors are assumed to follow the normal distribution of the mean 0 and precision \( \beta_i \), as in model 1. However, this model incorporates the term of the observation time interval as \( t_i \), which has the same value at every observation point within the same time interval. The function of the prior probability of coefficient \( w_i \) is given by the same function as in model 1, and the term \( t_i \) is also assumed to be given as the normal distribution of the mean 0 and precision \( \beta_i \). The distribution function of the prior probability of \( \beta_i \) is...
given as a uniform distribution and that of $\beta_2$ is given as a gamma function of the shape parameter $a = 1000$ and rate parameter $b = 1000$. As the distribution function of the posterior probability is proportional to the product of the likelihood function and the prior distribution, it results in the form shown in the Eq. (2).

$$f_2(\mathbf{w}, t, \beta_1, \beta_2) \propto k(\beta_1) \times m(\beta_2 | a, b) \prod_k h_k(\mathbf{w}_k) \times \prod_i \ell(t_i | \beta_2) \times \prod_{ij} g_2(y_{ij} | \mathbf{w}_i, t_i, \beta_1)$$  \hspace{2cm} (2)

The details of functions and the symbols are as follows:

$$g_2(y_{ij} | \mathbf{w}_i, t_i, \beta_1) = \text{Norm}(w_0 + w_i \cdot \text{WD}_{ij} + w_2 \cdot s_{ij} + w_3 \cdot \text{PR}_{ij} + \sum_{k=4}^{13} w_k \cdot x_{ik} + t_i \cdot \beta_1)$$

$$l(t_i | \beta_2) = \text{Norm}(0, \beta_2)$$

$$m(\beta_2 | a, b) = b^a \beta_2^{-a-1} \Gamma(a)^{-1} \exp(-b \beta_2)$$

where $t_i$ is the interval term of the $j$-th time interval, which takes a constant value at every observation point dependent only on the observation intervals, and $t_i$ expresses the vector consisting of $t_i$. The other parameters and terms are the same as in model 1.

Model 3: The random effect model with the time-integer

This model is an extended version of model 2, in which the regression coefficients are considered to change by observation point. The function of the posterior joint probability of $\mathbf{W}$ and $t$ is given in Eq. (3), and the likelihood function of temperature is given by Eq. (4). As randomness by observation point is allowed, the regression coefficients are expressed in the form of matrix $W_i$, in which the regression coefficients vector $\mathbf{w}_i$ of the $i$-th point form arrays in the columns. The two subscripts $i$ and $j$ of matrix $W_i$ refer to an observation point and an explanatory variable number, respectively.

The function of the prior probability of the regression coefficient vector $\mathbf{w}_i$ is given as a multivariate normal distribution that has parameters of the mean vector $\mu$ and precision matrix $\Lambda_w$, which is the inverse matrix of the variance–covariance matrix acquired from the result of the simulation of model 2. Thereby, regression coefficients are allowed to fluctuate by observation point, although they are restricted to move around the outskirts of the averages of each coefficient.

$$f_3(\mathbf{W}, t, \beta_1, \beta_2) \propto k(\beta_1) \times m(\beta_2 | a, b) \times \prod_i h_i(\mathbf{w}_i | \Lambda_w) \times \prod_i l(t_i | \beta_2) \times \prod_{ij} g_3(y_{ij} | \mathbf{w}_i, t_i, \beta_1)$$ \hspace{2cm} (3)

$$g_3(y_{ij} | \mathbf{w}_i, t_i, \beta_1) = \text{Norm}(w_{i0} + w_{i1} \cdot \text{WD}_{ij} + w_{i2} \cdot s_{ij} + w_{i3} \cdot \text{PR}_{ij} + \sum_{k=4}^{13} w_{ik} \cdot x_{ik} + t_i \cdot \beta_1)$$ \hspace{2cm} (4)

The details of functions and the symbols are as follows:

$$h_2(\mathbf{w}_i | \Lambda_w) = (2\pi)^{-14/2} |\Lambda_w|^{1/2} \exp\left[-(\mathbf{w}_i - \bar{\mathbf{w}})^T \Lambda_w (\mathbf{w}_i - \bar{\mathbf{w}})/2\right]$$

where $w_{i}$ indicates the coefficient to the $k$-th explanatory variable of point $i$ and $(-)^T$ in $h_i$ means the transposed matrix. The other parameters and terms are the same as in models 1 and 2.

The calculations of these models were performed using the MCMC simulation based on a hierarchical Bayesian model (Clark and Gelfand, 2006; Albert, 2009). The scripts were described and implemented in R2.14.1 and WinBUGS1.4.3 (R Development Core Team, 2011; Spiegelhalter et al., 2003). The MCMC simulation was performed using three chains with 3000 repetitions. Burn-in was set at 1000 and the thinning rate was set at 2. Statistical values were computed from the 1000 samples obtained from the averaged values of the three chains of simulations. For the predictions of temperature and calculations of $\mathbf{w}$ and $\Lambda_w$ in model 3, 1000 samples from the third chain were used.

4. Results and Discussion

Table 2. presents the averages of coefficient $w_i$ for the daily maximum temperature, daily minimum temperature, and daily mean temperature obtained from the above model simulations. The results from model 1 for daily maximum temperature show that the regression coefficients of daily wind velocity and daily precipitation have positive values and contribute to the rise of temperature, whereas these coefficients are negative and contribute to the reduction of temperature in the case of model 2. For the daily minimum temperature, the absolute value of the coefficient of daily wind velocity in model 1 is smaller than in either model 2 or model 3. It is considered that this is because of periodic changes inherent in the climatic phenomenon appearing indirectly in these coefficients. The meaning of each coefficient might become more appropriate and acute if regression models were built in which the observation periods were divided into shorter terms, rather than building models that use data pooled from the entire period.

As model 3 includes the premise of randomness of the regression coefficients, the average of each coefficient can be regarded as the average of the coefficients of the observation points. The differences to model 2 can be attributed to the distance from the central line of the river, distance from the coastline, designated commercial land use zone, mesh population, and mesh road density. The order is reduced to approximately 1/100 in model 3. This is because fluctuations of the values by observation point are canceled by each other and they take values near 0 as a result. This is why it should not be concluded that these factors have no influence on temperature. This problem will be discussed later in relation to the fluctuations of the coefficients by observation point.

The performance of each model was verified by comparison of actual temperatures with the simulated model 1 the correlation coefficients between the actual obtained from the three models (Table 3.). In measurements and predicted values of the daily values maximum temperature and daily mean temperature are high, i.e., 0.841 and 0.826, respectively. Furthermore,
Table 2. Results of Regression Coefficients Obtained from the Simulations

| Daily maximum temperature | Model 1       | Model 2       | Model 3       |
|---------------------------|---------------|---------------|---------------|
| Intercept                 | 2.80E+01      | 3.10E+01      | 3.11E+01      |
| Wind velocity             | (5.31E+01)    | (3.33E+01)    | (4.04E+01)    |
| sunshine hours            | 5.38E-01      | 2.37E-01      | 2.26E-01      |
| Precipitation             | (1.37E+01)    | (2.62E+01)    | (2.99E+01)    |
| Distance from river       | 4.16E+04      | 4.81E+04      | 3.74E+04      |
| Distance from coastline   | 5.75E-05      | 4.37E-05      | 4.15E-05      |
| Altitude                  | -1.18E-02     | -1.77E-02     | -1.20E-02     |
| NDVI                      | 3.64E+00      | 2.64E+00      | 2.64E+00      |

| Daily mean temperature    | Model 1       | Model 2       | Model 3       |
|---------------------------|---------------|---------------|---------------|
| Intercept                 | 2.19E+01      | 2.27E+01      | 2.27E+01      |
| Wind velocity             | (3.99E+01)    | (2.48E+01)    | (3.01E+01)    |
| sunshine hours            | 2.25E-01      | 1.62E-01      | 1.62E-01      |
| Precipitation             | (6.10E-02)    | (2.58E+02)    | (2.86E+02)    |
| Distance from river       | 1.01E+04      | 2.30E+04      | 2.90E+06      |
| Distance from coastline   | 2.26E-06      | 2.45E-05      | 1.99E-07      |
| Altitude                  | -5.96E-03     | -6.21E-03     | -6.00E-03     |
| NDVI                      | 1.50E+00      | 5.84E-01      | 4.85E-01      |

* In each column, the upper figure indicates regression coefficient, and the lower figure indicates t-value (regression coefficient divided by its standard error).

Table 3. Indices Showing the Performance of Each Model Simulation

| Daily maximum temperature | Model 1       | Model 2       | Model 3       |
|---------------------------|---------------|---------------|---------------|
| Correlation               | 0.84          | 0.905         | 0.94         |
| RMSE                      | 795.840       | 794.840       | 795.840       |
| DIC                       | 2784          | 2423          | 2066          |
| (Deviance)                | 2769          | 2389          | 2000          |

| Daily mean temperature    | Model 1       | Model 2       | Model 3       |
|---------------------------|---------------|---------------|---------------|
| Correlation               | 0.823         | 0.891         | 0.921         |
| RMSE                      | 1.31E+01      | 1.02E+01      | 0.876         |
| DIC                       | 2879          | 2453          | 2185          |
| (Deviance)                | 2878          | 2453          | 2185          |

Table 4. The Result of the Cross Validation of Each Model

| Daily maximum temperature | Model 1       | Model 2       | Model 3       |
|---------------------------|---------------|---------------|---------------|
| Correlation               | 0.832         | 0.891         | 0.921         |
| RMSE                      | 795.840       | 794.840       | 795.840       |
| DIC                       | 2784          | 2423          | 2066          |
| (Deviance)                | 2769          | 2389          | 2000          |

| Daily mean temperature    | Model 1       | Model 2       | Model 3       |
|---------------------------|---------------|---------------|---------------|
| Correlation               | 0.824         | 0.904         | 0.933         |
| RMSE                      | 1.31E+01      | 1.02E+01      | 0.876         |
| DIC                       | 2879          | 2453          | 2185          |
| (Deviance)                | 2878          | 2453          | 2185          |

* In each column, the upper figure indicates regression coefficient, and the lower figure indicates t-value (regression coefficient divided by its standard error).

The daily minimum temperature also shows a comparatively high value of 0.781. In model 2, the correlation coefficients of all the regression coefficients are approximately 0.9. However, in model 3, in which the random nature of each regression coefficient is presumed, the correlation coefficients exceed 0.9, and the value of the daily minimum temperature is considerably improved, attaining a value of 0.921.

The ratios with which the observed values fall into the 95% confidence intervals of the simulated values were also verified. In all models and for all temperature indices, 790 to 805 cases of the 840 fell within the specified range. In model 1, the widths of the 95% confidence intervals of the simulated values of daily maximum temperature, daily mean temperature were 4.96, 5.30, and 4.63, respectively. In model 2, these widths were reduced by approximately 1 °C, i.e., they were 3.99, 4.15, and 3.56, respectively. The reduction was even greater in model 3, decreasing to 3.23, 3.58, and 3.01, respectively. This means that in model 3, the reduction was approximately 1.7 °C compared with model 1. This demonstrates the improvement in reliability achieved by model 3.

With regard to the DIC, which illustrates the model performance, the value for model 3 is smaller than model 1. Particularly for the daily mean temperature, the value for model 3 is 1945, which is a reduction of approximately 700 compared with the value for model 1. This means that even when considering the penalties due to the increase in the number of parameters and the complexity of the model, the performance of the model is improved.

Table 4. shows the result of the cross validation of these three models, which exhibits their abilities as a prediction model. However, it should be noticed that figures in model 2 and model 3 are not results of the cross validation in the rigid sense of the meaning. Figures shown in the table are mean values of correlations and RMSE in two trials of the validation process. In the first trial, data of the first half periods are used as training data and data of the latter half...
Table 5. Lists of Coefficients of Variation

| Observation point | Distance from coastline | Designated commercial land use zone | Mesh road density |
|-------------------|------------------------|-------------------------------------|------------------|
|                   | Daily maximum temperature | Daily minimum temperature | Daily mean temperature |
| Intercept          | 0.015                  | 0.016                             | 0.012             |
| Wind velocity      | -0.104                 | 0.120                             | -0.275            |
| Sunshine hours     | 0.095                  | 0.106                             | 0.058             |
| Precipitation      | -0.330                 | -0.253                            | -0.253            |
| Distance from river| 5.628                  | 9.597                             | 5.687             |
| Distance from coastline| 4.675             | -10.022                           | -176.213          |
| Altitude           | -0.009                 | -0.019                            | -0.012            |
| NDVI               | 0.047                  | 0.370                             | 0.116             |
| Building areas     | 0.031                  | 0.076                             | 0.051             |
| Forest             | -0.056                 | -0.098                            | -0.081            |
| Agricultural areas | 0.300                  | -0.125                            | -0.199            |
| Commercial zone    | 15.285                 | 24.677                            | 23.747            |
| Population         | -15.680                | -14.096                           | -6.100            |
| Road density       | 9.900                  | 3.589                             | 3.155             |

Therefore, the author used mean values of temperature in each period of testing data in order to give “t”.

In Table 4, models 2 and 3 exhibit better performances than model 1 in terms of both correlation and RMSE. The RMSE of model 2 is improved by approximately 0.2 compared with model 1 in all cases, and that for model 3 is further improved compared with model 2. This also shows that they also work as predictors on new unknown cases. Unless considering random effects, the predictors will return temperatures with errors of 1.1–1.3 °C. However, the error can be reduced to 0.8–0.9 °C if random effects are considered using the Bayesian framework.

The fluctuations of each regression coefficient by observation point were verified in model 3. To perform the verification, the mean and standard deviation of each coefficient at each point were calculated. Then, the standard deviation was divided by the mean, which is called a coefficient of variation (Table 5). If the absolute value of this value is large, it means that the fluctuations are comparatively large compared with the size of the coefficient itself. It can be seen that the distance from the central line of the river, distance to the coastline, designated commercial land use zone, mesh population, and mesh road density all show large numbers, whereas climatic factors such as the daily wind velocity, daily number of sunshine hours, and daily precipitation periods are used for testing its ability of prediction. In the second trial, the data of the latter half periods are used as training data and data of the first half periods are used as testing data. In models 2 and 3, the term “t” which expresses time-series fluctuation is required to be provided beforehand. However it cannot be provided only from the training dataset because the periods of testing data are different from those of training data.

Table 6. Lists of Regression Coefficients of Each Point (i.e., Five Largest Points and Five Smallest Points)

| Observation point | Distance from the coastline | Designated commercial land use zone | Mesh road density |
|-------------------|-----------------------------|-------------------------------------|------------------|
| Daily maximum temperature |                     |                                 |                  |
| Five largest      | 34. Sennnari (school) | 3.74E-06                            | 5.723E-06          |
|                    | 34. Sennnari (school) | -3.442E-06                          | -0.330            |
|                    | 32. Suehiro park      | -2.60E-06                           | -1.91E-06          |
|                    | 34. Sennnari (school) | -2.17E-06                           | -1.805E-07         |
|                    | 35. Senri            | -1.63E-06                           | -3.055E-06         |
| Five smallest      | 15. Naruo branch     | -2.89E-06                           | 4.75E-06           |
|                    | 37. Ibaragi city office | -2.43E-06                      | -2.59E-06          |
|                    | 29. Osaka observation station | -6.27E-07                  | -2.48E-06          |
|                    | 42. Suta kita fire station | -2.10E-06                 |                  |
| Daily minimum temperature |                     |                                 |                  |
| Five largest      | 1. Kobe             | 1.33E-06                            | 3.74E-06           |
|                    | 43. Alpinky school   | 9.900                               | 4.63E-06           |
|                    | 27. Narita (school)  | -1.75E-06                           | -2.59E-06          |
|                    | 29. Osaka observation station | -6.27E-07                  | -2.48E-06          |
| Five smallest      | 13. Takarazuka Fureai | -2.89E-06                          | -4.75E-06          |
|                    | 27. Narita (school)  | -1.43E-06                           | -2.59E-06          |
|                    | 37. Ibaragi city office | -2.43E-06                      | -2.48E-06          |
|                    | 42. Suta kita fire station | -2.10E-06                 |                  |

Therefore, the author used mean values of temperature in each period of testing data in order to give “t”.

In Table 4, models 2 and 3 exhibit better performances than model 1 in terms of both correlation and RMSE. The RMSE of model 2 is improved by approximately 0.2 compared with model 1 in all cases, and that for model 3 is further improved compared with model 2. This also shows that they also work as predictors on new unknown cases. Unless considering random effects, the predictors will return temperatures with errors of 1.1–1.3 °C. However, the error can be reduced to 0.8–0.9 °C if random effects are considered using the Bayesian framework.

The fluctuations of each regression coefficient by observation point were verified in model 3. To perform the verification, the mean and standard deviation of each coefficient at each point were calculated. Then, the standard deviation was divided by the mean, which is called a coefficient of variation (Table 5). If the absolute value of this value is large, it means that the fluctuations are comparatively large compared with the size of the coefficient itself. It can be seen that the distance from the central line of the river, distance to the coastline, designated commercial land use zone, mesh population, and mesh road density all show large numbers, whereas climatic factors such as the daily wind velocity, daily number of sunshine hours, and daily precipitation
show numbers of less than 1. The fluctuations of the climatic factors are smaller than the geographic factors. Moreover, the value of the NDVI is also comparatively small. These findings indicate that the influences on temperature of climatic phenomena such as sunshine, wind velocity, and the amount of land with vegetation are comparatively small and have little geographic variation. This also implies that geographic factors such as the distance from the river or distance from the coastline are affected by the surrounding land use or the distance from the water surfaces. In other words, these coefficients vary according to the location or surrounding situation.

To consider these findings in more detail, the distance from the coastline, for which this value is comparatively large, and the designated commercial land use zone and mesh road density were investigated. The regression coefficients of these factors at all observation points were derived, and the five largest and five smallest points are listed for the three temperature indices (Table 6.).

The spatial distributions of the listed points, based on their distance from the coastline, were drawn on a map for daily maximum temperature and daily minimum temperature (Fig.2.). It can be seen that most of the five largest points for daily maximum temperature are those near to the sea. The distribution is more varied for cases with the five smallest points. This suggests that the cooling effect of sea breezes appears more evident near the sea.

An obvious tendency cannot be determined in the case of the five smallest points for daily minimum temperature. However, there is a clear tendency for the five largest points to be located in places near the edges of mountainous areas such as Nishiwaki, Kumatori, and Toyono town office, where there is plenty of surrounding green land. It is thought that the overnight minimum temperature at such locations further reduces because of cooler air flowing directly from the mountains.

With regard to the designated commercial land use zone, the tendency is that both the five largest and five smallest points are relatively common for the daily maximum, daily minimum, and daily mean temperatures. For example, the five largest points for daily maximum temperature, such as Senri, and the Neyagawa city office and Toyonaka city offices, tend to be located in central urban areas where there are many buildings, high commercial activity, and considerable traffic. Therefore, it is thought that the quantities of thermal storage in the concrete and artificial thermal emissions are large, which exacerbate the rise of temperature. The sites of Sennari (school), Kobe, Naruo branch office, Nishinomiya city office, and Osaka observation station are all listed as the five smallest points. A scrutiny of their surroundings reveals a large river near Sennari, which could function as a path for wind, and large tracts of green land such as temples and parks in the neighborhoods of the Naruo branch office, Nishinomiya city office, and Osaka observation station. These are obviously different from the locations of the five smallest points for the daily maximum temperature. The designated commercial land use zone in this study refers only to legal land use regulation, which also implies that the actual situation such as building usage, building density, and building height differ between observation points. Thus, this might be one of the sources of the geographic differences.

For road density, it is characteristic that the five largest points for daily minimum temperature are those surrounded by forests and farmland, e.g. Nishiwaki and the Toyono town office. This raises the possibility that trunk roads could play a role as some form of heat source, discharging heat into the surrounding area. However, the sites of the Kishiwada city office and Narita (school) have small values for daily maximum temperature, daily minimum temperature, and daily mean temperature. Analysis of the maps reveals that these areas have comparatively large tracts of green land or large rivers in their neighborhoods. Moreover, another common feature is the comparatively high density of surrounding residential areas. Thus, there is a possibility that roads function as pathways for wind, allowing cool winds to flow from the green land or rivers.

---

Fig.2. Spatial Distributions of the Five Largest Points and Five Smallest Points of the Regression Coefficients of the Distance from the Coastline, Overlain on the NDVI Map. The Left-Hand Figure Shows the Values for Daily Maximum Temperature and the Right-Hand Figure Shows the Values for Daily Minimum Temperature. On the Map, Gray Shading Indicates NDVI, which Corresponds to the Amount of Land with Vegetation. The Lines Denote the Rivers.
Rocks might sometimes work to elevate temperatures but sometimes play a role as conduits for cooling winds that reduce temperatures. Overall, it depends on the surrounding geographic situation, and it is supposed that in most cases both effects counteract each other.

5. Conclusions

In this study, summer temperature development was modeled and the regression coefficients were estimated using the hierarchical Bayesian model. This proposed model considers the random effects of the regression coefficients, and the effectiveness of the model was quantitatively assessed. Moreover, the geographic factors that influence the randomness of the regression coefficients were discussed. It is concluded that summer temperature can possibly be modeled with high accuracy by building a regression model based on the hierarchical Bayesian method. By considering the random nature of the regression coefficients as well as the time-relevant terms, the accuracy was further improved and the correlation coefficient was improved to approximately 0.94. The differences in the regression coefficients between observation points for climatic factors such as the daily wind velocity and daily precipitation were comparatively small. However, the differences for geographical factors such as the distance from the central line of the river, distance from the coastline, designated commercial land use zone, mesh population, and mesh road density tended to be larger. Verification of these factors revealed three typologies: 1) Factors for which the effects vary depending on the surrounding geographic situation or land use, such as the mesh road density; 2) Factors that do not always express "actual" conditions, such as legal land use regulations; 3) Factors that are themselves influenced by location, such as the distance from the coastline. Based on this, it is concluded that in some cases, it is better to realize that the influence of land use varies depending on the location or surrounding situation, and that it is necessary for land use planning to ascertain this point.

Future research should model temperature development in winter and estimate the influence on the results of changing the distribution function of the prior probability of the regression coefficients.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 26420631. The author would like to thank the reviewers.

References

1) Albert, J. (2009) Bayesian computation with R. Berlin: Springer. (Ishida, M. and Ishida, K. (translators) (2010) Japanese translated edition. Tokyo: Maruzen)

2) Ando, M., Nishida D., Murakawa, S., Kindaichi, S., Ishida M., and Yasugi, K. (2014) Analysis of affecting factors on the energy consumption in detached houses by the multi-level model. Study on the energy consumption of electrified housing in Hiroshima area. Journal of Environmental Engineering (Transactions of AJI), 698, pp.383-392. (Written in Japanese)

3) Clark, J. S. and Gelfand, A. E. (eds.) (2006) Hierarchical modeling for the environmental sciences. New York: Oxford University Press Inc.

4) Congdon, P. (2003) Applied Bayesian modeling. Chichester: WILEY.

5) Congdon, P. (2006) Bayesian statistical modeling. Chichester: WILEY.

6) Furutani, T. (2004) Bayesian geographically weighted regression model and its application for land price model estimation. Papers on city planning, 39(6), pp.787-792. (Written in Japanese)

7) Honjo, T., Kohira N., Ichimura, T., and Maruta, Y. (2000) Effectiveness of neural network for the analysis of urban green thermal environment. Journal of the Japanese Institute of Landscape Architecture, 63(5), pp.547-550. (Written in Japanese)

8) Irie T. (2003) Study on effect of open space in reducing heat island by presuming the temperature. Journal of the Japanese Institute of Landscape Architecture, 66(5), pp.889-892. (Written in Japanese)

9) Iagawa S., Ichinose, T., Katagiri, Y., Osawa, S., and Ishikawa M. (2012) The relationship between green coverage distribution and inhabitation of Orthoptera on the reclaimed land in Tokyo Bay Area. Journal of the Japanese Institute of Landscape Architecture, 75 (5), pp.621-624. (Written in Japanese)

10) Jeong, Y., Lee, G., and Kim, S. (2015) Analysis of the relation of local temperature to the natural environment, land use and land coverage of neighborhoods. Journal of Asian Architecture and Building Engineering, 14 (1), pp.33-40.

11) Kobayashi, Y. (2013) Evaluation of external economic effect of vegetation considering seasonal variation. Journal of the Planning Institute of Japan, 48(3), pp.1023-1028. (Written in Japanese)

12) Kotani, K., Maruta Y., and Yanai, S. (1996) An investigation on the distribution of air temperature and the effect of open space to mitigate the heat island phenomenon in the Tokyo city. Papers on city planning, 31, pp.85-90. (Written in Japanese)

13) Li, H., Gao, W., and Ojima, T. (2002) Numerical analysis about urban climate change by urbanization in Shanghai. Journal of Asian Architecture and Building Engineering, 1 (2), pp.143-148.

14) Onishi, A., Okuoka, K., Shi, F., and Morisugi, M. (2010) Understanding of seasonal and spatial characteristics of heat mitigation effect of vegetation by applying geographically weighted regression model. Reports of the City planning Institute of Japan, 9, pp.93-97. (Written in Japanese)

15) Paez, A., Uchida T., and Morisugi, M. (2008) Urbanization and the urban heat island effect from a spatial descriptive approach. Papers on city planning, 33, pp.67-72.

16) R Development Core Team (2011) R: A language and environment for statistical computing. http://www.R-project.org/: R Foundation for Statistical Computing.

17) Shado, H., Sugiyama, J., Izumi, H., and Oka T. (1996) A study on the air temperature distribution influencing by various land uses. Journal of Architecture and Planning (Transactions of AJI), 479, pp.49-56. (Written in Japanese)

18) Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003) WinBUGS User Manual Version 1.4. http://www.mrc-bsu.cam.ac.uk/software/bugs/: The BUGS Project.

19) Tanimoto, J. and Ishino, H. (1996) A study on the distributional characteristic of outdoor air temperature for the design of air conditioning in the Tokyo metropolitan using detailed numerical geographical information. Journal of Architecture and Planning (Transactions of AJI), 483, pp.27-32. (Written in Japanese)

Notes

The term Deviance refers to the likelihood of each model multiplied by −2. Each data source of temperature is as follows: http://www.jma.go.jp/jp/amedas/. This procedure is based on the empirical Bayesian method. Model 3 becomes unstable if used without contrivances because the number of explanatory variables is comparatively large compared with the number of samples. It is necessary to provide the prior distribution with more meaningful information to control excessive fluctuations. However, it is not objective to provide arbitrary information to the prior distribution without rational reasoning. This is why the concept of the empirical Bayesian method was used in this study.