

# Network optimization of urban heat island measurements -Effect of reduction of observation points-

Tsuyoshi Honjo<sup>1</sup>, Kiyoshi Umeki<sup>1</sup>, Hiroaki Yamato<sup>2</sup>, Takehiko Mikami<sup>3</sup>, C.S.B. Grimmond<sup>4</sup>

<sup>1</sup> Department of Environmental Science and Landscape Architecture, Faculty of Horticulture, Chiba University, Matsudo, Chiba, 271-8510, Japan, honjo@faculty.chiba-u.jp

<sup>2</sup> Chuo University, 1-3-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551, Japan

<sup>3</sup> Faculty of Liberal Arts, Teikyo University, Hachioji, Tokyo, 192-0395, Japan

<sup>4</sup> Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, UK

dated : 15 June 2015

## 1. Introduction

Many cities have sensor network for heat island measurement and it is possible to observe the spatial pattern of meteorological data in detail.

In the design and operation of a meteorological measurement network for spatial characterization of the urban heat island, there are some questions. In the planning stage, the optimal number of measurement points and their optimal distribution are important. In the measurement stage, questions include whether sensors should be substituted or not, if instruments are out of order. In the design and operation of a meteorological measurement network, a larger number of measurement points are better. But considering the labor and cost, a smaller number of points would be better if the same result can be obtained.

We analyze an existing meteorological measurement network around the Tokyo metropolitan area (Extended METROS) to consider the effect of reducing observation points according to the methods of Honjo et al. (2015). Long term analysis of the effect of reducing observation points is done by using data of 18 months.

## 2. Material and Methods

The measurement points of Extended METROS are shown in Fig.1. The area analyzed extends from 35.3° to 36.3°N latitude (about 90.4 km) and from 139.2° to 140.2°E longitude (about 111.3 km) and includes Tokyo Bay. Thermometers (T&D, TR5106 thermistor sensor with RTR-52a data logger) are housed in un aspirated Stevenson screens at 1.5 m above ground level sited in elementary schools. The temperature measurements, collected since 2006, consist of samples taken from approximately 200 sites. The actual number at any given time varies because of trouble with data collection, broken or missing measurement instruments, reallocation of measurement points, etc. The data obtained from May 2007 to October 2008 (18 months, every hour) were used for the analysis.

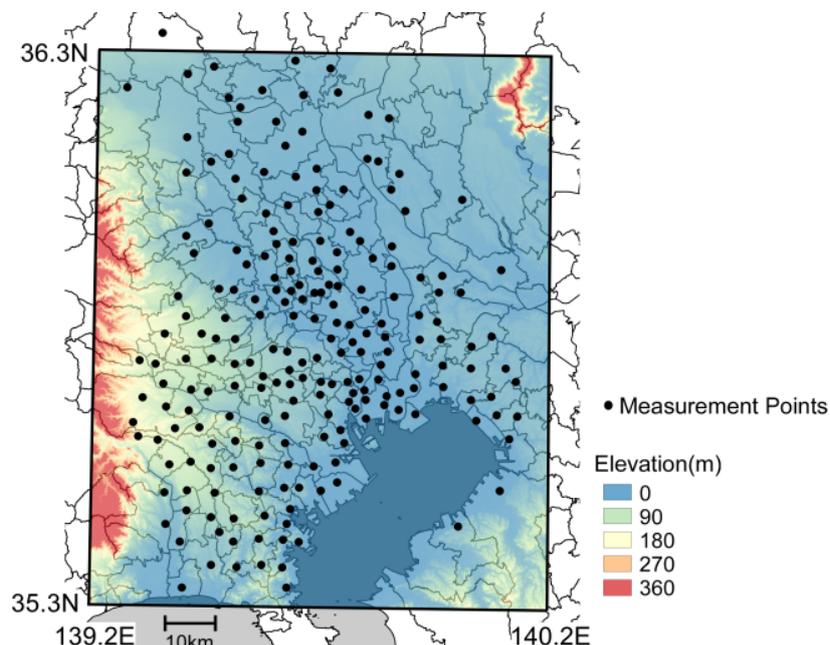


Fig. 1 The location of measuring points of the Extended METROS network in the Tokyo metropolitan area. Lines within the land indicates borders of local governments.

Interpolation is necessary to obtain a contour map or 2-D image of the distribution. In this study, a simple interpolation method, inverse distance weighting (IDW) algorithm was used.

With IDW the interpolated temperature  $T(x)$  was calculated from sampled data  $T_k$  by:

$$T(x) = \frac{\sum_{k=1}^m w_k(x) T_k}{\sum_{k=1}^m w_k(x)}$$

where the weighting function

$$W_k(x) = \frac{1}{d(x, x_k)^p}$$

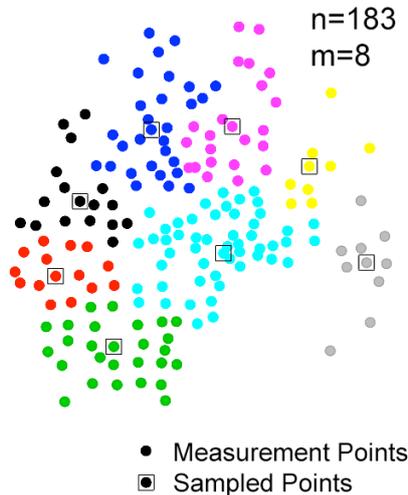
is related to  $x$ , the coordinate vector of an interpolated point,  $x_k$  is a measurement point,  $d(x, x_k)$  denotes the distance from  $x$  to  $x_k$ ,  $m$  is the number of sampled points, and  $p=2$  is used. Using the IDW method, spatial domains with  $201 \times 201 = 40401$  interpolated points were generated for analysis. These can also be visualized. To estimate the similarity between two spatial domains or interpolated images, normalized cross-correlation,  $R_{ncc}$  (hereafter the correlation) and root-mean-square error ( $RMSE$ ) were used:

$$R_{ncc} = \frac{\sum_{ix=1}^N \sum_{iy=1}^N T_1(ix, iy) T_2(ix, iy)}{\sqrt{\sum_{ix=1}^N \sum_{iy=1}^N T_1(ix, iy)^2 \sum_{ix=1}^N \sum_{iy=1}^N T_2(ix, iy)^2}}$$

$$RMSE = \sqrt{\frac{\sum_{ix=1}^N \sum_{iy=1}^N (T_1(ix, iy) - T_2(ix, iy))^2}{N^2}}$$

Where  $T_1(ix, iy)$  and  $T_2(ix, iy)$  are the interpolated values of deviation from mean air temperature of the original and sampled data, respectively;  $ix$  and  $iy$  are coordinates of the interpolated image; and  $N$  is the number of pixels in the  $x$ - $y$  dimension.

Selection of measurement points is undertaken using hierarchical clustering. Measurement points are classified into  $m$  categories. After the clustering and calculation of the center position of each category, the nearest point to the center position of each category was chosen as the sampled point and the temperature of that point was used as the representative value of the category (Fig. 2). The longitude and latitude of the measurement points were used for the classification by the clustering.



*Fig. 2 Example of choosing  $m=8$  points from  $n=183$  measurement points by sampling with hierarchical clustering. All points are classified into  $m$  categories expressed as different colors and center points of each category are selected as representative sampled points.*

The effect of reducing observation points from an existing meteorological measurement network is considered, using sampling with clustering. 10% to 90% of the data were reduced and sampled from the original data in 10% interval. (Corresponding sampling ratios were 0.9 to 0.1.) By using the sampled point with clustering, the interpolated data were obtained with IDW. The correlation and RMSE were also calculated from the interpolated data. Both are used as an index of the similarity of the images.

### 3. Results and Discussion

#### 3.1 Sampling with clustering and interpolation with IDW method

The measurement points, where all the data of observation period (May 2007 to October 2008) existed, was only 100 points out of 200 points. The numbers of points where no missing data existed in each month is shown in Table 1. In most of the months, the number of the points was more than 150. These monthly data of the no missing points were used for the further analysis.

Table 1 Number of measurement points where no missing data exist in each month.

month/year	no. of points	month/year	no. of points
05/2007	193	01/2008	178
06/2007	178	02/2008	190
07/2007	165	03/2008	194
08/2007	180	04/2008	179
09/2007	176	05/2008	165
10/2007	145	06/2008	177
11/2007	180	07/2008	174
12/2007	176	08/2008	163
		09/2008	155
		10/2008	129

The sampling with clustering and interpolation with IDW method were done hourly for the 10-90% cases and the original data during the observation period. Examples of relatively (a) good correlation and (b) bad correlation are shown in Fig. 3. The images are made from the original data (178 points), 50% (89 points) and 20% (36 points) of the data. Empirically, the images with the correlation which is more than 0.9 seem very similar. Calculation are done for 10 cases each hour (1 original and 9 reduction cases) and interpolated images (201 x 201 pixels) like Fig. 3 are made. The correlation and RMSE between sampled data and original data are calculated from the interpolated data.

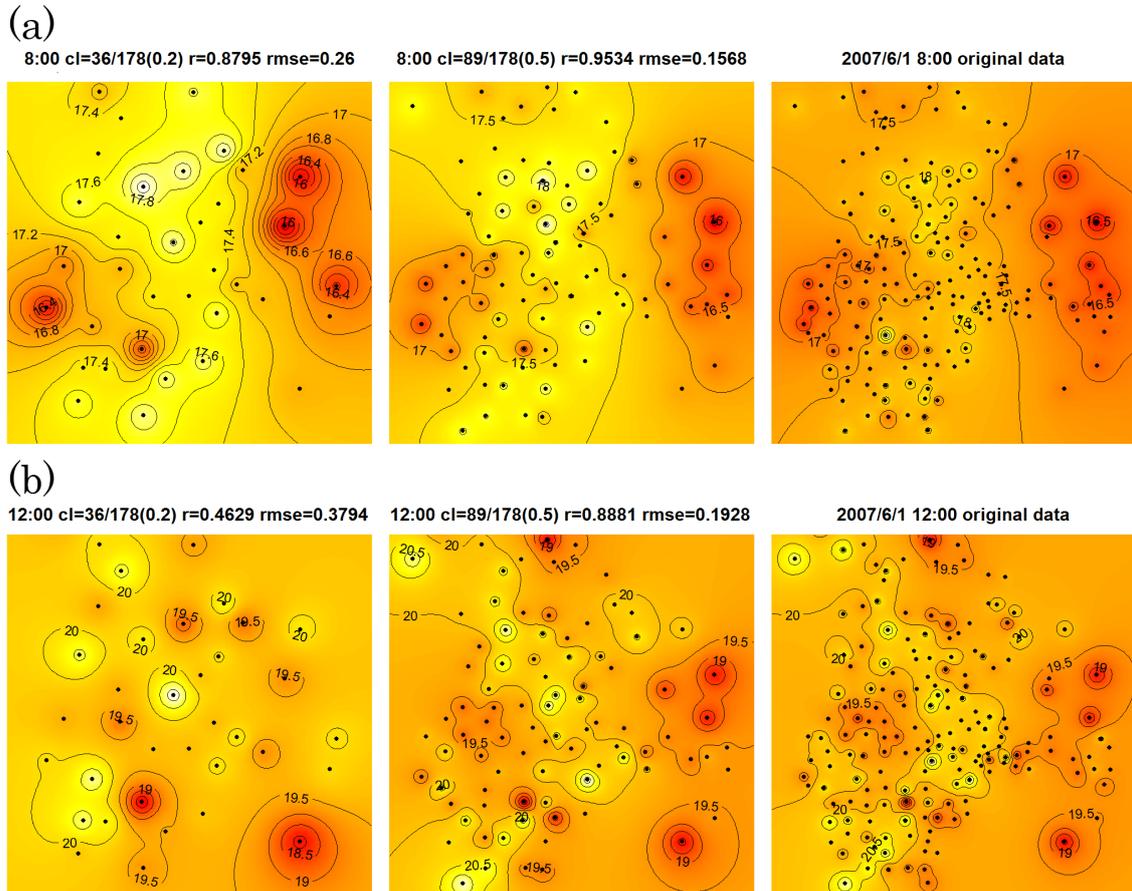


Fig. 3 examples of interpolated temperature distributions. Relatively (a) good correlation cases and (b) bad correlation cases. In the figure, r means the correlation with the image of the original data and the rmse means the RMSE in degree C. The area is same as the area of Fig. 1.

### 3.2 Monthly data analysis

The results of variation along the whole hourly time in August, 2007 are shown in Fig. 4. The cases of low sampling ratio (i.e., 0.1 and 0.2) shows low value of the correlation and high value of the RMSE. Fluctuations of the value in these cases are high. In the correlation of low sampling ratio, one low value peak in a day is clearly observed and these peaks also correspond the high value peak of the RMSE. With the sampling ratio of more than 0.7 (30% reduction), the correlations are mostly more than 0.9 and the RMSE are less than 0.2. The patterns of the variations of other months are similar to those of Fig. 4.

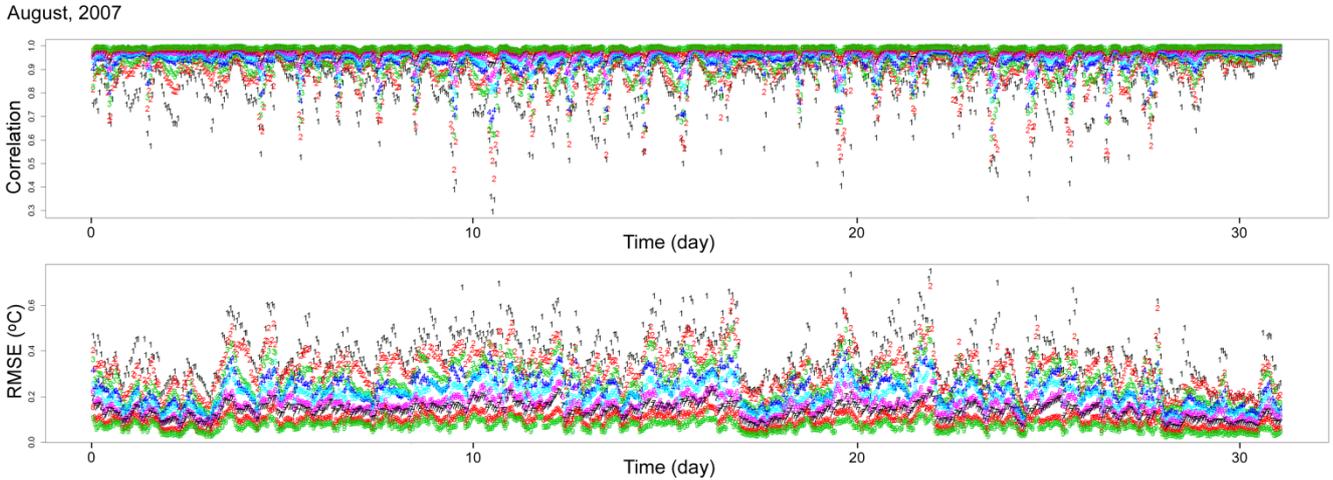


Fig. 4 Hourly correlation and RMSE in August, 2007. The number on the points in the figures corresponds the sampling ratio (1 is 0.1, 2 is 0.2 and 9 is 0.9).

The data were averaged for each 24 hours of the day and examples of two months (August, 2007 and February, 2008) are shown in Fig. 5. About the correlation, low value can be seen at 10 am in August and at around noon in February in the cases of low sampling ratio. In the cases of high sampling ratio, both the correlation and RMSE are nearly constant throughout the day.

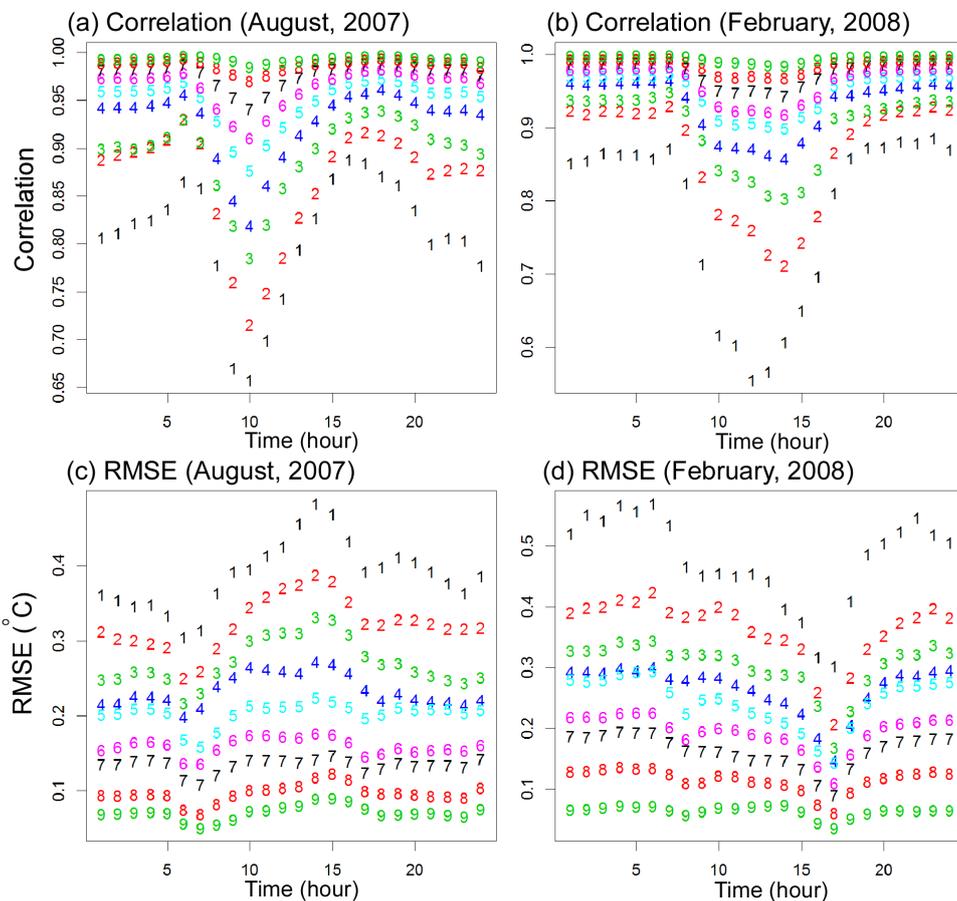


Fig. 5 Averaged correlations and RMSE of each hour of day in August, 2007 and February, 2008.

### 3.3 Characteristics of whole period

The results of whole months are shown in Fig. 6. Because number of points are different among months, sampled number of points are used for the x-axis. It is obvious that the correlation become low and RMSE high according to the reduction of points. Ranges of RMSE are broader than those of the correlation. High correlation (more than 0.95) in the average can be seen about the correlation even though reduction to half number of points.

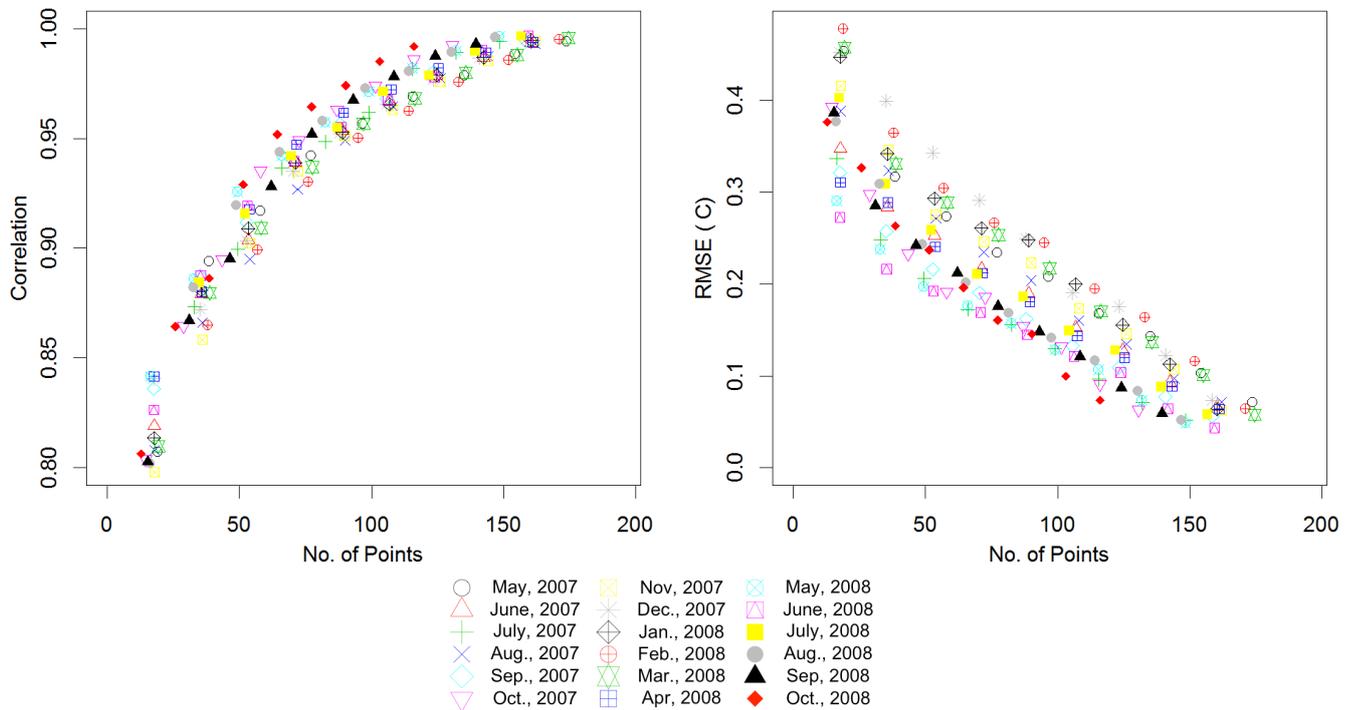


Fig. 6 Correlation and RMSE of each month in relation with the number of points for all the period. Marks of the points represents the months.

### 4. Conclusion

The methods presented here clearly show the effect of reducing observation points. It is difficult to precisely define the optimal number of points but we can find the allowable points considering the limit of the correlation and RMSE. The methods presented in this study can be applied in other meteorological measurement networks in evaluating the reduction of existing points of the networks.

### References

Tsuyoshi Honjo, Hiroaki Yamato, Takehiko Mikami, C.S.B. Grimmond, 2015: Network optimization for enhanced resilience of urban heat island measurements, Sustainable Cities and Society, in press. doi:10.1016/j.scs.2015.02.004