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

Tsuyoshi Honjo¹, Kiyoshi Umeki¹, Hiroaki Yamato², Takehiko Mikami³, C.S.B. Grimmond⁴
1 Department of Environmental Science and Landscape Architecture, Chiba University
2 Chuo University
3 Tokyo Metropolitan University
4 Department of Meteorology, University of Reading
Network for heat island measurement

- 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.
- What is the optimal number of measurement points and what is their optimal distribution?
- Whether sensors should be substituted or not, if instruments are out of order.
Objective of the study

- The effect of reducing observation points
- We analyze an existing meteorological measurement network around the Tokyo metropolitan area (Extended METROS).

Measurement points of Extended METROS.

Methods

1. Data of measurement Points

2. Selection of points (10% - 90%) by clustering.

3. Interpolation by IDW (Inversed Distance Weighing)

4. Similarity between the original data and the interpolated data
Data of measurement Points

- The data obtained from May 2007 to October 2008 (18 months, every hour) were used

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

<table>
<thead>
<tr>
<th>month/year</th>
<th>no. of points</th>
<th>month/year</th>
<th>no. of points</th>
</tr>
</thead>
<tbody>
<tr>
<td>05/2007</td>
<td>193</td>
<td>01/2008</td>
<td>178</td>
</tr>
<tr>
<td>06/2007</td>
<td>178</td>
<td>02/2008</td>
<td>190</td>
</tr>
<tr>
<td>07/2007</td>
<td>165</td>
<td>03/2008</td>
<td>194</td>
</tr>
<tr>
<td>08/2007</td>
<td>180</td>
<td>04/2008</td>
<td>179</td>
</tr>
<tr>
<td>09/2007</td>
<td>176</td>
<td>05/2008</td>
<td>165</td>
</tr>
<tr>
<td>10/2007</td>
<td>145</td>
<td>06/2008</td>
<td>177</td>
</tr>
<tr>
<td>11/2007</td>
<td>180</td>
<td>07/2008</td>
<td>174</td>
</tr>
<tr>
<td>12/2007</td>
<td>176</td>
<td>08/2008</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td></td>
<td>09/2008</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10/2008</td>
<td>129</td>
</tr>
</tbody>
</table>
Selection of points by clustering

points = 180/180 (100%)

points = 126/180 (70%)

points = 90/180 (50%)

points = 54/180 (30%)

points = 36/180 (20%)

points = 18/180 (10%)
Selection of points (10% - 90%) by clustering

Example of choosing 8 points from 183 by sampling with hierarchical clustering of coordinates. First, 183 points are classified into 8 categories expressed as different colors. Second, center points of each category are selected.
Interpolation by IDW (Inversed Distance Weighing)

Interpolated temperature $T(x)$ was calculated from measured data $T_k$.

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

weighting function:

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

$x$ : the coordinate vector of an interpolated point,

$x_k$ : measurement point

$d(x,x_k)$ : the distance from $x$ to $x_k$,

$m$ : the number of sampled points

$P$ : parameter of IDW ($p=2$ is used.)
Similarity between the original data and the interpolated data

To estimate the similarity between two interpolated temperature distribution $T_1$ and $T_2$, normalized cross-correlation, $R$ (the correlation) and root-mean-square error ($RMSE$) were used:

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

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

$T_1(ix,iy)$ : the interpolated values of air temperature from the original data
$T_2(ix,iy)$ : the interpolated values of air temperature from sampled data
$ix$ and $iy$ : coordinates of the interpolated image
$N$ : the number of pixels in the $x$-$y$ dimension.
Correlation and RMSE

low correlation case

2007/8/16 8:00 original data

s=0.2 R=0.6091 RMSE=0.422
Correlation and RMSE

low correlation case

2007/8/16 8:00 original data
Correlation and RMSE
high correlation case

2007/8/16 16:00 original data

s=0.2 R=0.9731 RMSE=0.4345
Correlation and RMSE

high correlation case

2007/8/16 16:00 original data
Correlation and RMSE on 16th, 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).
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 number on the points in the figures corresponds to the sampling ratio (1 is 0.1, 2 is 0.2, and 9 is 0.9).
Averaged correlations and RMSE of each hour of day in August, 2007 and February, 2008. 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).
Correlation and RMSE of each month in relation with the number of points for all the period. Marks of the points represents the months.
Conculsion

• The methods presented here clearly show the effect of reducing observation points.
• we can find the allowable points considering the limit of the correlation and RMSE.
• The methods in this study can be applied in other networks.
Thank you for your attention.
The location of measuring points of the Extended METROS network in the Tokyo metropolitan area.
Methods

Data
The data obtained from May 2007 to October 2008 (18 months, every hour) were used.

Measurement Points
Monthly data of the no missing points were used for the further analysis. 10% - 90% points are selected by using the clustering.

Interpolation
Selected data were interpolated by IDW
Grid data of 201 x 201 points were made.

Similarity between the original data and the interpolated data
The correlation and Root-Mean-Square Error (RMSE) were used. Similar images shows relatively high correlation and low RMSE.