Contribution of urbanization to warming in China

Ying Sun\textsuperscript{1,2}, Xuebin Zhang\textsuperscript{3\ast}, Guoyu Ren\textsuperscript{1,2}, Francis W. Zwiers\textsuperscript{4} and Ting Hu\textsuperscript{1}

\textsuperscript{1}National Climate Center, Laboratory for Climate Studies, China Meteorological Administration, Beijing 100081, China. \textsuperscript{2}Joint Center for Global Change Studies, Beijing 100875, China. \textsuperscript{3}Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario M3H 5T4, Canada. \textsuperscript{4}Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia V8W 2Y2, Canada. \textsuperscript{\ast}e-mail: Xuebin.Zhang@Canada.ca
1. Temperature data from observations

Homogenized monthly mean temperatures at 2419 Chinese stations (Supplementary Fig. 1, blue and yellow dots) for the period 1951-2013 are provided by the China’s National Meteorological Information Center (NMIC) to this study. The raw temperature data were homogenized using a method described in Xu et al.\(^1\) which is based on Wang et al.\(^2-3\) and implemented in RHtest\(^4\). Data prior to 1961 were not used due to limited availability (Supplementary Fig. 2) especially in western China. Station data for a particular month is treated as missing if more than 7 daily values are missing in that month. Data for a year are treated as missing if any monthly value in that year is missing. A station is retained for the subsequent analysis only if there are at least 20 years non-missing data during the 1961-1990 base period for the station. For any given year, more than 2000 stations are retained. Monthly temperature anomalies relative to the 1961-1990 climatology are computed and then averaged to obtain annual mean values for individual stations. The resulting gridded values are subsequently used to compute area-weighted regional mean temperature anomalies.

Because of differences in socio-economic development that influences the urbanization effect and aerosol emission between the west and the east of China\(^5\), as well as differences in climate, we divide China into western and eastern regions along the 105°E line. Regional averages for Eastern and Western China are computed by averaging all available gridded values within each region (Supplementary Fig.3). Eastern China (EC) has an East Asian monsoon climate. It is also economically more
developed with rapid urbanization starting in the late 1970s, as well as higher aerosol emissions. Western China (WC) has subtropical/temperate arid and semi-arid climates, with slower and later economic development and thus a smaller and later urbanization effect in temperature records. The division of China into these two regions provides a useful contrast in signal patterns for urbanization and other anthropogenic forcing that helps to distinguish their effects in observations via the optimal fingerprinting technique.

2. Temperature data from climate model simulations

Monthly mean surface air temperatures simulated by climate models participating in the CMIP5 are used for the estimation of the climate response to external forcing and natural variability. In total, we use 108 historical simulations forced with historical anthropogenic and natural forcings (ALL) by 23 climate models that also provided RCP4.5 simulations. These ALL simulations end in 2005. The corresponding RCP4.5 simulations for years 2006-2012 are used to extend the ALL simulations, except in instances when the RCP4.5 ensemble is smaller than the ALL ensemble, in which case the RCP4.5 ensemble mean was used to extend the ALL simulations. We use 36 historical natural forcing (NAT) simulations by 8 models that end in 2012 to estimate the NAT responses. Additionally, we use 33 historical greenhouse gases (GHG) simulations by 7 models that end in 2012 to estimate the GHG response. Unforced preindustrial control simulations from 41 models, which were divided into a total of 346 52-year segments, are used for the estimation of natural variability. Supplementary Table 1 lists these simulations. Model simulations are of different
spatial resolutions, and they are transferred to $5^\circ \times 5^\circ$ grid used for the observations.

Temperature values for model coastal grid boxes are influenced by both land and ocean while our observational data are strictly from land stations. Because different models have different resolutions and ocean/land masking as well as difficulties in defining representative areas for observing stations, it is difficult to precisely match the model grid boxes with land observations. Rather than determining the best approach for ocean/land masking for the model data, we evaluate the influence on the average temperature over Eastern China due to these coastal grid boxes. We compare regional averages using model grid boxes that are 100% land and that include all available grid boxes within the region. The difference in the ensemble mean for 19 models whose land/ocean fraction information is embedded in the temperature file is very small. The largest difference in the annual mean temperature anomalies is about 0.04°C (Supplementary Fig. 4). As a result, we use all model grid boxes within the region irrespective of the size of ocean fraction of the grid box.

3. Signal pattern estimates

3.1 Signal patterns of climate response to large-scale external forcing

The CMIP5 forced simulations are used starting from 1961. It is not necessary to mask the model output to mimic observational availability because the observational data have complete coverage over mainland China on the $5^\circ \times 5^\circ$ grid. Multi-model ensemble means of ALL, NAT and GHG simulations are used as signals. This is done
by first computing ensemble mean for individual models and then averaging these ensemble means.

3.2 Urbanization signal pattern

Long-term trend differences in temperatures between nominally urban and rural stations have previously been used to represent the effect of urbanization\textsuperscript{7-9}. However, the effect is likely underestimated when determined in this way because truly rural stations are difficult to find and some rural stations may have gradually urbanized with time\textsuperscript{10}. Instead of using the trend in the difference between urban and rural stations as the estimate of complete urbanization effect, we use the difference to construct an urbanization effect signal pattern and then estimate its magnitude using the optimal fingerprinting approach.

Ren et al.\textsuperscript{10} selected a set of rural stations using a procedure that takes into account the station history, population, the size of built-up areas etc. The locations of these stations are marked as yellow dots in Supplementary Fig. 1. These stations are used to produce a gridded rural temperature anomaly data set using methods identical to those used for gridding the full collection of stations. Similarly, we also produced gridded urban temperature anomaly dataset by using data from non-rural stations. The gridded rural and urban temperature anomalies are spatially averaged to produce regional average time series of rural and urban temperature anomalies for both Eastern and Western China (Supplementary Fig.3). To reduce the influence of climate variability over space on the urban and rural temperature difference series, we only use the
values from the grid boxes in which there is at least one rural station and one urban station (there are 22 such grid boxes in each region). Figure 1 shows the temperature differences between the urban stations and the rural stations. The urbanization effect should be small in the 1960s and 1970s due to a lack of development. It should increase monotonically and approach an upper bound over time as the urbanization effect should plateau once established\textsuperscript{11}. We thus fit the temperature differences to sigmoid curves to represent signal patterns of urbanization influence. We use sigmoid curves that are defined by a three-parameter logistic function:

\[ f(t) = \frac{L}{1 + e^{-k(t-t_0)}} , \]

where \( t_0 \) is the \( t \)-value of the sigmoid’s midpoint, \( L \) is the curve’s maximum value, and \( k \) represents the steepness of the curve. We assume a zero average urbanization effect during the 1960s and 1970s when fitting the curve by removing the mean difference for 1961-1980 from the regional series prior to the fitting. As the observations and model simulated signals are centered to the 1961-1990 mean, we center the fitted curves to their respective 1961-1990 mean as the urbanization signal pattern to be used in the detection and attribution analysis.

Both the observation vector and urbanization signal pattern are computed from the same set of station data. It is thus important to discuss potential circularity. For two variables \((x, y)\), and two combinations of them \(a = x + y\), \(b = x - y\), the second pair \((a, b)\) are just as independent as the first pair \((x, y)\). If we consider in our case the variables \((x, y)\) to be urban and rural station data, heuristically, the variable \(a \) would
act like the regional means in the observational vector whereas $b$ would act like the differences going into the urbanization fingerprint. Therefore the urbanization effect calculated as the difference between all urban and rural stations would be quite independent from the observation vector used in the regression.

4. **Estimate of covariance**

Fitting and testing the regression models requires two independent estimates of the covariance structure of internal climate variability\textsuperscript{12}. They are constructed using inter-ensemble differences from forced simulations and unforced control simulations. The data include inter-ensemble differences of forced simulations starting from 1961 and inter-ensemble differences from 52-yr chunks of data starting from 1898 of each forced simulation. Data from control simulations are divided into 52-yr chunks and similarly masked to be missing when observational data are missing. Regional averages are computed from those observationally masked data. A regularized covariance estimator\textsuperscript{13} is used to provide more robust covariance matrix estimates and is used for model fitting. Both the large number of inter-ensemble control simulation chunks and the use of a regularized estimate procedure ensure that the covariance matrix estimates are of full rank. The regression model is therefore fitted without resorting to an EOF truncation to reduce the dimension of the detection space. A resampling-based residual consistency test\textsuperscript{13} is used to compare model simulated internal variability with observations.

5. **Robustness to the sampling error in the URB signal**
The URB signal should be small at the beginning and increase with time until reaching an upper bound once urban areas are mature\textsuperscript{11}. Constrained by the small nature of the signal, the sampling errors in the URB signal should also be small at the beginning. The errors would increase with time as URB signal increases and as rural stations become increasingly urbanized. This error structure would be different from the model simulated noise and thus cannot easily be accounted for in the total least square framework. In the detection and attribution analysis, we are effectively using the ordinary least fitting for the URB signal by artificially specifying a very large number of “model runs” for that signal. Here we assess the impact of URB signal uncertainty by using a bootstrap procedure to evaluate the robustness of URB detection results against the sampling errors in the URB signal pattern.

The bootstrap procedure involves the resampling of the difference series between the temperatures observed at urban and rural stations. There are 22 5×5º grid boxes in both Eastern China and Western China with data from both urban and rural stations. We construct a bootstrap sample of regional means of the difference series for each region by averaging the differences series from 22 grid boxes that are separately drawn, with replacement, from the 22 grid boxes in each region. These regional means are then fitted to logistic curves to construct bootstrap URB signal patterns. This procedure was repeated 1000 times. Supplementary Fig. 5 displays the time evolution of the 90% range of the signal patterns. It is clear that sampling errors in the URB signal patterns increase over time and can be quite large towards the end of the period. As there is a large sampling uncertainty in the urban and rural temperature
difference, the estimate of urban warming based on the urban and rural temperature series should also have large sampling uncertainty. These bootstrap URB signal patterns are then used in the 2-signal and the 4-signal detection and attribution analysis in place of the original URB signal pattern. The scaling factor of the URB signal in the 1000 2-signal analyses has a median value 1.67 (90% confidence interval: 0.86 to 2.48). The URB signal was separately detected 866 times and the ALL signal was always detected. The URB signal scaling factor in the 1000 4-signal analyses has a median value 1.40 (90% confidence interval: 0.58 to 3.96). The URB signal was separately detected 434 times and the GHG signal was always detected. However, if the observation vector is replaced with the GCM generated noise, the detection rates are less than 5% as expected. This clearly indicates that the detection results for the URB signal are robust against the sampling errors in the URB signal.

References


Supplementary Table 1 | **List of multi-model simulations used in this study.**

Numbers represent the ALL, NAT, GHG simulation ensemble sizes or the number of 52-year chunks for the CTL simulations. The symbol * indicates that these model runs were used for noise estimation only.

<table>
<thead>
<tr>
<th>Model</th>
<th>ALL (23)</th>
<th>CTL (41)</th>
<th>NAT (8)</th>
<th>GHG (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS1-0</td>
<td>23</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCESS1-3</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCC-CSM1-1</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCC-CSM1-1-m</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BNU-ESM</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CanESM2</td>
<td>5</td>
<td>16</td>
<td>5 (to 2012)</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>CCSM4</td>
<td>6</td>
<td>8</td>
<td>4*</td>
<td></td>
</tr>
<tr>
<td>CESM1-BGC</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>CESM1-CAM5</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CESM1-CAM5-1-FV2</td>
<td>4*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CESM1-FASTCHEM</td>
<td>3 *</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CESM1-WACCM</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CMCC-CESM</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>CMCC-CM</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>CMCC-CMS</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>CNRM-CM5</td>
<td>10</td>
<td>14</td>
<td>6 (to 2012)</td>
<td>6 (to 2012)</td>
</tr>
<tr>
<td>CSIRO-Mk3-6-0</td>
<td>9</td>
<td>8</td>
<td>5 (to 2012)</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>EC-EARTH</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>FGOALS-g2</td>
<td></td>
<td></td>
<td>11</td>
<td>3*</td>
</tr>
<tr>
<td>FIO-ESM</td>
<td>3</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL-CM2p1</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>GFDL-CM3</td>
<td>5</td>
<td>8</td>
<td>3*</td>
<td></td>
</tr>
<tr>
<td>GFDL-ESM2G</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>GISS-E2-H</td>
<td>6</td>
<td>13</td>
<td>5 (to 2012)</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>GISS-E2-H-CC</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>GISS-E2-R</td>
<td>6</td>
<td>14</td>
<td>5 (to 2012)</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>GISS-E2-R-CC</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>HadCM3</td>
<td>10*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadGEM2-CC</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>4</td>
<td></td>
<td>4 (to 2012)</td>
<td>4 (to 2012)</td>
</tr>
<tr>
<td>INMCM4</td>
<td></td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>6</td>
<td>16</td>
<td>3 (to 2012)</td>
<td>3 (to 2012)</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>3</td>
<td>5</td>
<td>3 (to 2012)</td>
<td></td>
</tr>
<tr>
<td>IPSL-CM5B-LR</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>MIROC-ESM</td>
<td>3</td>
<td>10</td>
<td>3*</td>
<td></td>
</tr>
<tr>
<td>MIROC-ESM-CHEM</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>MIROC4h</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>MIROC5</td>
<td>5</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>3</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>3</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-P</td>
<td></td>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>MRI-CGCM3</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NorESM1-M</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NorESM1-ME</td>
<td></td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>SUM(models)</td>
<td>108(23)</td>
<td>346(41)</td>
<td>36(8)</td>
<td>33(7)</td>
</tr>
<tr>
<td></td>
<td>125(26)*</td>
<td></td>
<td>49(12)*</td>
<td></td>
</tr>
</tbody>
</table>

© 2016 Macmillan Publishers Limited. All rights reserved.
### Supplementary Table 1

List of multi-model simulations used in this study. Numbers represent the ALL, NAT, GHG simulation ensemble sizes or the number of 52-year chunks for the CTL simulations. The symbol * indicates that these model runs were used for noise estimation only.

<table>
<thead>
<tr>
<th>Model</th>
<th>ALL</th>
<th>NAT</th>
<th>GHG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS1</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>ACCESS1-3</td>
<td>3</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>BCC</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>BNU-ESM</td>
<td>9</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>CanESM2</td>
<td>5</td>
<td>16</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>CCSM4</td>
<td>6</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>CESM1-BGC</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CESM1-CAM5</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>CESM1-CAM5-1-FV2</td>
<td>4*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CESM1-FASTCHEM</td>
<td>3*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CESM1-WACCM</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMCC-CESM</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMCC-CM</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMCC-CMS</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNRM-CM5</td>
<td>10</td>
<td>14</td>
<td>6 (to 2012)</td>
</tr>
<tr>
<td>CSIRO-Mk3-6-0</td>
<td>9</td>
<td>8</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>EC-EARTH</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGOALS-g2</td>
<td>11</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>FIO-ESM</td>
<td>3</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>GFDL-CM2p1</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL-CM3</td>
<td>5</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>GFDL-ESM2G</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GISS-E2-H</td>
<td>6</td>
<td>13</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>GISS-E2-H-CC</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GISS-E2-R</td>
<td>6</td>
<td>14</td>
<td>5 (to 2012)</td>
</tr>
<tr>
<td>GISS-E2-R-CC</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadCM3</td>
<td>10*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadGEM2-CC</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>4</td>
<td>4 (to 2012)</td>
<td>4 (to 2012)</td>
</tr>
<tr>
<td>INMCM4</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>6</td>
<td>16</td>
<td>3 (to 2012)</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>3</td>
<td>5</td>
<td>3 (to 2012)</td>
</tr>
<tr>
<td>IPSL-CM5B-LR</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIROC-ESM</td>
<td>3</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>MIROC-ESM-CHEM</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIROC4h</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIROC5</td>
<td>5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>3</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-P</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRI-CGCM3</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>NorESM1-M</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>NorESM1-ME</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUM(models)</td>
<td>108</td>
<td>125</td>
<td>346</td>
</tr>
</tbody>
</table>

### Supplementary Figure 1

Map of observing stations and location of 143 rural stations. Spatial distribution of the 2419 observing stations (blue and yellow dots) whose data are used in the analysis along with the 5×5 grid used for the estimation of regional averages. Selection of rural stations (yellow dots) is based on refs. 10. Elevation (m) is indicated by the colors, with low elevations in green and high elevations in brown.
Supplementary Figure 2 | Availability of observing data. a) the number of observing stations with sufficient data to compute annual mean temperature. b) the number of 5×5° grid boxes with which there is at least one station available to compute grid box mean. Black, blue, and red lines show numbers for the whole China, Eastern and Western China, respectively.
a) Eastern China

![Graph showing temperature anomalies for Eastern China]

b) Western China

![Graph showing temperature anomalies for Western China]

Supplementary Figure 3 | **Regional mean temperature anomalies.** Annual mean temperature anomalies (°C) relative to 1961-1990 average for a) Eastern China and b) Western China. Blue and red lines show time series computed from rural and non-rural stations. Black lines show the differences between rural and non-rural stations.
**Supplementary Figure 4 | Multi-model mean temperature changes under ALL forcing for Eastern China.** Multi-model ensemble mean of annual temperature anomalies relative to 1961-1990 averages computed with land only model grids (blue line) and all model grids (red line) within the region. Model data are regridded to the common 5×5º gridded values by averaging values from available native model grids whose centers are within the 5×5º box. Data from 19 models with available land/ocean fraction data were used in the calculation.
Supplementary Figure 4 | Multi-model mean temperature changes under ALL forcing for Eastern China. Multi-model ensemble mean of annual temperature anomalies relative to 1961-1990 averages computed with land only model grids (blue line) and all model grids (red line) within the region. Model data are regridded to the common 5×5º gridded values by averaging values from available native model grids whose centers are within the 5×5º box. Data from 19 models with available land/ocean fraction data were used in the calculation.

Supplementary Figure 5 | Sampling uncertainty in URB signal. The median values (red curves) and the 5th and 95th percentile ranges (grey areas) of the URB signal from the 1000 Bootstrap samples for a) Eastern China, and b) Western China, respectively.