Supplement of Atmos. Meas. Tech., 15, 735–756, 2022
https://doi.org/10.5194/amt-15-735-2022-supplement

Supplement of

A high-resolution monitoring approach of canopy urban heat island using a random forest model and multi-platform observations

Shihan Chen et al.

Correspondence to: Yuanjian Yang (yyj1985@nuist.edu.cn)

S1. Specific inversion steps of related environmental variables

S1.1 Extraction of IS

According to related research (Son et al., 2017), Landsat 8 OLI multispectral data will be used to calculate the Normalized Difference Composite Index (NDCI) to extract IS area. The Normalized Difference Built-up Index (NDBI), Soil Adjusted Vegetation Index (SAVI) and Normalized Difference Water Index (NDWI) were calculated. Then NDCI was calculated, and IS area was obtained through k-means clustering. See formula 1-4 for details:

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \tag{S1}$$

Here, SWIR and NIR are the radiation intensity values received in the shortwave infrared band 1 (band 6) and near infrared band (band 5) of Landsat 8 OLI respectively;

$$SAVI = \frac{(NIR - R)(1 + L)}{(NIR + R + L)} \tag{S2}$$

Among them, L is generally selected as 0.25, 0.5 or 1. When L is selected as 0.5, the IS extraction effect is the best.

$$NDWI = \frac{G - SWIR}{G + SWIR} \tag{S3}$$

G is the green band (band 3) of the OLI sensor. When the NDWI value is greater than 0.2, it is considered as a permanent water body and masked.

$$NDCI = \frac{(NDBI + NDWI)/2 - SAVI}{(NDBI + NDWI)/2 + SAVI} \tag{S4}$$

Since the resolution of the land satellite is 30 meters, an urban pixel is usually composed of IS, water bodies and vegetation. The mixed pixel was converted into a certain pure pixel by fuzzy clustering and threshold method. Then, the distribution of IS was finally obtained.

In the references, the accuracy of IS classification is 0.9. In this study, Taking IS in 2017 as an example, it is compared with the IS produced by Xin Huang et al., and the correct rate is 94%, indicating the relatively high accuracy.

S1.2 Remote sensing indices and surface albedo calculation

NDVI, gNDVI and SAVI can well reflect the vegetation coverage in the study area and is closely related to temperature changes. The specific calculations of NDVI and gNDVI are as follow:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{S5}$$

$$gNDVI = \frac{NIR - G}{NIR + G} \tag{S6}$$

Among them, R is the reflectance of OLI red band (band 4). The distribution of Nanjing's NDVI was show in Figure 2d.

NDMI is used to determine vegetation water content. It is calculated as a ratio between the NIR (band 5) and SWIR (band 6) values in traditional fashion:

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \tag{S7}$$

The NDBI uses the NIR and SWIR bands to emphasize manufactured built-up areas. It is ratio based to mitigate the effects of terrain illumination differences as well as atmospheric effects. The calculating method of SAVI and NDBI is shown in Text S1.1.

The air near the surface is generally heated by the long-wave radiation reflected from the surface, so surface albedo has a strong correlation with AT. Surface albedo is the definition of the ratio of reflected emission to the incidence of target object, that is, the ratio of total radiant energy emitted in each direction per unit time and unit area to the total incident radiant energy. This article uses the method of Liang. Landsat8 OLI data were chosen to invert the surface albedo, with the formula being as follow:

$$albedo = 0.356B + 0.130R + 0.373NIR + 0.085SWIR1 + 0.072SWIR2 - 0.0018 \quad (S8)$$

Among them, B is the reflectance of OLI band 2, and SWIR2 is the reflectance of OLI band 7.

In the reference, the quartile of calculated albedo is 5%. In this study, based on 21 July 2017 albedo, two areas of 1km×1km water body and vegetation pure pixels are selected to calculate the average, witch are 6.86% and 15.05% respectively (both in the normal range).

S1.3 Distance calculation from the city center calculation

The closer to the city center, the more significant the UHI effect is. Therefore, the distance from it has a significant impact on AT. There are many ways to define the city center. In this article, geometric center was chosen to be the city center. The central formula are as follows:

$$\bar{x} = \frac{\int xf(x)dx}{\int f(x)dx}, \bar{y} = \frac{\int yf(y)dy}{\int f(y)dy} \quad (S9)$$

Among them, $f(x)$ and $f(y)$ represent the weights of x and y respectively, , which are 1 in the calculation of the geometric center. $\bar{x}$ and $\bar{y}$ represent the calculated geometric center coordinates. Then, ignoring the curvature of the earth, the distance were calculated from each weather station to the city center:

$$dis = \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2} \quad (S10)$$

Here, x and y represent the horizontal and vertical coordinates of the weather station, and dis represents the distance from the city center to the weather station.

S2. Stepwise linear regression and geographically weighted regression

Stepwise linear regression is the step-by-step iterative construction of a linear regression model that involves the selection of independent variables to be used. In this study, backward selection method was used. its basic steps are: 1. Start with all explanatory variables in the model. 2. Remove the variable with the largest P-value, which means the least statistically significant. 3. The new (p-1)-variable model is established and the variable with the largest P-value is removed. 4. Continue until a stopping rule is reached. For instance, stop when all remaining variables have a significant P-value defined by some significance threshold.

A GWR model takes the following form:

$$Y_i = \beta_0(u_i, v_i) + \sum_{j}^{P} \beta(u_i, v_i)X_{ij} + \varepsilon_i \quad (S11)$$

where Y_i, X_{ij}, and ε_i are the dependent variable, the j explanatory variable (subscripted as j), and the random error at point i (subscripted as subscript i) respectively. The location is denoted by the coordinates (u_i, v_i) of a given point i. The coefficients $\beta_j(u_i, v_i)$are weights function on the location, and $\beta_0(u_i, v_i)$ is the geographical intercept.

S3. Table and caption

Table S1. band ranges and the main use of Landsat 8 OLI.

Sensor	NO.	Wavelength (μm)	Spatial resolution(m)	Usage
OLI	1	0.433–0.453	30	Coastal monitor
	2	0.450–0.515	30	Water penetration and discrimination of soil and vegetation
	3	0.525–0.600	30	Vegetation identification
	4	0.630–0.680	30	Observing roads, bare soil and vegetation
	5	0.845–0.885	30	Biomass estimation and wet soil identification
	6	1.560–1.660	30	Distinguish roads, bare soil, water and fog-cloud identification
	7	2.100–2.300	30	Rock and mineral identification
	8	0.500–0.680	15	Resolution enhancement
	9	1.360–1.390	30	Cloud detection

S4. Figures and captions

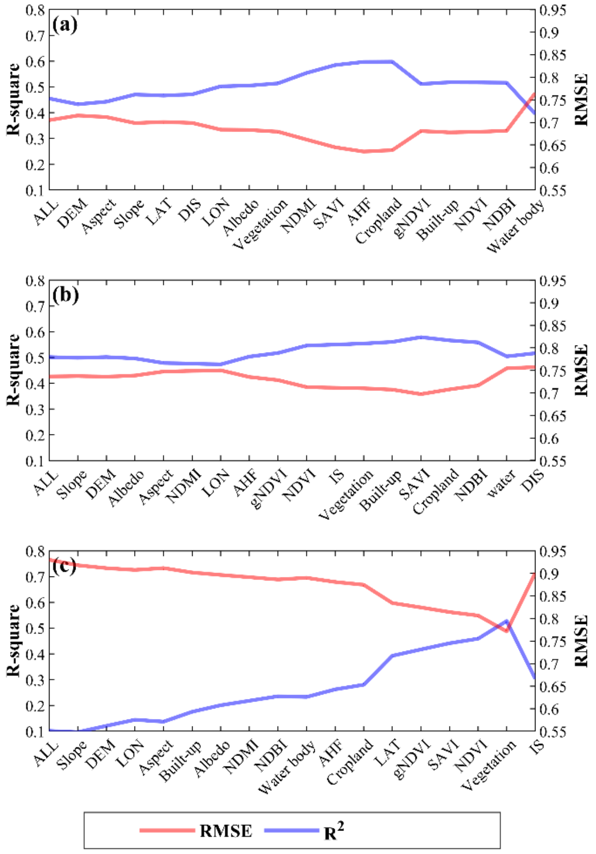

Figure S1. The performance of the RF models under different variable combinations: (a) 11 August 2013; (b) 2 September 2015; (c) 21 July 2017

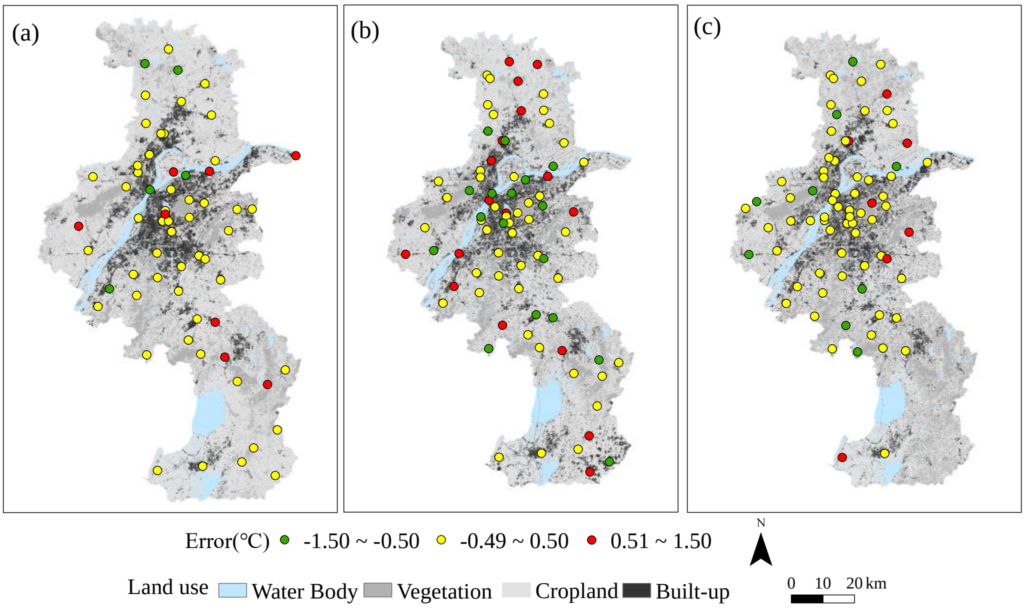

Figure S2. the predicted absolute error of the air temperature(AT) by random forest: (a) 11 August 2013; (b) 2 September 2015; (c) 21 July 2017.

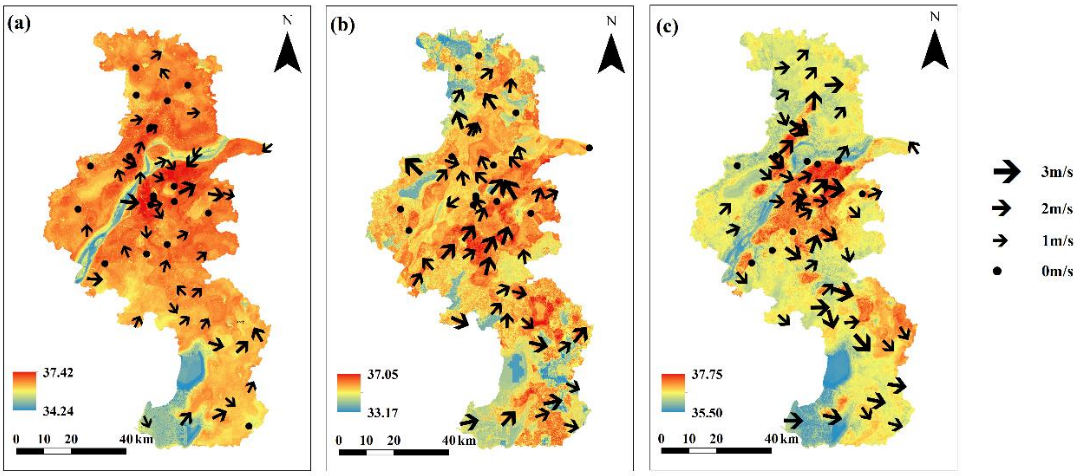

Figure S3. Spatial distribution of air temperature (AT) and wind vector field in Nanjing: (a) 12 August 2013; (b) 13 August 2013; (c) 14 August 2013.