**Code description** **on** **multivariate low-quality spatio-temporal data processing methods**

**1. Abstract of code**

We provided a set of scripts implemented in MATLAB R2013a for spatio-temporal serial data analysis procedures. The codes of all methods have been uploaded on our general pipeline available for reusing at the GitHub (<https://github.com/370982/parakeet.git>) in folder MatlabCode. The input data and output data of incomplete data imputation and denoising of time series were all XLS files, and the codes of NRA method and NLM algorithm were running in MATLAB, including comparison with other algorithms, the calculation of evaluation indices. Spatial population distribution data were clipped from the Jing-Jin-Ji Region and then converted into GeoTiff format before running the procedures in MATLAB software and needed to keep the output files with the same geo-referencing information. Extraction and conversion of the population data was done using ArcGIS 10.2. All the methods could been conducted as long as MATLAB software installing correctly.

**2. Missing value imputation method**

**2.1 NRA method**

When using NRA method to conduct multiple missing data imputation, complete population time series was introduced as an auxiliary dependent variable and to establish a nonlinear relationship. First, correspondence to the linear relationship between two variables was determined using a scatter plot. Then, the functional form for the nonlinear relationship was preliminarily determined to predicate a nonlinear function model for conversion to a linear regression model through mathematical transformation. Finally, a linear regression model based on MATLAB tools was fitted to determine the significance of the regression equation using an F-test. In addition, the relative error of the simulation experimental results was controlled at approximately 5%. Using the NRA method, the parameter of *‘RelaError’* is the relative error and *‘F’* is the F value of F-test.

**2.2 Improved NLM algorithm**

When using an improved NLM algorithm to conduct single missing data imputation, the single missing value must be input 0. Here, a complete population time series was introduced as a source for Gaussian kernel weight and weighted traversal of time series data. Parameter *‘kernel’* is normalized kernel weight matrix and parameter *‘f’* is the size of neighborhood window set to 5. The Gaussian kernel parameter *‘h’* is not a global variable and was determined by a specific experiment, and it directly determined the degree of filtering. Incomplete time series data have similar varying trends, strong correlations and the same lengths in population time series; thus, weight is obtained using the Gaussian kernel function provided by the population time series. Missing data imputation is provided by weighted traversal of all time series data.

**3. NLM algorithm for denoising**

We used the NLM algorithm to denoising time series and the NLM algorithm has two steps: (1) adding the noise; (2) removing the noise. Additive white Gaussian noise obeying a normal distribution is a basic noise model used in information theory. Consequently, we assumed that time series have no prior knowledge of a system, and then Gaussian white noise was added to original time series data, and the SNR was 20 dB. Due to the units of original data not being unified, they had to be normalized before adding the noise. Parameter *‘f’* is the size of neighborhood window set to 5. Weights for the NLM algorithm were determined using the Gaussian kernel function, and the Gaussian kernel parameter reached 0.15 has well experimental effects when using the NLM algorithm.

**4. Gaussian pyramid method for scaling up**

The Gaussian pyramid method was used to scale up the raster data with a 1 km resolution of population distribution. Local areas were selected as experimental areas for scale-up and the extraction of the population data was done using ArcGIS 10.2. By constantly Gaussian blurring of multi-scale on population data and down-sampling, different resolutions of population data are obtained. Parameter *‘f’* is the Gaussian template and we used a Gaussian template of 3\*3 in the study. To ensure the elements of the Gaussian template are between 0 and 1, the template must be normalized. Parameter *‘N’* is the coefficient of down-sampling and they were set to 2, 4 and 8. Accordingly, the population data were scaled up to 2 km, 4 km and 8 km.