ST-MVL: Filling Missing Values in Geo-sensory Time Series Data

The document describes sample datasets and codes we have used in paper [1].

The sample datasets consist of three parts:
a) "pm25_ground.txt" contains the PM2.5 readings from 36 air quality monitoring stations in Beijing from 2014/05/01 to 2015/04/30. It is used as a ground truth file, though it has some missing entries. Each column stands for a sensor and each row denotes the reading of those sensors at a time stamp. The readings in a row are separated by commas, with the first field denoting a time stamp. A blank field in a row denotes a missing value.
b) "pm25_missing.txt" is a missing file, which is generated by deliberately adding additional missing entries to pm25_ground.txt. Those missing entries are targets of our algorithm to fill. The schema of this dataset is the same as "pm25_ground.txt". please refer to [1] for the details on the method with which we added missing values to the ground truth file. Random adding missing entries violates the nature of data missing.
c) "pm25_latlng.txt" describes location information of air quality sensors. Each row stands for a sensor, and columns, separated by comma, are defined as: sensor_id, latitude, longitude.

The code is built in Visual Studio 2013, consisting of two major functions:
1. To replicate our experimental results, please run "verify" function with the aforementioned three pm25 datasets. We use 3rd, 6th, 9th and 12th months data as a test set and the rest for training. For more details, please refer to paper [1] and the code.
2. If you want to apply our algorithm to your own datasets, please prepare two files with the same schema as "pm25_missing.txt" and "pm25_ latlng.txt" respectively, and then run the "stmvl" function. For more details, please refer to the code.

Please cite the following two papers when using the code and dataset.

[1] Xiuwen Yi, Yu Zheng, Junbo Zhang, Tianrui Li. in Proceedings of the 25th International Joint Conference on Artificial Intelligence, IJCAI 2016, June 1, 2016.
[2] Yu Zheng, Licia Capra, Ouri Wolfson, Yang Hai. Urban Computing: Concepts, Methodologies, and Applications. ACM Transaction on Intelligent Systems and Technology (ACM TIST), 5(3), 2014.

More datasets can be found on the homepage of urban computing:
https://www.microsoft.com/en-us/research/project/urban-computing/