Supplement of Atmos. Chem. Phys., 21, 14493–14505, 2021
https://doi.org/10.5194/acp-21-14493-2021-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Clustering diurnal cycles of day-to-day temperature change to understand their impacts on air quality forecasting in mountain-basin areas

Debing Kong et al.

Correspondence to: Guicai Ning (ninggc09@lzu.edu.cn)

K-means clustering

K-means is one of the most commonly used unsupervised learning algorithms that treat the renowned clustering problem (MacQueen, 1967;Hartigan and Wong, 1979;Mokdad and Haddad, 2017). This is, by automatically partitioning the given data set into a certain number of groups selected a priori (assume k clusters). The aim of the K-means algorithm is to divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized. Then, the initial cluster centers are iteratively refined as follows.

Each data point is assigned to its neighboring cluster centroid based on the Euclidean distance metric.

Each cluster centroid is then re-calculated to be the mean of its constituent data points. This can be achieved by minimizing an objective function known as a squared error function. It is defined as:

$$J(v) = \sum_{i=1}^{k} \sum_{j=1}^{c_i} (||x_i - v_j||)^2$$

where

k: is the number of cluster centers;

c_i: is the number of data points in the i^{th} cluster;

$||x_i - v_j||$: is the Euclidean distance between x_i and v_j;

v_j: is the data points in the i^{th} cluster;

x_i: is the centroid vector of the i^{th} cluster.

When there is no further change in assignment of data point to clusters, the K-means algorithm converges to the optimal solution.

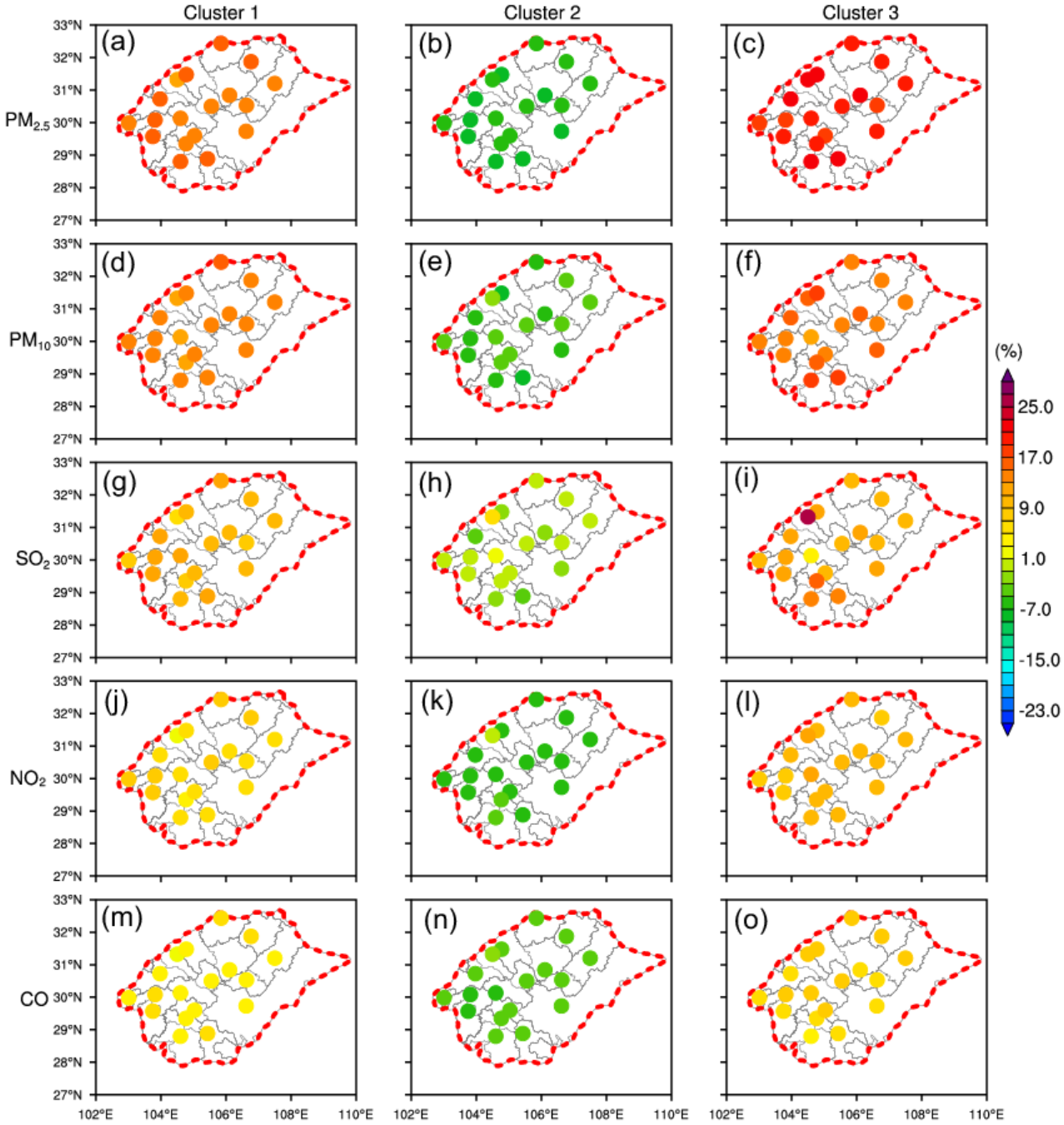

Figure S1 Spatial distribution of percentage values of the day-to-day changes in surface $PM_{2.5}$ (**a–c**), PM_{10} (**d–f**), SO_2 (**g–i**), NO_2 (**j–l**), and CO (**m–o**) concentrations following the three identified diurnal cycles within one day.

References

Hartigan, J. A., and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28, 100-108, doi: 10.2307/2346830, 1979.

MacQueen, J.: Some methods for classification and analysis of multivariate observations, Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1967, 281-297.

Mokdad, F., and Haddad, B.: Improved infrared precipitation estimation approaches based on k-means clustering: application to north Algeria using MSG-SEVIRI satellite data, Adv. Space Res., 59, 2880-2900, doi: 10.1016/j.asr.2017.03.027, 2017.