A genetic algorithm implementation for spatio-temporal variogram modelling to determine air quality monitoring network representativeness

Abstract

One way to improve air quality index estimations is through monitoring network optimization, i.e. determination of optimal sitting criteria and the number of monitoring stations to be included in the network. In this work, we used a genetic algorithm based strategy for parameter search to obtain the best fitted variogram model for air contaminants. This variogram is then used to define a coverage area - representativeness - of the monitoring network for each air pollutant in the Valley of Mexico Metropolitan Area.

1 Introduction

To better estimate population exposure to air pollution and its health effects is necessary to collect reliable data from air monitoring networks. Thus, decision making in public health policy will relay on how accurate measurements are done. Improvement in exposure assessment can be done by establishing an optimal sitting criteria and the number of monitory stations necessary in that network. Spatial representativeness is defined as the area containing just correlated sample measurements [1]. The same idea is used to define temporal representativeness, i.e. the amount of time that measurements show correlation. We propose the use of the variogram to estimate spatial and temporal representativeness of an air quality monitoring network as the joint areas of individual monitoring stations. In Geostatistical procedures such as spatial interpolation, genetic algorithms (GA) have been implemented as parameter search strategy. In particular, using ordinary Kriging interpolation method for ground water hydrology, GA have proven to be more efficient than traditional and heuristic parameter selection [2]. Thus, in this work, we applied a GA-based strategy for variogram parameter search using publicly available information on air contaminants.

2 Materials and methods

Air pollutant data from the Air Quality monitoring network located throughout the Valley of Mexico Metropolitan Area (VMMA) was downloaded using the [R] package 'aire.zmvm'. Information was available for carbon monoxide (CO), ozone (O_3), nitrogen oxide (NO), nitrogen oxides (NO_X), sulphur dioxide (SO_2), particulate matter $PM_{2.5}$, PM_{10} and coarse particles PM_{CO}. A genetic algorithm was used to find the best variogram model to fit the data, where the chromosome was implemented by the combination of covariance structure (*metric, separable, productSum, sumMetric* and *simpleSumMetric*) and spatio-temporal variance models (*Exponential, Gaussian* and *Spheric*) with the corresponding parameters (*nugget, sill* and *range*). The fitness function was the minimal wRMSE over the proposed semivariogram chromosome.

3 Results

By optimizing search parameters for the covariance structure, variogram model and its initial values, we were able to determine the best fitted variogram per contaminant. The *range* parameter from each

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

pollutant, which indicates the correlation length is depicted in Figure 1. Panel A) shows the spatial correlation range while panel B) the temporal component.

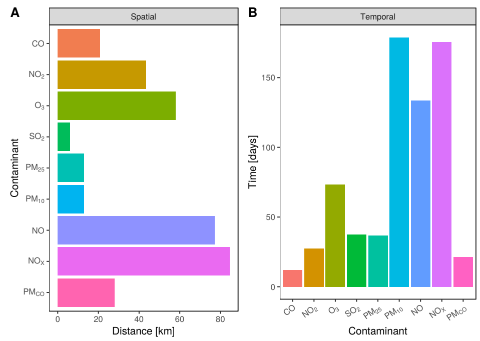

Figure 1: Range parameters for each contaminant. A) Spatial correlation range (km). NO and NO_X have the highest values while SO_2 the smallest. B) Temporal correlation range (days). The largest values correspond to PM_{10} and NO_X and the lowest to CO.

The representativeness area calculated for individual stations, was obtained from the spatial correlation lengths. The combinations of these areas stands for the representativeness area of the whole monitoring network.

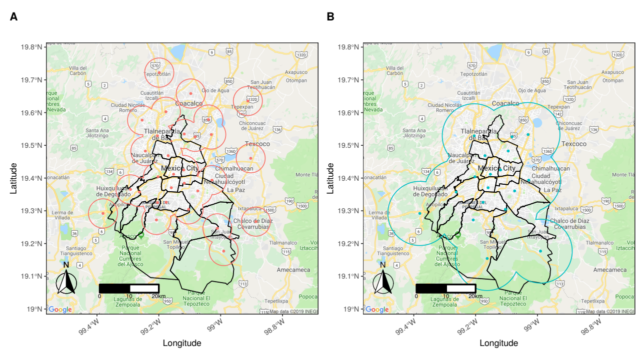

Figure 2: Representativeness area for A) sulphur dioxide SO_2 and B) particulate matter $PM_{2.5}$. The covered area is obtained by joining those of the individual monitoring stations.

4 Conclusion

As in previous reports, representativeness area is shown as pollutant geography - dependant. Here, we have implemented a GA to optimize parameter search space to construct the best spatio-temporal variogram. With the correlation range obtained from this procedure, we were able to generate the monitoring network covered area for each pollutant in the VMMA.

References

[1] Beauchamp, M., Malherbe, L., de Fouquet, C. & Létinois L. (2018) A necessary distinction between spatial representativeness of an air quality monitoring station and the delimitation of exceedance areas, *Environmental monitoring and assessment* **190** (7):441.

[2] Abedini, M.J., Nasseri, M. & Burn, D.H. (2012) The use of a genetic algorithm-based search strategy in geostatistics: application to a set of anisotropic piezometric head data, *Computers & Geosciences* **41** 136-46.