Received 9 February 2024; | Revised 9 February 2024; | Accepted 9 February 2024

STATISTICS FOR HIGH DIMENSIONAL DATA REPORT

Air quality analysis in Madrid city with Hidden Dynamic Geostatistical model

Andrea Arici* | Gabriele Marchesi**

[1]Department of Management, Information and Production, University of Bergamo

Correspondence
*Mat. 1060222, Email:
a.arici4@studenti.unibg.it
**Mat. 1068521, Email:
g.marchesi15@studenti.unibg.it

Summary

Air quality monitoring in modern city is becoming gradually more and more important in the last years both for the health of planet and of the people. In this study, the Hidden Dynamic Geostatistical Model from (Calculli, Fassò, Finazzi, Pollice, & Turnone 2015) is used to better understand the distributions of air pollutants over the city of Madrid. The model's parameter, which are estimated through the EM-algorithm, and the interpolation process are implemented using the D-STEM software. The goal of this study is to understand how the concentration of the considered pollutants are distributed across the entire Madrid surface. Different types of models have been considered to understand the correlation between the pollutants and the weather.

KEYWORDS:
D-STEM, spatio-temporal model, Hidden Dynamic Geostatistical Model, PM_{10}, NO_x

1 | INTRODUCTION

One of the most relevant problems in modern metropolis is pollution. In particular air pollution, which is not only a huge factor in climate change but also for diseases in humans. Higher levels of air pollutants increase the risk of respiratory problems, which have a mortality rate 61% higher than the cardiovascular diseases (Vanos, Hebbern, & Cakmak 2014). Amongst all the air pollutants, we decided, in this study, to focus on nitrogen oxides NO_x and particulate matter PM_{10} (the number 10 means the diameter of the particles in μm) .

NO_x is the primary air pollutant and it's a mixture of NO, N_2O, NO_2. It's mainly generated by burning fossil fuels in industrial plants and in motor vehicles. NO_x also participate in the formation of smog consisting of products resulting from the interaction with organic compounds and, together with sulfur dioxide, contribute to the formation of acid rain (Teixeira, Feltes, & Santana 2008). This pollutant have significant harmful effects on the respiratory tract, such as reduced defences against microorganisms and increased bronchial response in persons with asthma (Organization 2006).

PM_{10} is also a complex mixture of substances from different origins such as elemental carbon, organic carbon and water soluble compounds and has been proven to be directly associated with mortality in elderly adults living in polluted cities. An increase in PM_{10} equal to 100 $\mu g/m^3$ was associated with an increase in overall mortality equal to approximately 13% (Saldiva et al. 1995). The aerodynamics of PM can also affect the severity of adverse health effects. In general, lower sized PM (< 100 nm diameter), is associated with more serious adverse effects after both short-term exposure to an elevated concentration of pollutants or long-term exposure. (Franchini & Mannucci 2009).

Thus the goal of this work is trying to better understand how the particulate matter (PM_{10}) and the nitrogen oxides (NO_x) are spatially distributed and correlated in a particular area of subject using the HDG model. The other main task is trying to interpolate the concentration of these pollutants all over the subject area, starting from the limited data made available by the monitoring stations at their specific coordinates. The testing of this approach and of the model is done by conducting the study on the city of Madrid in the 2019 year. Since, generally, concentrations of pollutants are influenced by meteorological phenomena, information about the weather has also been taken into account.

2 | DATASET DESCRIPTION

In this study two main datasets have been utilized, one for the air pollutants and one for the weather. All the data have been downloaded from the Madrid City Council open portal website[1]. There are a total of 24 stations across Madrid measuring the air pollutants. Unfortunately not every pollutant is measured in each of those stations. The data about pollution have 24 stations measuring the NO_x and only 12 for PM_{10}. The concentration for both pollutants is measured in micrograms per cubic meter $[\mu g/m^3]$.

The meteorological data includes values about temperature, relative umidity, wind speed and barometric pressure in 24 different stations. Some of those stations are in the same location as the pollutant ones. See figure 2 for the locations of the monitoring stations. Like for the pollutant not every station measures all the variable mentioned above, but only a subset of them. There are 23 stations for both temperature and relative humidity but there are only 9 for the wind speed and 8 for the barometric pressure. More informations about both weather and pollutants can be found in the table1 .

The number of stations indicated above, for both pollutants and weather, have measurements for each month of the year (with some missing data). Stations where the selected pollutants and weather data had over eight months of consecutive missing data have been removed. The observations for all the variables taken into account still contained some missing values (table 2 for other info). The data at the source were divided in hourly measurements for each month of the year 2019, starting from the hour 00:00 of January 1 until 23:00 of December 31. Each data point was associated with the station ID, latitude, longitude and altitude of the sensor that made the measurement. The dataset containing the metadata about the stations was also available on the same website as the others. For the analysis in this report, only one observation per day has been taken into account: the measurement at the hour 14:00. An additional dataset, retrieved from the AWS Open Data Terrain Tiles API[2], contains all the altitudes of the whole surface of Madrid and surroundings with a resolution of the order of a thousandth of a degree. The data coming from the API allowed us to get the specific altitude given the latitude and longitude of the point of interest in the study.The picture 1 shows the altitude in the city of Madrid. All data have been then log-standardized for the analysis.

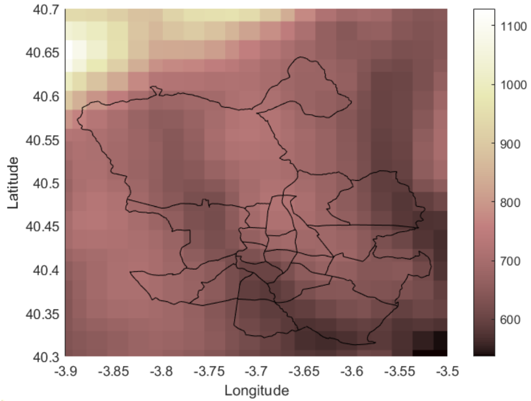


FIGURE 1 Map of the elevations [m] in Madrid area. The lighter area in the top left of the picture corresponds to the mountain range of the "Sierra de Guaderrama"

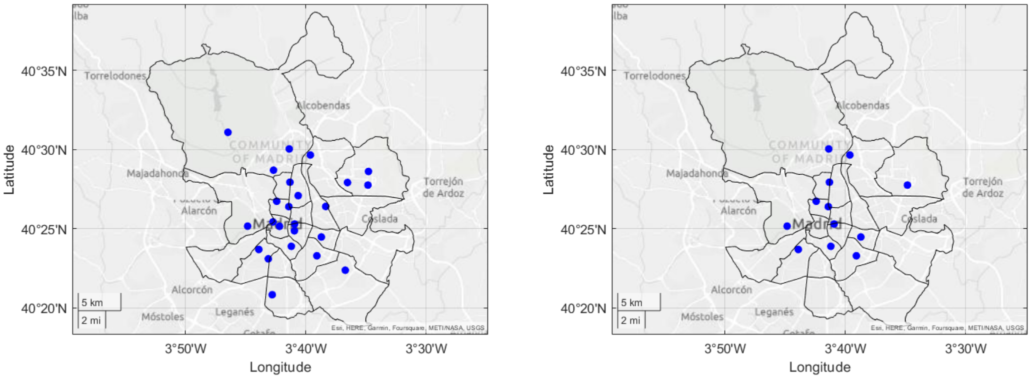
[1] https://datos.madrid.es/portal/site/egob/menuitem
[2] https://registry.opendata.aws/terrain-tiles/

Variable	Description	Measure unit	Name in report
NO_x	Concentration of all nitrogen oxides combined	$[\mu g/m^3]$	NO_x
PM_{10}	Concentration of the particulate matter of diameter smaller or equal than 10 μm	$[\mu g/m^3]$	PM_{10}
Temperature	Temperature on the surface	[°C]	Temp
Relative humidity	Ratio of partial pressure of water vapor in air to the saturation vapor pressure of water at the same temperature	-	Umid
Wind speed	Velocity of the wind	$[m/s]$	Wind
Barometric pressure	Pressure within the atmosphere of Earth	$[mb]$	Press
Elevation/Altitude	Altitude at a specific coordinate w.r.t the sea level	[m]	Alt
Latitude	Latitude of coordinate	[°]	Lat
Longitude	Longitude of coordinate	[°]	Lon
Weekend	Dummy variable which is 1 when the specific observation falls on Sunday or Saturday, 0 otherwise	-	Week

TABLE 1 Summary of all the response variables (first six) and features (last four) in the datasets. The features doesn't include missing data

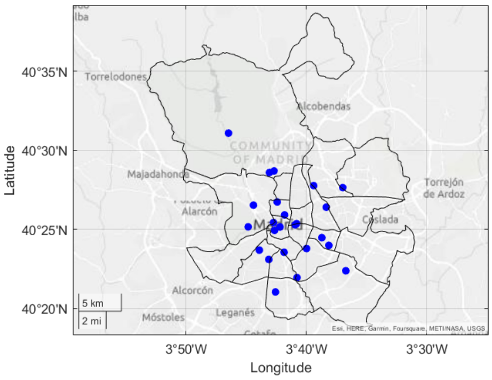
Year	Pollutant	Number of stations	Mean $(\mu g/m^3)$	Standard deviation $(\mu g/m^3)$	Missing (%)
2019	PM_{10}	12	19.495	4.677	3.995
	NO_x	24	46.546	15.967	2.009

TABLE 2 Informations about the pollutants used in the study



(a)

(b)



(c)

FIGURE 2 Position of the monitoring stations in the Madrid city: (a) NO_x (24 sites); (b) PM_{10} (12 sites); (c) meteorological stations (24 sites)

3 | METHODOLOGY

3.1 | The multivariate hidden dynamic geostatistical model

In this section we describe the hidden dynamic geostatistical model by (Calculli et al. 2015) and the related notation. Let **y**(**s**,t) be the p-variate response variable at the location on the sphere $\mathbf{s} \in D \subset \mathbb{S}^2$ at the discrete time t where D is the geographic region of interest. The hidden dynamic geostastical model is a particular spatio-temporal stochastic process defined by the subsequent equations:

$$\mathbf{y}(\mathbf{s},t) = X_\beta(\mathbf{s},t)\boldsymbol{\beta} + \mathbf{z}(\mathbf{s},t) + \boldsymbol{\varepsilon}(\mathbf{s},t) \tag{1}$$

$$\mathbf{z}(\mathbf{s},t) = \mathbf{G}\mathbf{z}(\mathbf{s},t-1) + \boldsymbol{\eta}(\mathbf{s},t) \tag{2}$$

$$\boldsymbol{\varepsilon}(\mathbf{s},t) \sim N_p(\mathbf{0},\boldsymbol{\Sigma}_\varepsilon) \tag{3}$$

$$\boldsymbol{\eta}(\mathbf{s},t) \sim GP_p(\mathbf{0}, V\rho(||\mathbf{s}-\mathbf{s}'||;\theta)) \tag{4}$$

where $\boldsymbol{\beta}$ is a vector of fixed effects, $X_\beta(\mathbf{s},t)$ is the fixed effects matrix. $\mathbf{z}(\mathbf{s},t)$ is a p-variate latent variable with Markovian dynamics ruled by the transition matrix **G**. $\boldsymbol{\epsilon}(\mathbf{s},t)$ is the p-variate Gaussian measurement error independent in space and time, $\boldsymbol{\eta}(\mathbf{s},t)$ is a p-variate sequence of Gaussian random process, independent in time with **0** mean and variance-covariance matrix given by:

$$V\rho(||\mathbf{s}-\mathbf{s}'||;\theta))$$

where $||\mathbf{s}-\mathbf{s}'||$ is the distance between two points **s** and **s**′ which belongs to the region D, **V** is the correlation matrix and $\rho(||\mathbf{s}-\mathbf{s}'||;\boldsymbol{\theta})$ is a valid spatial correlation function of parameter θ. The model parameter set to be estimated is

$$\Psi = \{\boldsymbol{\beta}, \boldsymbol{\Sigma}_\varepsilon, \mathbf{G}, \mathbf{V}_{lt}, \theta\}$$

where $\mathbf{V}_{lt}$ is the lower triangular submatrix of **V**. Note that some restrictions are imposed in the parameter: $\theta > 0$, $|g_i| < 1$ and ${\sigma_i}^2 > 0$, where g_i are the element on the diagonal of the **G** matrix and ${\sigma_i}^2$ on the diagonal of the $\boldsymbol{\Sigma}_\varepsilon$ matrix.
In our study the spatial correlation function is the following one:

$$\rho(||\mathbf{s}-\mathbf{s}'||;\theta)) = \exp\left(-\frac{||\mathbf{s}-\mathbf{s}'||}{\theta}\right)$$

3.2 | Mapping capability of HDG model

Given the the parameter set $\hat{\Psi}$ estimated through the EM-algorithm implemented in D-STEM, new sites $\mathcal{S}_0$ which are disjointed with the ones used in model estimation and time $t = 1, ..., T$, the prediction for the i-th response variable is:

$$\hat{\mathbf{y}}_\mathbf{t}(\mathcal{S}_0,t) = \mathbf{X}_{\beta,i}(\mathcal{S}_0,t)\hat{\boldsymbol{\beta}}_i + \mathbf{z}_t^{T,i}(\mathcal{S}_0) \tag{5}$$

where $\mathbf{z}_t^{T,i}(\mathcal{S}_0)$ is the output of the Kalman smoother in the sites $\mathcal{S}_0$ for each i-th response variable implemented in D-STEM. The software provided also the standard deviations of the prediction as described in (Calculli et al. 2015).

3.3 | Model selection and Leave One Gauge Out Cross-Validation

The validation method used to perform model selection is the Leave One Gauge Out Cross-Validation (LOGOCV), proposed in (Calculli et al. 2015) based on prediction performance by looking at the cross validation root mean squared error (CRMSE) of the prediction in the test set. The LOGOCV is an iterative procedure in which one gauge at a time is removed from all the available ones and used in testing, while the others are used in the training set. This is done for all the station in the dataset and the prediction is performed over all the time horizon taken into account in the study. So in our case the prediction on the test gauge are done for all the 365 days of the 2019 year and then confronted with the actual measured value by the monitoring station. The RMSE for every model is then calculated by averaging the RMSE in every gauge for each day obtaining the CRMSE. From a computational point of view this can be very expensive since for every gauge, a new model is estimated by the EM-algorithm and then the prediction on the test gauge is performed by making the kriging only on the specific test gauge. In our case due to the limited number of stations and the time horizon being relatively small, this procedure is manageable in a little time.

For model selection two types of models are considered: *Full* models, which includes all the time-invariant covariates (*week-end, latitude, longitude and altitude*) and *Selected* models, where only the covariates that are significant at 5% level in the *Full* model are included. Different model dimension are taken into account, see table 3 , this was done to examine the eventual correlations between both pollutants and the weather variables.

Model	Response Variables
1-variate NO_x	NO_x
1-variate PM_{10}	PM_{10}
2-variate	NO_x - PM_{10}
6-variate	NO_x - PM_{10} - $UMID$ - $TEMP$ - $WIND$ - $PRESS$

TABLE 3 HDG model with the relative response variables used in model estimation

3.4 | Comparison with linear model

To verify the predictive capability of the HDG model, a comparison with a classical linear regression model with an autoregressive part is made. Two linear models were built, one for each pollutant, using the coefficients that were significant in the bivariate model. The goal is to confirm the fact that using a more complex spatio-temporal model, like HDG, instead of a simple autoregressive linear model, would give better predictions in terms of CRMSE. The same procedure of the LOGOCV discussed in section 3.3 is applied also to the auto-regressive linear model.

The linear regression follows the classical form

$$\mathbf{y} = \mathbf{X}_{\boldsymbol{\beta}} \cdot \boldsymbol{\beta} + \mathbf{e} \tag{6}$$

where the $\mathbf{y}$ is a $(N \cdot n) \times 1$ vector obtained by stacking the N observation of the pollutant for each one of the n stations, $\mathbf{X}_{\boldsymbol{\beta}}$ is the matrix $(N \cdot n) \times d$ that contains the regressors, where d is the number of regressors included in the model and $\mathbf{e}$ is the modeling error of the same dimensions of the $\mathbf{y}$. The matrix is obtained by stacking the N observations for the relative regressor for each of the n stations for all the d regressors. In this way is possible to jointly estimate the value of the coefficients $\boldsymbol{\beta}$ for all the monitoring stations for the whole period all at once.

4 | RESULTS

As the first three rows of table 4 show, all the CRMSE are quite similar for the HDGM, with the lowest one represented by the bivariate model. Although the multivariate model with meteorological variables could give a better interpretability to the model due to the known interactions between those air pollutants and the weather, the values of the CRMSE shows that it is the worst from a prediction point of view in both PM_{10} and NO_x, furthermore the presence of more response variables in this model led us to choose a simpler model. In both pollutants, the bivariate model shows a better predicting performance with a slightly better CRMSE for both NO_x and PM_{10}, than the univariate model and this fact led us to choose the bivariate as the best model to reach our goal. Moreover, looking at the values of the log-likelihood in the table 9 for the bivariate model can be seen that the difference between the *Full* and *Selected* model is very small, this means that the latter model has a fit capacity which is almost equal to the first one, for this reason the model with less parameters has been chosen for our study (Bivariate selected).

	Full		**Selected**	
CRMSE	NO_x	PM_{10}	NO_x	PM_{10}
1-variate	16.083	9.968	16.138	9.872
2-variate	16.073	9.929	16.029	9.847
6-variate	16.798	9.894	16.901	9.906
linear	-	-	27.779	12.760

TABLE 4 Cross-validation root mean squared errors for the *Full* and *Selected* models of the NO_x and PM_{10} for the 6-variate, 2-variate, 1-variate and linear auto-regressive model case.

By looking at the CRMSE values of the linear model (last row of the same table), is clear that for both the pollutants, they are worse than the ones obtained with the spatio-temporal model, positively confirming the presence of spatial correlation between the observations and also justifying the use of a more complex model such as the HDG.

As shown in table 5 , the coefficients which were significant in the hdg model are also significant in the linear model, confirming the fact that the choices for the $\hat{\boldsymbol{\beta}}$ coefficients in the *Selected* were consistent.

Response variable	$\hat{\beta}_{lat}$	$\hat{\beta}_{alt}$	$\hat{y}_{t-1}$
NO_x	$-0.136_{(0.008)}$	$0.156_{(0.008)}$	$0.741_{(0.007)}$
PM_{10}		$-0.112_{(0.013)}$	$0.568_{(0.013)}$

TABLE 5 Linear model coefficients with log-standardized values. Standard deviations in parenthesis. $\hat{y}_{t-1}$ is the autoregressive coefficient

The results of the estimation of $\boldsymbol{\beta}$, $\mathbf{G}$ and $\boldsymbol{\Sigma}_\varepsilon$ through the EM-algorithm are shown in table 10 .

For the PM_{10} variable we can see that the *elevation* coefficient is negative, meaning that to higher altitude corresponds lower concentration of the pollutant. This match our expectations about the PM_{10} behaviour. Surprisingly, the NO_x seems to behave in the opposite way with respect to the PM_{10}, by increasing with the altitude.

By looking at the value of $\hat{g}_i$ can be seen that PM_{10} has an higher value than NO_x which means that the first one has a higher persistence in the air than the second through time.

The values of $\hat{\sigma}_i^2$ in table 10 show good (low) value for NO_x , while for the PM_{10} it is much higher, meaning that the model doesn't have a great fit capability for this pollutant. This is confirmed by looking at the R^2 in table 8 : the NO_x has a much higher value than the PM_{10}, corresponding to worse prediction capability for the second pollutant.

The estimated $\hat{\theta}$ for this model is 11 km (std = 0.02 km) which is expected for the distances that we are taking into account, this proves that there is a good spatial correlation for the pollutants over the city and a quite smooth behaviour.

The value of the estimated variance-covariance matrix $\hat{V}$ and the relative correlation matrix $\hat{V}_{corr}$ is presented in table 6 and 7 . The two pollutants don't have a strong correlation but it is still of good magnitude and, more than that, positive, which is in line with expectations: when a pollutant is very present in the air we expect that the levels of the other air pollutants will also follow, more or less, the same trend. Moreover, looking at the log-standardized variance-covariance matrix $\hat{V}$ can be seen that the values are quite high for both pollutants, 0.492 for NO_x and 0.584 for PM_{10} showing that our estimates are heavily influenced by the values assumed by the latent variable. This result is in line with our expectations since our $\hat{\boldsymbol{\beta}}$ values don't help much with the prediction of the concentrations.

Response variable	NO_x	PM_{10}
NO_x	$0.492_{(0.014)}$	$0.233_{(0.013)}$
PM_{10}	$0.233_{(0.013)}$	$0.584_{(0.024)}$

TABLE 6 Log-standardized $\hat{\mathbf{V}}$ variance covariance matrix for the bivariate *Selected* model. Standard deviations in parenthesis

Response variable	NO_x	PM_{10}
NO_x	$1.000_{(0.000)}$	$0.433_{(0.030)}$
PM_{10}	$0.433_{(0.030)}$	$1.000_{(0.000)}$

TABLE 7 $\hat{\mathbf{V}}_{\mathbf{corr}}$ correlation matrix for the bivariate *Selected* model. Standard deviations in parenthesis

	Full		**Selected**	
R2	NO_x	PM_{10}	NO_x	PM_{10}
1-variate	0.862	0.520	0.861	0.541
2-variate	0.863	0.535	0.863	0.544
6-variate	0.848	0.538	0.821	0.539
linear	-	-	0.560	0.293

TABLE 8 Cross-validation R squared for the *Full* and *Selected* models of the NO_x and PM_{10} for the 6-variate, 2-variate, 1-variate and linear auto-regressive model case.

Model	Full	Selected
1-variate NO_x	2002.256	1992.266
1-variate PM_{10}	-728.204	-729.590
2-variate	1408.832	1402.115
6-variate	14721.621	14739.954

TABLE 9 Log-likelihoods value of the *Full* and *Selected* models for the 6-variate, 2-variate and 1-variate case.

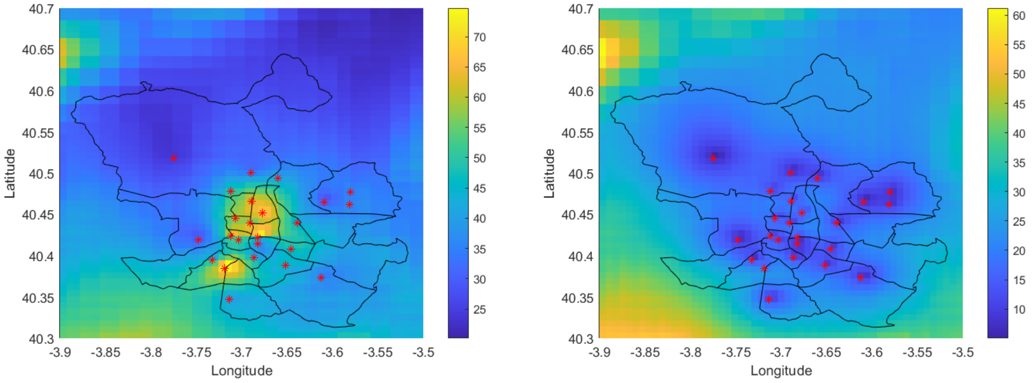
Response variable	$\hat{\beta}_{lat}$	$\hat{\beta}_{alt}$	$\hat{g}_i$	$\hat{\sigma}_i^2$
NO_x	$-0.133_{(0.040)}$	$0.093_{(0.023)}$	$0.684_{(0.004)}$	$0.035_{(0.002)}$
PM_{10}		$-0.112_{(0.026)}$	$0.568_{(0.019)}$	$0.197_{(0.008)}$

TABLE 10 Bivariate *selected* model coefficients in log-standardized values. Standard deviations in parenthesis. $\hat{g}_i$ and $\hat{\sigma}_i^2$ are respectively the value on the diagonal of the $\mathbf{G}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ matrices of the corresponding response variable (NO_x or PM_{10})

These parameters of the Bivariate model were used to perform the kriging on a regular grid with a resolution of 0.629 km^2 that covers the entire city.

This process provided the predictions $\hat{\mathbf{y}}_{\mathbf{i}}$, and their standard errors, for each point of the grid. The map of the results has been overlapped with the map of the city, allowing us to see the predicted concentrations in each district of Madrid.

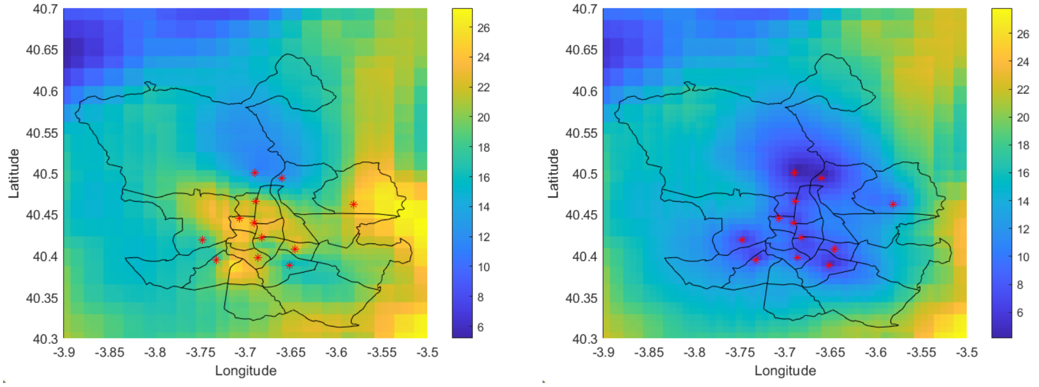
The figures 3 a 3 b show the spatial predictive capability of the Bivariate *Selected* model for the NO_x pollutant. This is the result of kriging done using the model parameters vector $\hat{\Psi}$ estimated by the EM-algorithm applied to the fine regular



(a)

(b)

FIGURE 3 Average NO_x concentration [$\mu g/m^3$] predictions (a) and standard deviations of the predictions [$\mu g/m^3$] (b) over the 2019 year



(a)

(b)

FIGURE 4 Average PM_{10} concentration predictions [$\mu g/m^3$] (a) and standard deviations of the predictions [$\mu g/m^3$] (b) over the 2019 year

grid mentioned above. In particular these images show the value of the average concentrations and standard deviations of the predictions during the entire 2019 year. Looking at these images can be noticed that the predictions precision is higher near the monitoring stations and tends to gradually decrease when we move away from them. This might be due to the fact that there are only two regressors (latitude and altitude) for predicting the values and they are not able to explain much about the response variable and also because the latent variable tends to go to zero when we are far away from the measurements due to the negative exponential function.

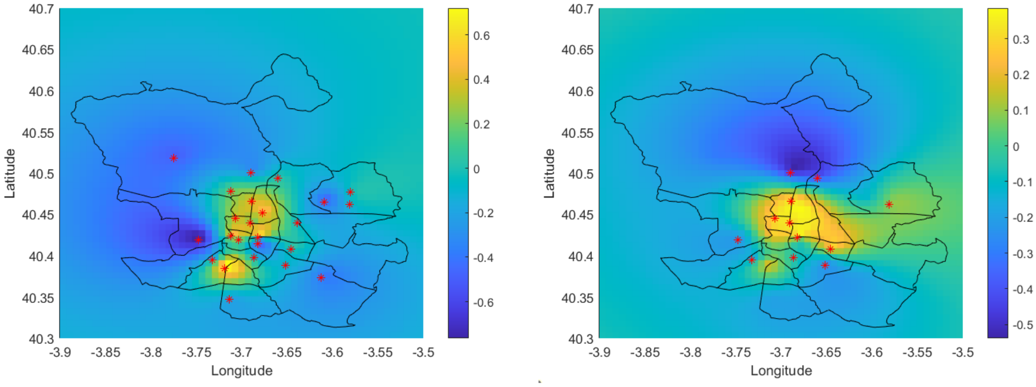
In fact by looking also at the latent variable of the NO_x 5 a there is a hotspot in the centre of Madrid, where the majority of the stations are located. When we move to the borders of the city, where there aren't any stations, the values tends to be quite low, and on the border of grid, they are at their minimum.

By looking at the PM_{10} graphs 4 a 4 b, we can see that, similarly to the NO_x, as soon as we move from the stations, the precision drops. Also since we have a lower number of stations, so its more difficult to cover the entire city with good predictions. Even though, there is some correlation between the NO_x and the PM_{10}, we can't notice any significative difference of the predictive precision in the locations of the NO_x stations.

Looking at the latent variable for the PM_{10} case 5 b, we also see a similar behaviour to the one for the NO_x.

In the upper part of Madrid, we lack of measuring stations since there is the natural park of "Fuencarral-El Pardo". PM_{10} is really difficult to predict while for the NO_x, we have one station which allow us to have an idea of the pollutant concentrations in the middle of the park.

Also looking at the concentration prediction graphs we notice that the location of the hotspot is different for the two pollutants. However, this consideration need to be taken with a grain of salt, since both predictions have high standard values.



(a)

(b)

FIGURE 5 Average log-standardized NO_x (a) and PM_{10} (b) latent variables over the 2019 year

5 | CONCLUSIONS

In conclusions looking at the obtained results, we can assert that using a spatio-temporal model led us to better results than a classical autoregressive linear model, due to the existing spatial correlations of the pollutants' concentrations over Madrid. The goal of this work can be considered only partly achieved: good performances were reached in the proximity of the stations, obtaining a good interpolation with satisfactory precision of concentrations between them, but when we move away from them, the significant regressors (like the *elevation*) in the model were not able to help the predictions far from the monitoring stations, obtaining quite poor results in the furthest area. Furthermore the latent variable exponentially drop to lower values as soon we move from the monitoring areas, explaining why in the outer part of Madrid we get lower prediction values. The medium level correlation between NO_x and PM_{10} in our data helped only partly in the combined prediction of them due to a correlation value which wasn't very high, but still helped w.r.t the univariate HDG model showing slightly better prediction capabilities in terms of CRMSE.

The meteorological variables unfortunately didn't increase the prediction capability w.r.t the other model taken into account forcing us to not consider them in our study.

References

Calculli, C., Fassò, A., Finazzi, F., Pollice, A., & Turnone, A. (2015). Maximum likelihood estimation of the multivariate hidden dynamic geostatistical model with application to air quality in apulia, italy. *Environmetrics*, *26*(6), 406–417.

Franchini, M., & Mannucci, P. M. (2009). Particulate air pollution and cardiovascular risk: short-term and long-term effects. In *Seminars in thrombosis and hemostasis* (Vol. 35, pp. 665–670).

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: a review. *Frontiers in public health*, *8*, 14.

Organization, W. H. (2006). *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. Author.

Saldiva, P. H., Pope III, C. A., Schwartz, J., Dockery, D. W., Lichtenfels, A. J., Salge, J. M., … Bohm, G. M. (1995). Air pollution and mortality in elderly people: a time-series study in sao paulo, brazil. *Archives of Environmental Health: An International Journal*, *50*(2), 159–163.

Teixeira, E. C., Feltes, S., & Santana, E. R. R. d. (2008). Estudo das emissões de fontes móveis na região metropolitana de porto alegre, rio grande do sul. *Química Nova*, *31*, 244–248.

Vanos, J. K., Hebbern, C., & Cakmak, S. (2014). Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 canadian cities. *Environmental Pollution*, *185*, 322–332.

□