21st International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
27-30 September 2022, Aveiro, Portugal

SPATIAL MODEL FOR DAILY AIR QUALITY HIGH RESOLUTION ESTIMATION

H21-137

Morgan Jacquinot[1], Romain Derain[1], Alexandre Armengaud[1], Sonia Oppo[1]
[1]AtmoSud, Air Quality Observatory in South Region, France, www.atmosud.org
Contact: morgan.jacquinot@atmosud.org

INTRODUCTION

AZUR is a modelling platform that creates daily high-resolution concentration cartographies for pollutants like PM10, PM2.5 and NO_2 on millions grid cells in short time computing compared to deterministic model. It produces cartography until day+2 at 25m of resolution, considering the punctual measurements and forecasts. The input data of this platform are:

- annual concentration maps coming from ADMS-Urban model mix with geostatistical method (Seaton et al. 2022).
- punctual measurement and forecast simulated by the Eurlerian Chemistry Transport model CHIMERE (Menut et al., 2021).

This work examines the relationships between daily and annual nitrogen dioxide values in order to provide a statistical model capable of fine-scale estimation of daily concentrations over large areas. The reduced computation time allows the provision of daily maps used by the French air quality observatories.
In this study we worked on all 22 nitrogen dioxide measurement stations of the PACA region. The park is characterized by 6 stations under traffic influence and 16 rural, urban, and peri-urban background stations.
The daily value of interest for nitrogen dioxide is the daily hourly maximum. In addition, we have annual averages estimated on a regular grid covering the whole PACA region. They are produced with the ADMS Urban model associated with a post processing by kriging. Annual spatial variations are provided by the large number of temporary campaigns carried out using passive tubes in addition to the 22 fixed stations of the park.
The relationships between daily and annual values are studied (Gressent et al., 2020). We perform an analysis of the pairs of measuring stations for which we calculate the ratio of their daily and annual values respectively. From these results we highlight that this relationship depends on the range of values of the daily measurements considered. This range of values is represented by the rank of the daily measurements when they are seen as percentiles of their annual distribution.
Fine scale daily maps are usually computed from dispersion models involving multiple parameters related in particular to emissions which require regular updating and high computation time. The aim of this work is to provide a simpler model, easily adaptable to other regions and capable of estimating nitrogen dioxide concentrations with a good approximation.

1 RELATIONSHIP BETWEEN ANNUAL AVERAGE AND HOURLY MAXIMUM

In this part we study the relations between the annual means and the daily values in nitrogen dioxide on the whole of the couples of stations of measurements of the PACA region. These relations depend on the rank p of the daily values.

Let us consider an annual history of the daily values of a station of measurement of nitrogen dioxide. For all the pairs of stations s_i.and $s_{i'}$ and for a fixed rank p, we calculate the ratio of their daily deciles q_p :

$$\frac{q_p(s_{i'})}{q_p(s_i)}$$

As well as the ratio of their annual average y:

$$\frac{y(s_{i'})}{y(s_i)}$$

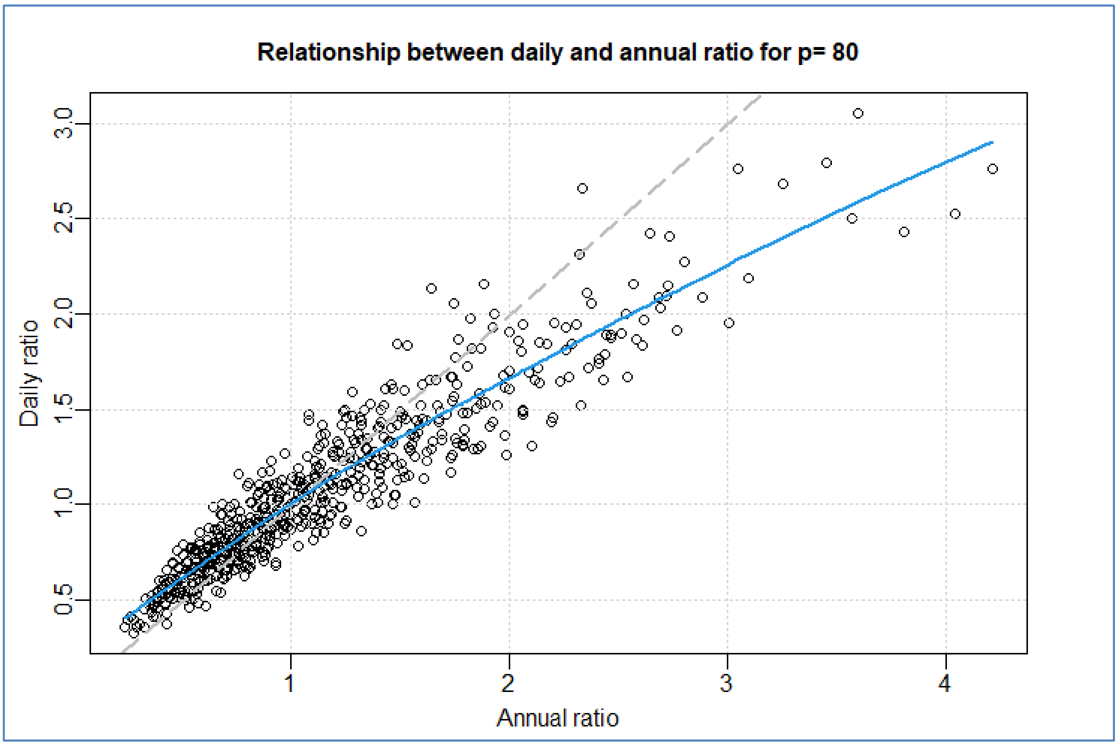

Figure. 1 : Relationship between the ratios of the annual means and the ratios of daily deciles of rank 80, in red a spline function fit.

Figure 1 shows that for the daily deciles q_{80} their ratios are lower than the ratios of the means, indeed the curve of adjustment is under the bisector represented in dotted line. By calculating these ratios for several deciles, we can observe how these relationships vary (Figure 2).

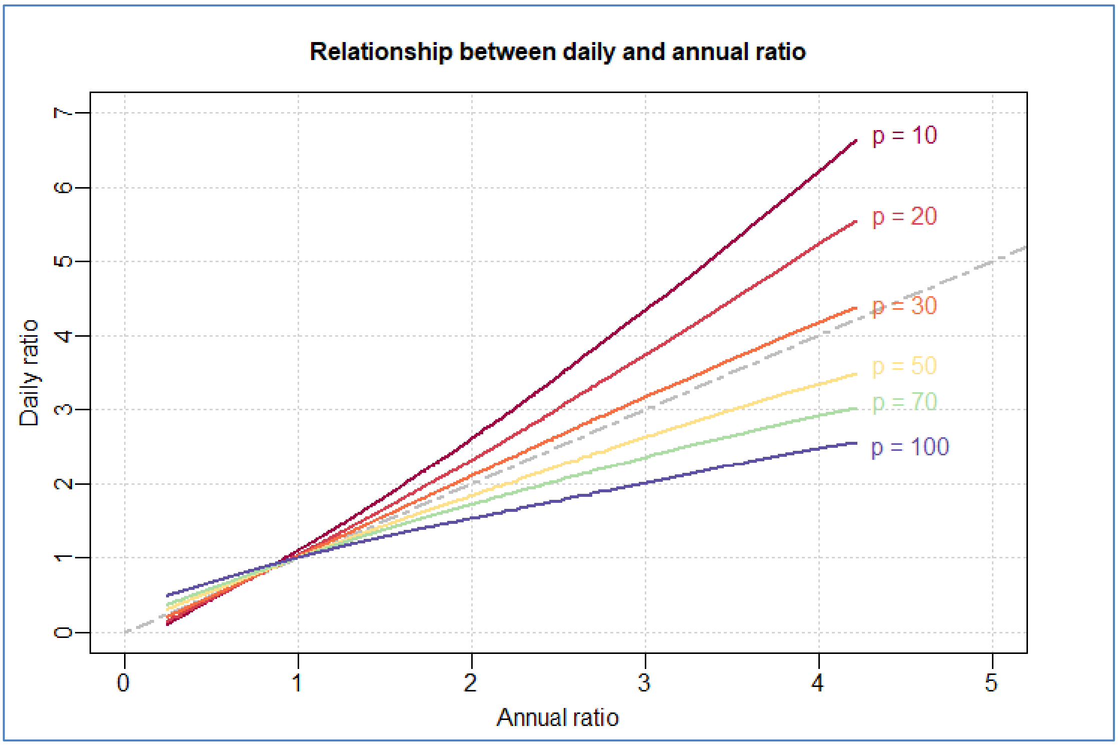

Figure 2 : Relationship between the ratios of annual averages and daily deciles for 6 deciles, represented by their spline fit.

The representation of these relationships for different daily deciles (Figure 2) shows that they evolve as follows:

- - For the deciles of low rank (q $_{\{10, 10, 30\}}$) the daily ratios are higher than the annual ratios
- - For the deciles of higher rank (q $_{\{50, 70, 100\}}$) the daily ratios are lower than the annual ratios.

Consider a pair of stations with an annual ratio of 4 (Figure 2). On days with high NO_2 levels in the air, the daily ratio is equal to 2.5 (q_{100} curve). On days with low NO_2 levels, the daily ratio is equal to 6.5 (q_{10} curve). In general, when NO_2 levels increase, the daily ratios decrease. For two measuring stations the ratio of their daily deciles, therefore, depends on the ratio of their annual averages as well as the rank p. The following relationship can be written:

$$\frac{q_p(s_{i'})}{q_p(s_i)} = f\left(\frac{y(s_{i'})}{y(s_i)}, p\right) \quad (1)$$

We suggest for this function f a polynomial of degree n. Equation (1) then becomes:

$$\frac{q_p(s_{i'})}{q_p(s_i)} = \sum_{j+k<n} \beta_{j,k} \left(\frac{y(s_{i'})}{y(s_i)}\right)^j p^k \quad (2)$$

With $\beta_{j,k}$ coefficients and $p \in [0,100]$.

2 USING THE MODEL

Let s_0 be a point of the mesh. We have an evaluation of its annual value $y(s_0)$. The daily value at point s_0 is the unknown to be determined. We note $\hat{q}_{s_i}(s_0)$ the estimate of this value made from the station s_i. Equation (2) then becomes:

$$\hat{q}_{s_i}(s_0) = q_p(s_i) \sum_{j+k<n} \beta_{j,k} \left(\frac{y(s_0)}{y(s_i)}\right)^j p^k \quad (3)$$

This formulation assumes that if the day value measured at the station is the p th percentile in the annual daily value distribution then the estimated day value at the grid point corresponds to the percentile of the same rank. This assumption is verified when the point s_0 is in the area of representativeness of the station s_i.This translates into the fact that the station's area of representativeness must be homogeneous in terms of weather and emission variations.
Note that the average emissions need not be identical at the grid point and at the station due to the presence of the term $\frac{y(s_0)}{y(s_i)}$ in the equation (3) which allows for spatial variations in the annual average pollution level.
A mesh point may be in the representativeness area of several neighboring measurement sites, or it may be on the border of several areas without belonging entirely to one of them. Therefore, for a given mesh point, each station s_i whose area of representativeness contains s_0 produces an estimate $\hat{q}_{s_i}(s_0)$. The overall estimate at s_0, denoted $\hat{z}(s_0)$, given by equation (4), is the inverse distance-weighted average of the estimates $\hat{q}_{s_i}(s_0)$.

$$\hat{z}(s_0) = \sum_{s_i \in E_{s_0}} \lambda_i \hat{q}_{s_i}(s_0) \quad with \sum_{s_i \in E_{s_0}} \lambda_i = 1 \quad (4)$$

With :

- E_{s_0}: set of stations whose area of representativeness contains the mesh point s_0,
- λ_i: weights depending on the distance of s_i to s_0,

The weights λ_i are calculated from the inverse square distance of the stations s_i at the mesh point s_0.

3 RESULTS

In order to evaluate the performance of the model, we calculate cross-validation estimates for the pollutants NO_2 and PM10 over the year 2019. The objective variables are the daily hourly maximum for NO_2, and the daily average for PM10.
The annual values used correspond to the year 2018, the ranks are calculated from the distribution of day values from 2016 to 2018.
In order to compare the results obtained with the suggested method, we perform a kriging with external drift using the annual mean, in global neighborhood on the same set of stations. The adjustment of the daily variogram is done automatically with a zero-nugget effect.
Tables 1 and 2 present the results of the leave-one-out-cross-validation by group of stations, on the one hand the background sites, on the other hand the sites under the influence of road traffic. For NO_2, for the background sites, the suggested method has an advantage of 4.7 % with an RMSE of 15.79 against 15.04. For the sites under traffic influence, the kriging method is better by 9.87 % with an RMSE of 17.41 against 19.12. Concerning PM10, the difference is null for the background sites and 19.93 % for the traffic sites, in favor of the proposed method.

	RMSE Proposed Method	RMSE Krigeage	RMSE deviation %	Correlation (R) Proposed method	Correlation (R) Krigeage
Background stations	15.04	15.79	4.76%	0.81	0.79
Traffic stations	19.12	17.41	-9.87%	0.78	0.79

Table 1 : NO_2 scores of the two methods by leave-one-out-cross-validation for all stations in the fleet.

	RMSE Proposed Method	RMSE Krigeage	RMSE deviation %	Correlation (R) Proposed method	Correlation (R) Krigeage
Background stations	4.61	4.62	0.09%	0.85	0.84
Traffic stations	6.24	7.79	19.93%	0.85	0.76

Table 2 : PM10 scores of the two methods by leave-one-out-cross-validation for all stations in the fleet.

4 CONCLUSION

We have suggested a statistical method of spatial estimation for ambient air quality. It allows to build daily maps for different pollutants from an annual map and a network of measurement stations. The scores obtained are close to those of a kriging with external drift for the background stations, with more marked differences for the sites under the influence of road traffic.
The inverse distance interpolation implies a fixed representativeness from one day to another. This leads us to consider an improvement of the model for the calculation of the weights. The calculation of a variogram on the rank of daily values could replace the inverse distance proposed by equation (4). This would imply variable range on a daily basis as in the case of kriging.
The presented method can be easily adapted for forecast mapping by replacing the daily value measured at the stations by the forecasted value. The calculation of the rank is then made from this forecasted value within the distribution of the measured daily values. Finally, the model can also be adapted for PM10 and PM2.5 and even for ozone if the domain has a well defined annual spatial structure.

5 REFERENCES

- Gressent A., L. Malherbe, A. Colette , H. Rollin, R. Scimiab: Data fusion for air quality mapping using low-cost sensor observations, Environment International Volume 143, October 2020, 105965.
- Seaton M, O'Neill J, Bien B, Hood C, Jackson M, Jackson R, Johnson K, Oades M, Stidworthy A, Stocker J and Carruthers D,2022: A Multi-model Air Quality System for Health Research: Road model development and evaluation. Environmental Modelling and Software, DOI: 10.1016/j.envsoft.2022.105455.
- Menut L., B. Bessagnet, R.Briant, A.Cholakian, F.Couvidat, S.Mailler, R.Pennel, G.Siour, P.Tuccella, S.Turquety, and M.Valari, The CHIMERE v2020r1 online chemistry-transport model, Geoscientific Model Development, 14, 6781-6811, 2021.