

Aplicación de un método de combinación de medidas y resultados de modelos en la evaluación de la calidad del aire en España

Application of a method for combining measured data and modelling results in air quality assessment in Spain

Fernando MARTÍN, Inmaculada PALOMINO & Marta G. VIVANCO

División de Contaminación Ambiental, Departamento de Medio Ambiente
CIEMAT- Ministerio de Ciencia e Innovación, Spain
fernando.martin@ciemat.es

Received: 23 June 2009

Accepted: 21 September 2009

RESUMEN

Se presenta la aplicación de un método para combinar las mediciones y los resultados de concentraciones de contaminantes atmosféricos obtenidos con el modelo CHIMERE. Este método tiene por objeto proporcionar información más realista de la calidad del aire en España en respuesta a los requerimientos de la legislación nacional y europea. El método consiste en utilizar técnicas de regresión lineal e interpolación *kriging* para corregir los resultados del modelo ajustándose mejor a las mediciones. Se aplica de forma separada a la contaminación rural y urbana obteniendo mapas para cada caso, para luego combinarlos según el grado en que cada zona del territorio puede considerarse rural o urbana. Se muestran y discuten los resultados obtenidos para la media anual de PM₁₀ en España para 2004.

Palabras clave: Calidad del aire; PM₁₀; Modelo CHIMERE; *Kriging*; regresión lineal.

ABSTRACT

The use of a method to combine air pollution measurements and CHIMERE model data is presented. This method is to response to the European and Spanish regulatory requirements to provide reliable information about the air quality in Spain. The method consists of using linear regression and kriging interpolation to correct the model results improving the fit to the observations. It was separately applied to the rural and urban air pollution yielding maps for each case, which then were combined by taking into account the distribution of rural and urban areas in the domain. The results for the annual mean of PM₁₀ for 2004 in Spain are shown and discussed

Key words: Air quality; PM₁₀; CHIMERE model; kriging; linear regression.

SUMMARY: 1. Introduction. 2. Methodology. 3. Results and discussion. 4. Summary and conclusions. 5. Acknowledgements. 6. References

1. INTRODUCTION

The atmosphere is an extremely complex reactive system in which numerous physical and chemical processes occur simultaneously. Ambient measurements only provide information about the atmospheric conditions at a particular time and location. Moreover, measurements alone cannot be directly used by policymakers to establish an effective strategy for solving air quality problems. Mathematical models provide the necessary framework to integrate of all the understanding of individual atmospheric processes (chemistry, transport, removal, etc) and study of their interactions. However, due to the large number of uncertainties in air quality modelling, such as emission data or model parameterizations, the models has to be validated and their uncertainties must be determined and reduced. Concerning the air quality assessment in a territory, the combination of measurements with validated models rises as a more adequate approach because more advantage is taken from the accuracy of the measurements and the good spatial coverage of the models results.

The annual air quality assessment in every EU member state is mandatory due to the European directives. The main tool for this assessment is the use of measured data in air quality stations. However, the use of supplementary techniques as modelling is allowed to complement the measured data. In Spain, air quality assessment has been carried out following the requirements of the European legislation. Modelling is being used to generate more consistent maps of the air quality variables: exceedances of limit values and averaged concentrations (Martín et al., 2004, Vivanco et al., 2007).

In this paper, the new method to combine measurements and modelling results for air quality assessment is shown. Some preliminary tests are also discussed.

2. METHODOLOGY

2.1 Description of the measurements-model combination methodology

Combining model results and observations has been commonly applied in meteorological modelling, but several studies have been also done to combine observations and air quality models. Denby et al. (2005) review different methodologies to combine and assimilate observations and models. Tarrason et al. (1998) developed a methodology to combine both EMEP monitoring and modeling data. According to this, the difference between modelled and monitored concentrations at each measurement point is taken to create a difference field by interpolating all the difference values. Finally, the model and difference field are added together. The German system “FLADIS” (Wiegand and Diegmann, 2000) combines interpolated measured data and the result of both statistical and physical models, based on emission, meteorological and topographical data. The combination is carried out using a linear weighting factor.

Martín et al. (2005) proposed to use a methodology based on defining an influence area for each observation depending on certain factors: the features of the

station, the distance between the grid point and the measurement point and the wind flow. This is based in the assimilation techniques (Benjamin and Seaman, 1985) used in meteorological models. However, this methodology was used in some annual air quality assessments in Spain, but it shows to be less efficient because the correction of models results are done only in areas next to the stations giving a less realistic map of the real distribution of pollutants. This was more evident in areas with very few and scattered observations.

The methodology used in this study is based on the idea that the real concentration of an atmospheric pollutant C in a station k can be expressed as

$$C_k = M_k + e_k + s_k \quad (1)$$

where M_k is a concentration estimate (i.e., by a dispersion model), e_k is the systematic error of the estimate (i.e., modelling error) and s_k is the inherent error or measurement error. The question is how to reduce the model error e_k , that is, how to correct the model results to provide a best fit to observations and to get a more realistic map of the spatial distribution of pollutant concentrations. Among the several options, the linear regression and the kriging interpolation methods are the most interesting (Fiala, 2009).

The linear regression technique assumes that a better estimate of the concentration C'_k can be obtained by

$$C'_k = aM_k + b + r_k \quad (2)$$

where a and b are the regression coefficients and r_k is the residual error which includes the measurement error and the non-solved part of the modelling error. This method corrects the concentration estimates by taking into account any influence of the concentration values on them.

The kriging interpolation technique assumes:

$$C'_k = \sum_{i=1}^n \lambda_i M_{k_i} + r_k \quad \sum_{i=1}^n \lambda_i = 1 \quad (3)$$

being λ_i the weights assumed on the basis of a variogram in order to minimize the mean-square-error, they range between 0 and 1. The variogram is a function representing how a measured variable varies with distance. In our study, the variogram can be computed by plotting the values of the concentration differences (or the model residuals) between pairs of stations against the distances between them. The resulted scatter plot can be fitted to simple functions, such as logarithm, exponentials, etc. This method corrects the concentration estimates by taking into account any influence of the distance or spatial representativeness of the air quality stations on the concentration estimates.

In this paper, both methods were applied separately and sequentially; firstly the linear regression and then, the kriging method based on a variogram of the concentration residuals resulting from the linear regression method. The results for annual PM10 concentrations are discussed in order to determine what method provides better results. It is done separately for rural and urban stations obtaining maps of pollutant concentrations correcting the model estimates with measured data. Finally, both maps are merged by computing a weighted average in the every grid

cell using weights related to population or land use in order to estimate whether a cell is mostly urban or rural.

2.2 Description of the models and model setup

Simulations of photochemical compounds were carried out using a regional version of the CHIMERE chemistry-transport model (Bessagnet et al., 2004; Hodzic et al., 2005). This model calculates the concentration of 44 gaseous species and both inorganic and organic aerosols of primary and secondary origin, including primary particulate matter, mineral dust, sulphate, nitrate, ammonium, secondary organic species and water. This model is being used for the annual simulations of air quality in Spain since 2004 (Martín et al., 2004, Vivanco et al., 2007). It has been evaluated using measured data of ambient pollutant concentrations from a large number of Spanish stations (Vivanco et al., 2006, Vivanco et al., 2009a and b) and compared with other models such as CMAQ (Baldasano et al., 2008). The model was shown to be suitable for air quality assessment as the uncertainty statistics were lower than the maxima established by the EU directives and the EPA criteria. The impact of the spatial computing resolution was also discussed in Vivanco et al. (2008).

The MM5 model was the meteorological processor used to feed the CHIMERE model. The models were run over European domains and then, over Iberian Peninsula ones. Domains for MM5 simulations are larger than the CHIMERE ones. The spatial resolution for MM5 simulations was 36 and 19 km for the European and Iberian Peninsula domains, respectively. MM5 simulations were forced by the National Centre for Environmental Prediction model (GFS) analyses at both scales. Regarding the CHIMERE simulations, the modelling domains are shown in Fig. 1. For the European domain, a 0.5° horizontal resolution and 14 vertical sigma-pressure levels extending up to 500 hPa were used. And for the Iberian Peninsula domain, the horizontal resolution was 0.2° resolution, using a one-way nesting procedure where coarse-grid simulations force the fine-grid ones at the boundaries without feedback.

For both domains emissions were derived from the annual totals of the EMEP database for 2003 (Vestreng et al., 2005). Original EMEP emissions were disaggregated taking into account the land use information, in order to get higher resolution emission data. The spatial emission distribution from the EMEP grid to the CHIMERE grid is performed using an intermediate fine-grid at 1 km resolution. This high-resolution land use inventory comes from the Global Land Cover Facility (GLCF) data set (<http://change.gsfc.nasa.gov/create.html>). Boundary conditions for the coarse domain were provided from monthly climatology from LMDz-INCA model (Hauglustaine et al., 2004) for gases concentrations and from GOCART model (Chin et al., 2002) for particulate species.

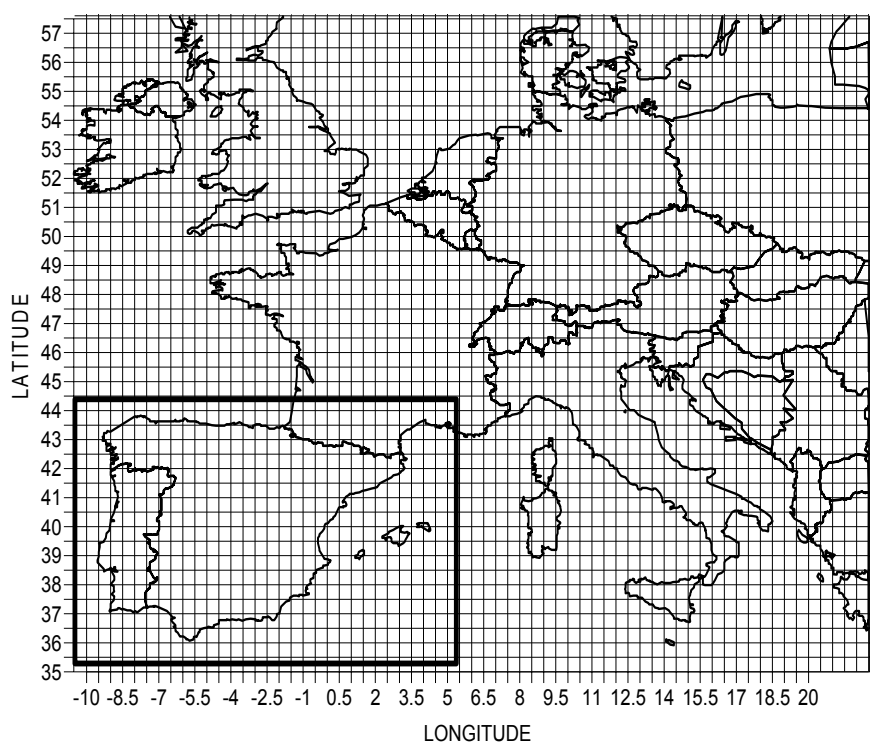


Figure 1. Map showing the computing domains used with the CHIMERE model.

3. RESULTS AND DISCUSSION

The spatial distribution of annual mean concentrations of PM₁₀ in Spain was computed using the CHIMERE model. It is plotted along with the measured PM₁₀ concentrations in stations in Fig. 2. Differences between observations and model results are detected, especially in urban areas. It is due to the spatial representativeness of many urban stations (less than 1 km²) which is much less than the spatial resolution of the model results (20x20 km²). Hence, the model can hardly fit the urban data in some cases. It can be solved by increasing the spatial resolution for modelling but the computational load increases a lot, and it is only possible in smaller domains unless supercomputers are used. Other option is to correct the model results to fit the measured data in stations. This is the common approach in air quality assessment in the EU member states based on the use of dispersion modelling along with air quality stations data. This is the way to take advantage of the accuracy of the measurements and the spatial coverage of the model results.

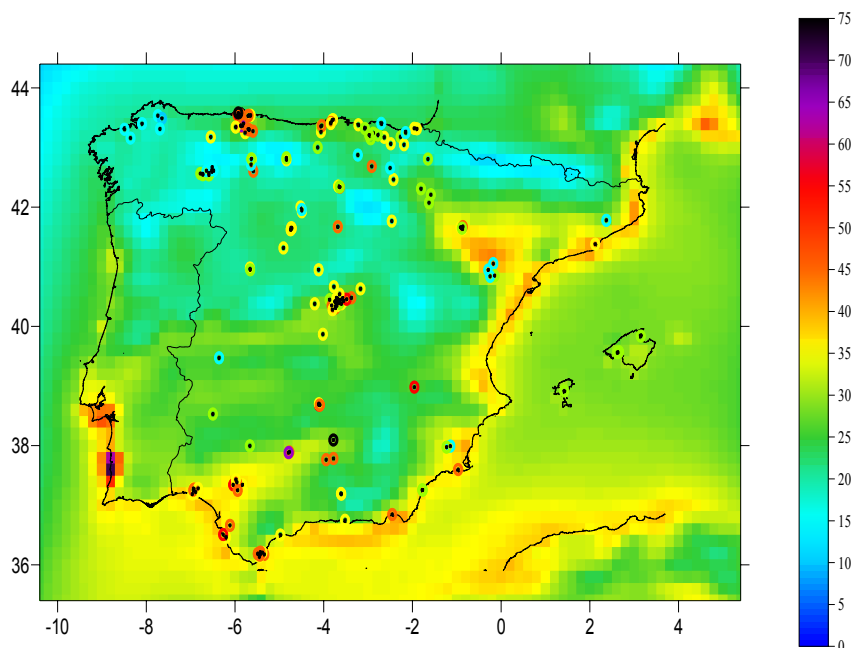


Figure 2. Map showing the annual mean concentrations ($\mu\text{g m}^{-3}$) of PM₁₀ in Spain for 2004 computed by models (isolines) and by air quality stations (colour dots).

As shown in Fig. 2, there are two main distribution patterns of concentrations. One is related to rural background pollution, which has smooth spatial variation. The model results fit basically this pattern due to the relatively coarse spatial resolution of $20 \times 20 \text{ km}^2$. The other one is related to the urban contribution that is highlighted by many of the measured data from monitoring stations which are mostly in urban areas.

Hence, the corrections of the model results must be done separately for rural areas and for urban areas. Firstly, the relation between modelled and observed concentration was investigated. A large set of stations with a percentage of valid data higher than 80% was used. Linear regression models (eq. 2) were determined for the groups of urban and rural stations. The results show that the correlation coefficient is very low for urban stations, but not for the rural ones, especially for the background rural stations, where correlation is quite good (Fig. 3). It is due that the rural background stations have a spatial representativeness quite similar to the spatial resolution of the CHIMERE model estimates ($20 \times 20 \text{ km}^2$).

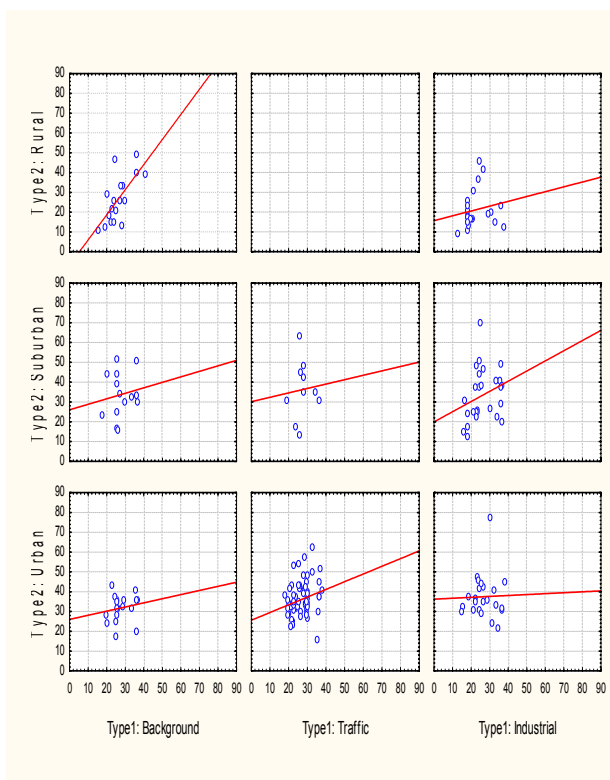


Figure 3. Scatter plots of model (x) and observed (y) concentration data ($\mu\text{g m}^{-3}$) for rural stations (upper), suburban (middle row) and urban (lower) and for background (left), traffic (middle column) and industrial (right) stations.

Linear regression equations were applied to correct the CHIMERE model estimates and the residuals for the rural stations and the urban ones were computed. These residuals and those of the CHIMERE model were analyzed in order to estimate how the residuals changes against the distance (Fig. 4). This information was used to estimate the needed weights for spatial interpolation of the residuals following eq. (3) for the kriging method, which was applied to the residuals of the CHIMERE model and to the residuals of the linear regression. Among several variograms used (linear, logarithmic, etc), the spherical type seems to provide better results.

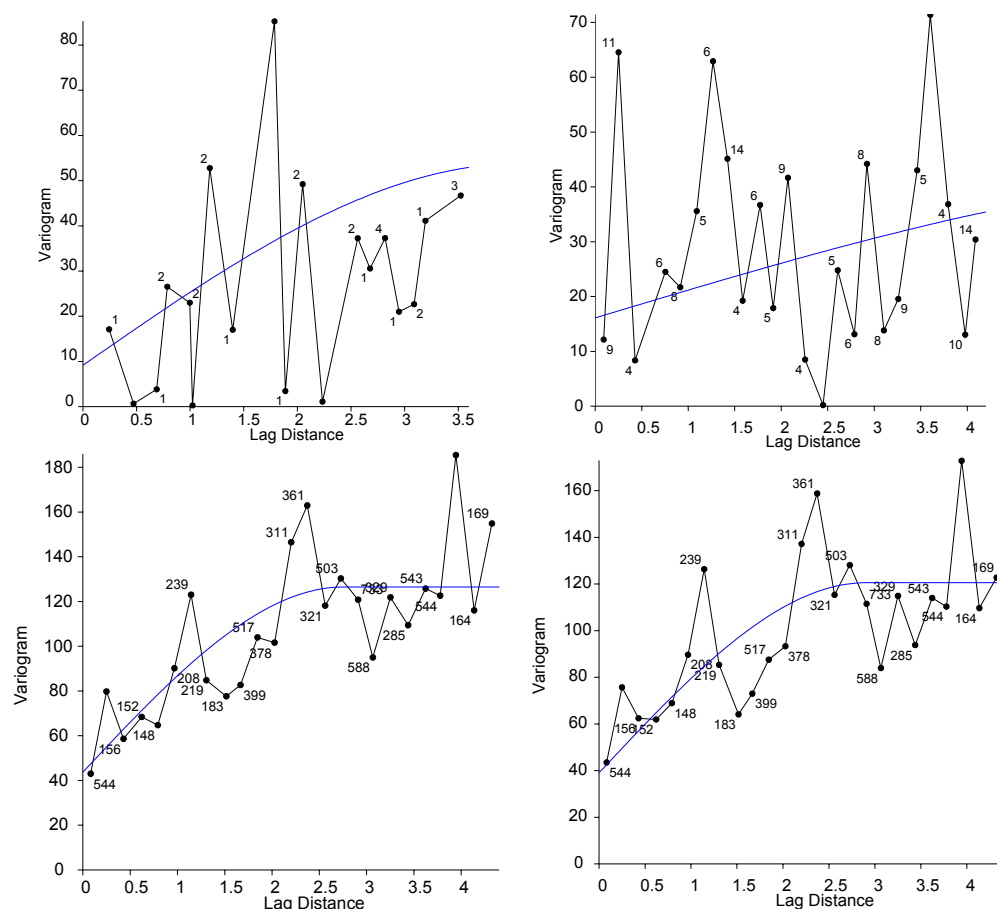


Figure 4. Variation of the CHIMERE (left) and linear regression (right) residuals of concentrations ($\mu\text{g m}^{-3}$) against the distance for rural (upper) and urban stations (lower). The spherical variograms are also plotted (blue lines).

The results of these interpolation techniques are shown in Fig. 5. The differences between rural and urban maps are very important. In urban maps, the concentrations are higher. The maxima are in several areas of the Mediterranean coast, the Ebro Valley, southern areas of the peninsula and some industrial and urban areas. When only kriging is used the differences between rural and urban maps are more notable. The kriging rural map shows areas in the Pyrenees with very low concentrations close to $5 \mu\text{g m}^{-3}$, which is much lower than the rural background measurements (Querol et al., 2006). On the other hand, the kriging method with urban stations seems to slightly overpredict the concentrations in some grid cells. However, the maps resulted from linear regression plus kriging show smoother results in rural maps. It is due to the correction done in the CHIMERE model results by the linear regression as the slope coefficient is clearly lower than 1, which reduces the concentration variability. It must be pointed out that linear regression correction

is more suitable for rural map because of a nice correlation between observations and model data for rural stations, while it was not so for urban ones. Then, it should be better to use linear regression correction plus kriging of residuals for the rural map, but only kriging for model residuals for the urban map. It must be noted that this procedures has been applied only for Spanish territory, not for Portugal. It explains the few changes in the rural and urban map for this country.

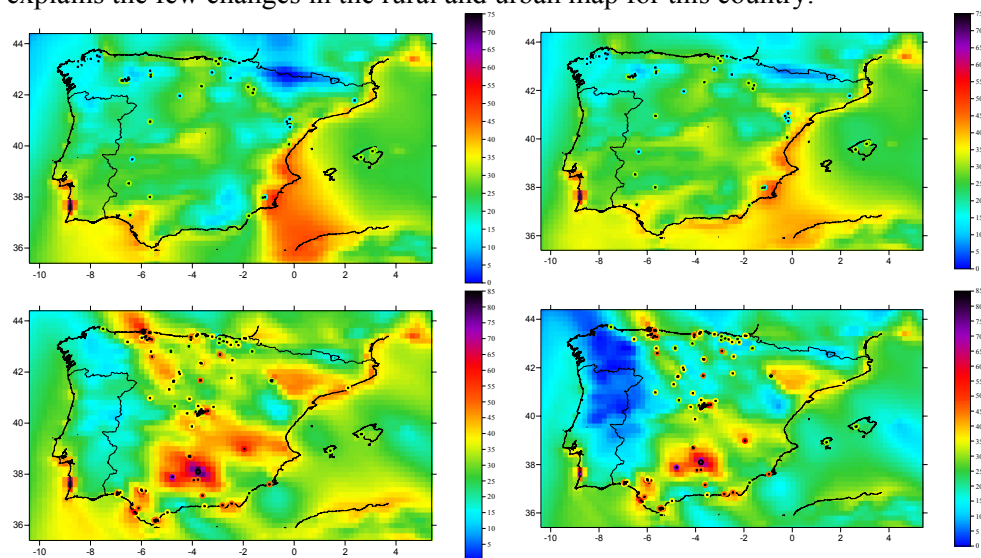


Figure 5. Maps showing the concentration ($\mu\text{g m}^{-3}$) maps resulted of applying the linear regression technique plus the kriging method for the regression residuals (right) and only kriging for the model residuals (left) for rural (upper) and urban stations (lower).

The final PM10 concentration map must rise from merging both maps (rural and urban). The procedure to merge them consists in computing a weighted average of the rural and urban concentration in every grid cell. The question is what surrogated indicator should be used. Two indicators were proposed: one, based on the land use data information from the Global Land Cover Facility (GLCF) data set of NASA and the other one was the population data from Spanish Institute of Statistic (*Instituto Nacional de Estadística*, INE).

In the first case, the percentage of urban land use in every grid cell was used to estimate the weight, assuming the value of 1 for the grid cell with the maximum percentage of urban areas and 0 for those with the minimum. A linear dependence was assumed for the others

For the second case, the way for obtaining the weights was similar assuming a linear dependence with the population but assigning a weight of 1 to cells with population higher than 250000 inhabitants and 0 to those with the minimum of population.

In Fig. 6, the resulted PM10 concentration maps obtained with both methods are shown. It seems that the urban areas are better represented when population data

and kriging method are used rather than land use data and the linear regression plus kriging. It should be due to some errors in the land use data. When linear regression is not used, some unrealistic concentrations are estimated such as Pyrenees areas with very low PM10 concentrations and the sea between Murcia-Alicante and Algeria coasts with high concentrations.

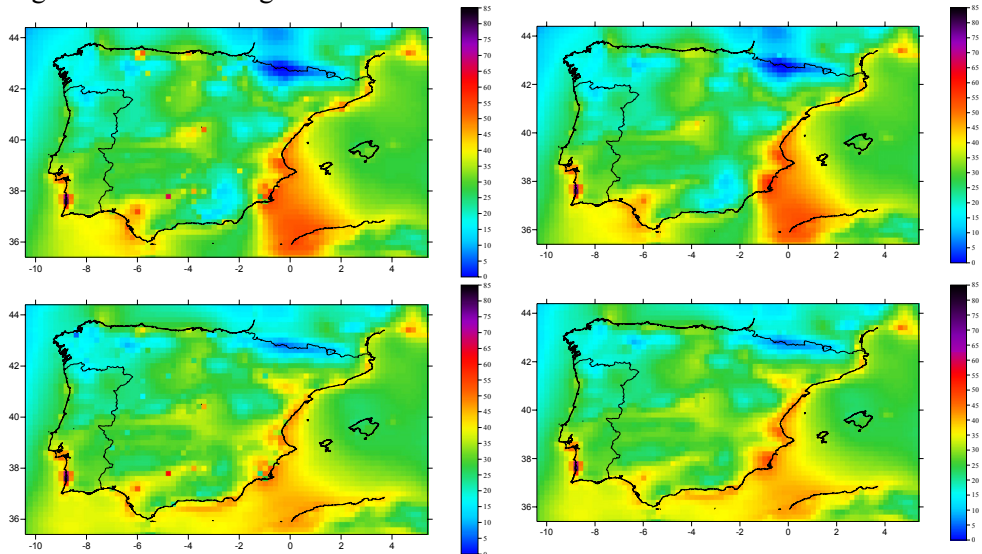


Figure 6. PM10 concentration maps ($\mu\text{g m}^{-3}$) resulted of using the land use (right) and population (left) methods for merging the rural and urban concentrations maps resulted of using only the kriging method (upper) and the linear regression plus kriging (lower) shown in Fig 5.

Statistics (RMSE and Normalized Bias) were computed for checking how good the methods are in estimating the PM10 corrected concentration for a set of urban and rural stations not used in the described process of combination of measurements and model data (Table 1). Best results were obtained when linear regression and kriging were applied to the CHIMERE model data for rural stations and only kriging for urban ones and using population as surrogate indicator for merging urban and rural maps. The corresponding map is depicted in Fig. 7. It seems to correct the shortcomings observed in maps of Fig. 6 and be more realistic. The minima values are about $10 \mu\text{g m}^{-3}$ in remote areas of mountain areas. The maxima (about $40\text{--}50 \mu\text{g m}^{-3}$) are in the Mediterranean coast and the Ebro and Guadalquivir valleys along with some urban and industrial areas. The urban areas are clearly remarked with higher values than those in the surroundings.

	CHIMERE model		CHIMERE model + kriging for urban and rural stations		CHIMERE model + linear regression and kriging for urban and rural stations		CHIMERE model + linear regression and kriging for rural and kriging for urban stations	
Popula- tion criteria	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
Rural	0.26	5.65	0.23	5.67	0.16	8.26	0.25	7.02
Urban	-0.17	9.55	-0.12	9.50	-0.20	10.10	-0.12	8.76
All	-0.01	8.35	0.01	8.31	-0.07	9.48	0.01	8.17
Land use criteria	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
Rural	0.26	5.65	0.23	5.58	0.26	6.95	0.27	6.95
Urban	-0.17	9.55	-0.12	10.48	-0.15	9.53	-0.13	9.64
All	-0.01	8.35	0.01	9.01	0.00	8.68	0.01	8.76

Table 1. Normalized Bias and Root Mean Square Error (RMSE) computed with a set of checking urban and rural stations not used for concentration mapping for the CHIMERE model results, three types of combination of model and observed data and two methods for merging rural and urban maps.

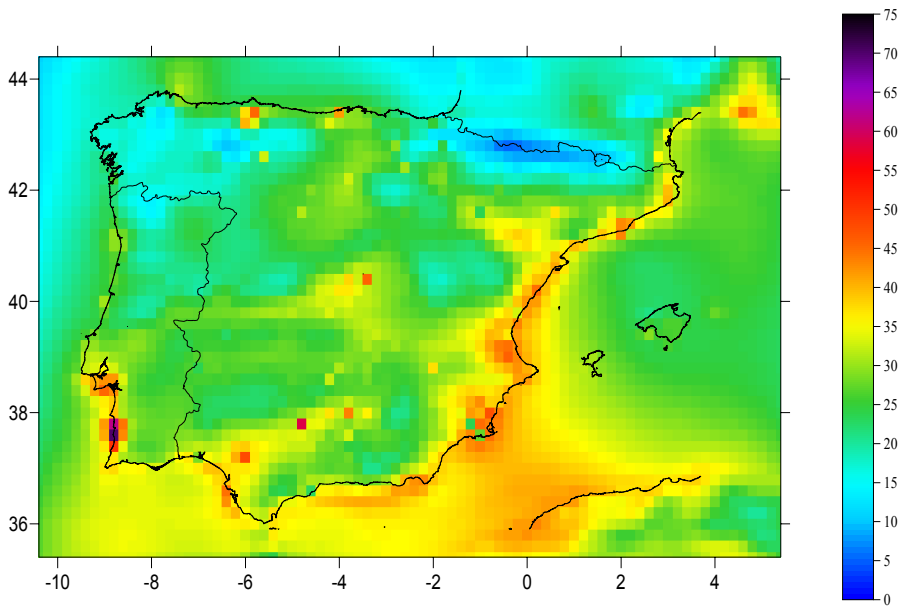


Figure 7. Final PM10 concentration map ($\mu\text{g m}^{-3}$) resulted of merging the rural map resulted of the linear regression plus kriging methods and the urban map from the kriging method using the population as surrogate indicator.

4. SUMMARY AND CONCLUSIONS

The use of a method to combine air pollution measurements and model data estimated with the CHIMERE model are shown in order to provide reliable information about the air quality in Spain in response to the European and Spanish regulatory requirements. The method consists of using linear regression and kriging interpolation to correct the model results improving the fit to the observations. It was separately applied to the rural and urban air pollution yielding maps for each case, which then were combined by taking into account the distribution of rural and urban areas in the domain. The methodology was applied to the annual mean of PM₁₀ for 2004 in Spain.

As expected the differences between rural and urban maps are very important with higher concentrations in urban maps. When only kriging is used the differences between rural and urban maps are more notable. However, the maps resulted from linear regression plus kriging show smoother results. Linear regression correction is more suitable for rural map because of a nice correlation between observations and model data for rural stations, while it was not so for urban ones.

Rural and urban maps must be merged to provide final PM₁₀ annual mean concentration maps for 2004. The best result was obtained when the rural map resulted of the linear regression plus kriging methods and the urban map from the kriging method were merged using the population as surrogate indicator.

However, more studies will be done in the near future for other pollutants and more years. Other methods will be tested comparing in order to improve the results and mapping as needed for correct reporting on air quality assessment in Spain. An important task is to investigate how the spatial distribution of air quality stations can influence on the results of applying the proposed methodology in this paper.

5. ACKNOWLEDGEMENTS

This work is being financially supported by Ministerio de Medio Ambiente, Medio Rural y Marino under a collaboration agreement for the use of atmospheric modelling for air quality assessment in Spain.

Los autores quieren expresar un recuerdo especial de la profesora Elvira Zurita, que fue una maravillosa persona, profesora, investigadora y amiga. Ella estará siempre en nuestros corazones.

6. REFERENCES

- BALDASANO J.M., P. JIMÉNEZ-GUERRERO, O. JORBA, C. PÉREZ, E. LÓPEZ, P. GÜERCA, F. MARTÍN, M.G. VIVANCO, I. PALOMINO, X. QUEROL, M. PANDOLFI, M.J. SANZ & J.J. DIEGUEZ (2008). Caliope: an operational air quality

- forecasting system for the Iberian Peninsula, Balearic Islands and Canary Islands – first annual evaluation and ongoing Developments. *Adv. Sci. Res.*, 2, 89–98
- BENJAMIN, S.G. & N.L. SEAMAN (1985). A simple scheme for objective analysis in curved flow. *Mon. Wea. Rev.*, 113, 1184–1198.
- BESSAGNET, B., A. HODZIC, R. VAUTARD, M. BEEKMANN, L. ROUIL & R. ROSSET (2004). Aerosol modeling with CHIMERE –first evaluation at continental scale. *Atmos. Environ.*, 38, 2803–2817.
- CHIN, M., P. GINOUX, S. KINNE, B.N. HOLBEN, B.N. DUNCAN, R.V. MARTIN, J.A. LOGAN, A. HIGURASHI & T. NAKAJIMA (2002). Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sunphotometer measurements. *J. Atmos. Sci.*, 59, 461–483.
- DENBY, B., HORÁLEK, J., WALKER, S.E., EBEN, K. & J. FIALA (2005). Interpolation and Assimilation Methods for European Scale Air Quality Assessment and Mapping Part I: Review and Recommendations, ETC/ACC Technical Paper 2005/7.
- FIALA J. (2009). Spatial assessment of PM₁₀ and ozone concentrations in Europe (2005) EEA Technical report No 1/2009. ISSN 1725-2237. 52 pp.
- HAUGLUSTAINE, D.A., F. HOURDIN, L. JOURDAIN, M.A. FILIBERTI, S. WALTERS, J.F. LAMARQUE & E.A. HOLLAND (2004). Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: description and background tropospheric chemistry evaluation. *J. Geophys. Res.*, 109, doi:10.1029/2003JD003957.
- HODZIC, A., R. VAUTARD, B. BESSAGNET, M. LATTUATI & F. MORETO (2005). On the quality of long-term urban particulate matter simulation with the CHIMERE model. *Atmos. Environ.*, 39, 5851–5864.
- MARTÍN F., I. PALOMINO & M. GARCÍA (2004). Aplicación de un modelo de dispersión para la evaluación de la calidad del aire en España. Año 2003 y actividades del 2º Semestre de 2004. Informe para la Dirección General de Calidad y Evaluación Ambiental. Ministerio de Medio Ambiente. 19 Noviembre 2004. Ref: 05/2004.
- MARTÍN F., I. PALOMINO & M. GARCÍA (2005). Aplicación de un modelo de dispersión para la evaluación de la calidad del aire en España. Año 2004 y actividades del 2º Semestre de 2005. Informe para la Dirección General de Calidad y Evaluación Ambiental. Ministerio de Medio Ambiente. Ref: 07/2005.
- QUEROL X. & 24 co-authors (2006). Atmospheric particulate matter in Spain: levels, composition and source origin. CSIC and Ministerio de Medio Ambiente. 39pp.
- TARRASÓN, L., A. SEMB, A.-G. HJELLBREKKE, S. TSYRO, J. SCHAUG, J. BATNICKI & S. SOLBERG (1998). Geographical distribution of sulphur and nitrogen compounds in Europe derived both from modelled and observed concentrations. The Norwegian Meteorological Institute, Oslo, Norway (EMEP/MS-C-W Note 4/98).
- VESTRENG, V., K. BREIVIK, M. ADAMS, A. WAGENER, J. GOODWIN, O. ROZOVSKAYA & J.M. PACYNA (2005). Inventory Review 2005, Emission Data Reported to LRTAP Convention and NEC Directive, Initial review of HMs and POPs, Technical Report MSC-W 1/2005, ISSN: 0804-2446.
- VIVANCO M.G., I. PALOMINO, F. MARTÍN & B. CASADO (2006). Actividades 2º semestre 2006. Evaluación de la calidad del aire en España mediante modelos. Informe para la Dirección General de Calidad y Evaluación Ambiental. Ministerio de Medio Ambiente. CIEMAT. Ref: 09/2006.
- VIVANCO M.G., I. PALOMINO & F. MARTÍN (2007). Air quality assessment in Spain for 2002–2004 using modelling techniques. Proceedings of the 6th International Conference on Urban Air Quality, Cyprus, 27–29 March 2007.

- VIVANCO, M.G., M. CORREA, O. AZULA, I. PALOMINO & F. MARTÍN (2008). Influence of model resolution on ozone predictions over Madrid area (Spain). International Conference on Computational Science and Its Applications. Italy.
- VIVANCO M.G., I. PALOMINO, F. MARTÍN & B. CASADO (2009a). Evaluación de la calidad del aire en España utilizando modelización combinada con mediciones, Re-evaluación 2006. Encomienda de Gestión MMA-CIEMAT para evaluación de la calidad del aire mediante modelización. Unidad de Contaminación Atmosférica. CIEMAT. Madrid. 15 Abril 2008, Ref: 17/2009.
- VIVANCO M.G., I. PALOMINO, R. VAUTARD, B. BESSAGNET, F. MARTÍN, L. MENUT & S. JIMÉNEZ (2009b). Multi-year assessment of photochemical air quality simulation over Spain. *Environmental Software & Modelling*, 24, 63–73.
- WIEGAND, G. & DIEGMANN, V. (2000). FLADIS - A system for extending air pollution point data to continuous spatial information (according to EU Council Directive 96/62/EC); In: Air Pollution VIII, WIT Press, 191-200, ISBN 1-85312-822-8.