Preprints of the 17th IFAC Symposium on System Identification
Beijing International Convention Center
October 19-21, 2015. Beijing, China

Improving the PM_{10} estimates of the air quality model AURORA by using Optimal Interpolation ⋆

Oscar Mauricio Agudelo * **Peter Viaene** ** **Bart De Moor** *

* *KU Leuven, Department of Electrical Engineering (ESAT) - STADIUS / iMinds Medical IT, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. E-mail: {mauricio.agudelo, bart.demoor}@esat.kuleuven.be.*
** *VITO - Flemish Institute for Technological Research, Environmental Modeling Unit, Boeretang 200, BE-2400 Mol, Belgium. E-mail: peter.viaene@vito.be*

Abstract: This paper presents the results of using a data assimilation technique known as Optimal Interpolation (OI) for improving the PM_{10} estimates of the air quality model AURORA. Ground-based measurements provided by IRCEL (the Belgian Interregional Environment Agency) have been used in the data assimilation process. AURORA has been set up to cover a domain consisting of Belgium, Luxemburg, and small parts of France, Germany and the Netherlands with a grid resolution of 5×5 km^2. In this study, the characterization of the error covariances between the grid-cells has been done by using the Hollingsworth-Lönnberg method. The simulation results show that the optimal interpolation technique manages to significantly reduce the error of the PM_{10} estimates, which indirectly leads to an improvement of the $PM_{2.5}$ field.

Keywords: Data assimilation, State estimation, Optimal interpolation, Air quality

1. INTRODUCTION

Particulate Matter (PM) is the term used for solid or liquid particles found in the air. Depending on the size of the particles we can distinguish two groups: $PM_{2.5}$ (up to 2.5 micrometers in aerodynamic diameter) and PM_{10} (up to 10 micrometers in aerodynamic diameter). The inhalation of these particles has harmful effects on human health such as asthma, lung cancer, cardiovascular issues, birth defects, and premature death (Pope III et al. (2002)), and therefore accurate estimation of PM is of vital importance.

Despite the big effort in creating models that accurately describe the distribution of air pollutant concentrations, considerable differences between the model estimates and the measurements can be observed in many cases (e.g. Vautard et al. (2012)). The complexity of the mechanisms involved, and the uncertainties in the model parameters and model inputs such as boundary conditions, meteorological patterns and emissions, are among others the main causes of these discrepancies. Here is where data assimilation comes into play. In data assimilation the outputs of a numerical model are combined with observational data for improving the quality of the estimates, while taking into account the errors or uncertainties of both the model and the measurements. These errors are defined in statistical terms and they play a very important role in the success or failure of the assimilation process.

Among the existing data assimilation techniques we can mention the variational methods (3D-var and 4D-var), the Kalman-filter based methods (e.g., Ensemble Kalman Filter, Ensemble Transform Kalman Filter, etc.) and the statistical methods such as Optimal Interpolation (OI). Unlike the Kalman-filter based methods where the forecast error covariances are time-dependent (they are estimated at each assimilation step), in the optimal interpolation method these covariances are time-invariant (the covariances are approximated by using an empirical covariance function), which makes this technique a very attractive option when computational resources are limited. Some recent applications of this method in air quality domain can be found in Wu et al. (2008), Tombette et al. (2009), Candiani et al. (2009) and Kumar et al. (2012).

In this paper we apply the optimal interpolation technique to the air quality model AURORA, to improve its PM_{10} concentration estimates for a period where a PM_{10} pollution episode took place, that is, elevated PM_{10} levels were registered. It is important to remark that for this period the estimations made by the model do not properly follow the observational data, and this poses a serious challenge for the data assimilation method.

This paper is organized in the following way. In Section 2, we present a brief description of AURORA as well as the network of air quality stations used for assimilation and validation purposes. Section 3 presents the optimal inter-

⋆ This work was supported by: • Research Council KUL: CoE PFV/10/002 (OPTEC), PhD/Postdoc grants • Flemish Government: iMinds Medical Information Technologies SBO 2015 • Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017).

Copyright © IFAC 2015

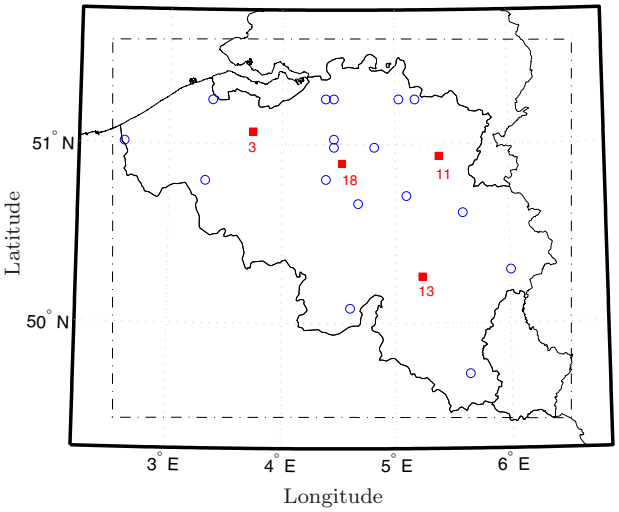

Fig. 1. Geographical location of the stations measuring PM_{10}. The blue circles and red squares represent assimilation and validation stations respectively. The dash-dot line delimits the area covered by AURORA.

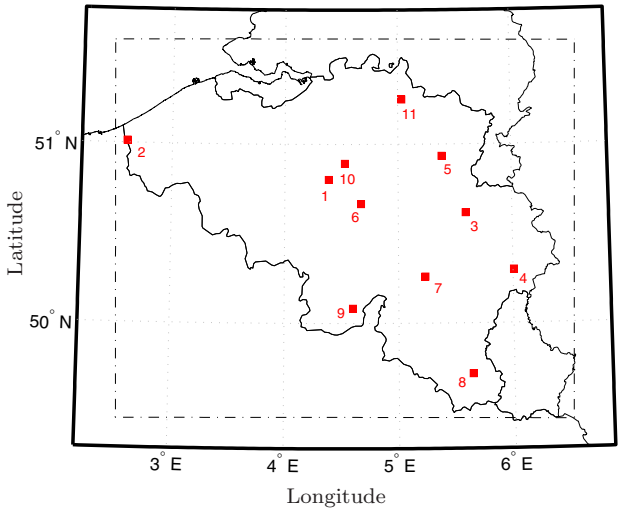

Fig. 2. Geographical location of the $PM_{2.5}$ stations (red squares). The dash-dot line delimits the area covered by AURORA.

polation technique and the approach followed for modeling the error covariances between the model variables. The technical details of the assimilation experiments and the simulation results are presented in Section 4. Finally in Section 5 some concluding remarks are given.

2. DESCRIPTION OF AURORA

AURORA (Air quality modeling in Urban Regions using an Optimal Resolution Approach) is a three-dimensional Eulerian model developed by VITO (Flemish Institute for Technological Research) for simulating air pollution in the lower troposphere at urban or regional scale (Mensink et al. (2001)). AURORA includes modules that describe (i) the transport by diffusion and advection of a number of trace gas components emitted both by anthropogenic and biogenic sources (ii) the (photo)chemistry of this gas mixture resulting in formation of secondary pollutants and (iii) the loss of these components by both dry and wet deposition. The thermodynamic equilibrium between the gas phase and the inorganic aerosol is modeled using a module based on ISORROPIA (Nenes et al. (1998)). Both the coarse (particles that have an aerodynamic diameter between 2.5 and 10 micrometers) and fine (particles with an aerodynamic diameter smaller than 2.5 micrometers) fractions of the species that form PM_{10} are modeled separately in AURORA. These species include both primary (emitted) and secondary (formed from others components) particles. The primary particles considered in AURORA are elementary carbon, and primary particulate matter (e.g., dust or sea salt). The secondary particles include inorganic particles consisting of ammonium, nitrate and sulphate, and secondary organic aerosols which are formed by condensation of oxidation products with a lower volatility.

In this study AURORA has been configured to cover a region that includes Belgium, Luxembourg, and small parts of the Netherlands, France and Germany (see the area delimited by the dash-dot line in Fig. 1). The horizontal domain was divided into 59×49 grid-cells with a resolution of 5 km. Regarding the vertical domain, 22 layers were used to span an altitude of 3000 m starting with a lower layer thickness of about 20 m. The model was set up for simulating a period of 12 days starting on January 20th, 2010. This period is of special interest since within it a PM_{10} episode took place, that is, elevated PM_{10} concentrations were registered.

The air quality data for carrying out the assimilation experiments were provided by IRCEL, the Belgian Interregional Environment Agency (IRCEL (2011)). Since the spatial resolution of the model is 5 km, only measurements from rural and urban-background stations were deemed representative and taken into account. In Fig. 1 we can see the location of the PM_{10} air quality stations used in this work. These 21 stations have been split into two groups: 17 assimilation stations and 4 validation stations. The first group is used in the assimilation process to obtain the optimal estimate of the state. The second group is not used in the assimilation, but only to verify the results. The validation stations have been selected in such a way that they are evenly distributed along the spatial domain and surrounded by nearby assimilation stations. Given that $PM_{2.5}$ is part of PM_{10}, we also evaluate the impact of assimilating PM_{10} measurements by checking the improvement in the estimation of the $PM_{2.5}$ concentrations. Again the data were provided by IRCEL, and due to the resolution of the model we only considered measurements coming from rural and urban-background stations. Fig. 2 shows the 11 stations finally selected.

In the rest of this paper the deterministic model AURORA will be succinctly presented as:

$$\mathbf{x}(k+1) = \mathbf{M}_{\mathrm{A}}(\mathbf{x}(k)) \tag{1}$$

where $\mathbf{x}(k)$ is the state vector containing the concentrations of different pollutants at every grid-cell, and $\mathbf{M}_{\mathrm{A}}(\cdot)$ is the nonlinear state-space operator that calculates the

Table 1. Performance of the data assimilation techniques when PM_{10} measurements are assimilated

Species	Kind of station	Model or Algorithm	Correlation	Mean Bias [$\mu g/m^3$]	RMSE	
					value [$\mu g/m^3$]	reduction [%]
PM_{10}	Assimilation	AURORA	0.36	-12.7213	33	-
		EnKF	0.94	-0.0161	7.7973	76.37
		OI	0.95	-0.0589	6.8227	79.31
	Validation	AURORA	0.44	-12.8412	34.2636	-
		EnKF	0.91	0.7483	13.4132	60.85
		OI	0.92	0.4230	13.0477	61.92
$PM_{2.5}$	-	AURORA	0.41	-11.1752	29.5165	-
		EnKF	0.93	-2.3667	11.0398	62.6
		OI	0.95	-2.4928	9.9925	66.15

concentrations at time $k+1$ from the concentrations given at time k.

3. OPTIMAL INTERPOLATION

Optimal interpolation (Daley (1996)) is a multivariate statistical method that is computationally efficient and is applicable to atmospheric or numerical forecast models. In optimal interpolation, observations are used together with the simulation results of a model in order to produce an improved estimate (this is referred to as *analysis*) of the true state of the atmosphere.

The analysis, or analyzed state vector, $\mathbf{x}_\mathrm{a}(k)$, is given by the best estimate (linear and unbiased) in the least-squares sense and is the solution of the following optimization problem:

$$\mathbf{x}_\mathrm{a}(k) = \operatorname{argmin}\, J\left(\mathbf{x}(k)\right) \tag{2}$$

where $J\left(\mathbf{x}(k)\right)$ is the cost function to be minimized

$$\begin{aligned} J\left(\mathbf{x}(k)\right) = &\left(\mathbf{x}(k) - \mathbf{x}_\mathrm{b}(k)\right)^T \mathbf{B}^{-1} \left(\mathbf{x}(k) - \mathbf{x}_\mathrm{b}(k)\right) + \\ &+ \left(\mathbf{y}(k) - \mathbf{H}\mathbf{x}(k)\right)^T \mathbf{R}^{-1} \left(\mathbf{y}(k) - \mathbf{H}\mathbf{x}(k)\right), \end{aligned} \tag{3}$$

$\mathbf{x}_\mathrm{b}(k)$ is the background state vector (i.e., the model estimates), $\mathbf{y}(k)$ is the vector that contains the measured concentrations at time k, $\mathbf{H}$ is a linear observation operator that maps some of the elements of the state vector $\mathbf{x}(k)$ to the ground-based measurements available at time k, $\mathbf{B}$ is the background error covariance matrix (it accounts for modeling errors) and $\mathbf{R}$ is the covariance matrix that accounts for both the measurement and representativity errors of the observations.

The solution of the optimization problem (2) is given by

$$\mathbf{x}_\mathrm{a}(k) = \mathbf{x}_\mathrm{b}(k) + \mathbf{K}\left(\mathbf{y}(k) - \mathbf{H}\mathbf{x}_\mathrm{b}(k)\right) \tag{4}$$

where $\mathbf{K}$ is the so-called Kalman gain defined as

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T \left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{5}$$

Notice that unlike other approaches (e.g., the Ensemble Kalman Filter (Evensen (1994))) where the background error covariance matrix $\mathbf{B}$ is estimated at every assimilation step, in the optimal interpolation technique the entries of this matrix do not change in time and are approximated by using an empirical covariance function, which makes OI computationally much cheaper. However the characterization of this matrix is a task that is far from trivial.

The OI algorithm can be summarized as follows:

- Set the initial conditions ($\mathbf{x}_\mathrm{a}(k=0)$)

for $k = 1, \ldots$

Forecast step

- Compute the model estimate or forecast $\mathbf{x}_\mathrm{b}(k)$ by evaluating $\mathbf{x}_\mathrm{a}(k-1)$ in (1),
$$\mathbf{x}_\mathrm{b}(k) = \mathbf{M}_\mathrm{A}\left(\mathbf{x}_\mathrm{a}(k-1)\right)$$

Analysis step

- Calculate the state estimation (analysis) by using (4),
$$\mathbf{x}_\mathrm{a}(k) = \mathbf{x}_\mathrm{b}(k) + \mathbf{K}\left(\mathbf{y}(k) - \mathbf{H}\mathbf{x}_\mathrm{b}(k)\right)$$
where
$$\mathbf{K} = \mathbf{B}\mathbf{H}^T \left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

end for

3.1 Estimation of the error covariance matrices

In this study, the error covariance matrices $\mathbf{B}$ and $\mathbf{R}$ have been determined using the Hollingsworth-Lönnberg technique (Daley (1996); Hollingsworth and Lönnberg (1986)). Let $\mathbf{y}_i^\mathrm{M}$ and $\mathbf{y}_i^\mathrm{O}$ be vector time series of model simulated values and observations respectively, of a given pollutant at a station i out of N_s monitoring stations. Furthermore, let $\bar{y}_i^\mathrm{M}$ and $\bar{y}_i^\mathrm{O}$ be the average of $\mathbf{y}_i^\mathrm{M}$ and $\mathbf{y}_i^\mathrm{O}$ respectively. Based on the explanation given in Tilmes (2001) and Kumar et al. (2012), the method can be summarized as follows:

- Modify the model estimations such that $\bar{y}_i^\mathrm{M} = \bar{y}_i^\mathrm{O}$,
$$\tilde{\mathbf{y}}_i^\mathrm{M} = \mathbf{y}_i^\mathrm{M} + \mathbf{1} \cdot b_i, \quad \text{for } i = 1, \ldots, N_\mathrm{s}.$$
Here, $\tilde{\mathbf{y}}_i^\mathrm{M}$ is the unbiased output of the model, $b_i = \bar{y}_i^\mathrm{O} - \bar{y}_i^\mathrm{M}$ is the model bias and $\mathbf{1}$ is a vector of 1's of appropriate dimension.

- At every station, compute the error sequence between the unbiased model estimations and the measurements,
$$\mathbf{e}_i = \mathbf{y}_i^\mathrm{O} - \tilde{\mathbf{y}}_i^\mathrm{M}, \quad \text{for } i = 1, \ldots, N_\mathrm{s}.$$

- Calculate the correlations between the error sequences found in the previous step.

- Plot these correlations as a function of the distance r between their related stations, and fit a curve $R(r)$ through these points.

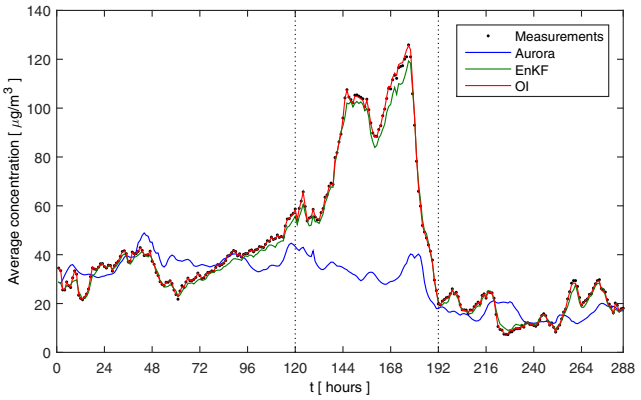

Fig. 3. Average of the PM_{10} concentration over the assimilation stations. Starting date: January 20th, 2010 at midnight. The dotted vertical lines delimit the period of the PM_{10} episode.

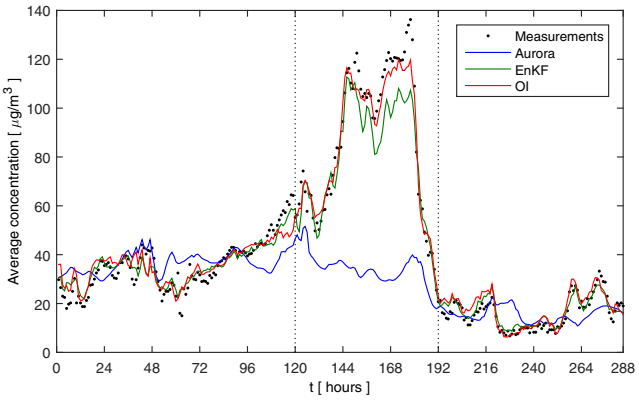

Fig. 4. Average of the PM_{10} concentration over the validation stations. Starting date: January 20th, 2010 at midnight. The dotted vertical lines delimit the period of the PM_{10} episode.

- Extrapolate $R(r)$ to zero distance, that is, find the intercept with the y-axis $R_z = \lim_{r\to 0} R(r)$.

- Find E_o^2 and E_m^2, the variances in the observation and background(model) errors, from the following equations:

$$R_z = \frac{E_m^2}{E_o^2 + E_m^2}, \qquad \frac{1}{N_s}\sum_{i=1}^{N_s} \overline{\left(\mathbf{y}_i^O - \tilde{\mathbf{y}}_i^M\right)^2} = E_o^2 + E_m^2, \tag{6}$$

where the overbar denotes the average and $(\cdot)^2$ should be interpreted as an element-wise exponentiation.

- Compute the error covariance matrices $\mathbf{R}$ and $\mathbf{B}$,

$$\mathbf{R} = E_o^2 \cdot \mathbf{I}, \qquad \mathbf{B} = E_m^2 \cdot \mathbf{\Gamma} \tag{7}$$

where $\mathbf{I}$ is the identity matrix, and $\mathbf{\Gamma}$ is a matrix whose entries are the correlations between the grid-cells of the model. For two grid-cells located in two different layers l_a and l_b, their correlation would be given by

$$f\left(r, l_a, l_b\right) = \frac{1}{R_z} R(r) \cdot \mathbf{S}_{l_a, l_b} \tag{8}$$

where r is the horizontal distance between the grid-cells, $\mathbf{S}$ is a symmetric positive definite matrix containing the correlations between the different layers, and $\mathbf{S}_{l_a,l_b}$ is the entry of the matrix $\mathbf{S}$ associated with the l_ath row and the l_bth column. Regarding the correlation matrix $\mathbf{S}$, this can be estimated in a straightforward way from a free-run of the model for the period of interest.

4. ASSIMILATION EXPERIMENTS

As it was mentioned in Section 2, a 12 day period starting on January 20th, 2010 at midnight was selected to carry out the assimilation experiments. These experiments were performed on a single PC with a 3 GHz Intel Core 2 Quad processor and 4 GB of RAM memory. In this study we have used the Root Mean Square Error (RMSE), the correlation coefficient and the mean bias as statistical measures for evaluating the performance of the optimal interpolation method. These statistical measures or indicators are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N_s N_h}\sum_{i=1}^{N_s}\sum_{k=1}^{N_h}\left(y_i^M(k) - y_i^O(k)\right)^2} \tag{9}$$

$$\text{Correlation} = \frac{\sum_{i=1}^{N_s}\sum_{k=1}^{N_h}\left(y_i^M(k) - \bar{y}^M\right)\left(y_i^O(k) - \bar{y}^O\right)}{\sqrt{S^M \times S^O}} \tag{10}$$

$$\text{Mean bias} = \frac{1}{N_s N_h}\sum_{i=1}^{N_s}\sum_{k=1}^{N_h}\left(y_i^M(k) - y_i^O(k)\right) \tag{11}$$

with

$$S^M = \sum_{i=1}^{N_s}\sum_{k=1}^{N_h}\left(y_i^M(k) - \bar{y}^M\right)^2,\ S^O = \sum_{i=1}^{N_s}\sum_{k=1}^{N_h}\left(y_i^O(k) - \bar{y}^O\right)^2,$$

$$\bar{y}^M = \frac{1}{N_s N_h}\sum_{i=1}^{N_s}\sum_{k=1}^{N_h} y_i^M(k), \quad \bar{y}^O = \frac{1}{N_s N_h}\sum_{i=1}^{N_s}\sum_{k=1}^{N_h} y_i^O(k),$$

where N_s is the number of stations, $y_i^M(k)$ and $y_i^O(k)$ are the estimated and measured PM_{10} concentrations respectively in the ith station at time k, and N_h is the number of simulated hours. For the specific case of a single air quality station, the RMSE is given by:

$$\text{RMSE}_{\text{station}} = \sqrt{\frac{1}{N_h}\sum_{k=1}^{N_h}\left(y^M(k) - y^O(k)\right)^2} \tag{12}$$

where $y^M(k)$ and $y^O(k)$ are the estimated and measured PM_{10} concentrations respectively in the station of interest at time k.

The error covariance matrices were estimated by applying the Hollingsworth-Lönnberg method (see Section 3.1) to the parts of the error sequences that do not cover the PM_{10} episode. For modeling the error correlations we fitted different functions and we selected the one that provided the best fit (smallest RMSE),

$$R(r) = 0.3724 \cdot e^{-0.02541 \cdot r} + 0.3891 \cdot e^{-0.001458 \cdot r} \tag{13}$$

for which the intercept with the y-axis is $R_z = 0.7615$. From (6), we found that the variance of the PM_{10} obser-

vation errors E_o^2, and the variance of the model (background) errors E_m^2, are 29.35 $\mu g^2/m^6$ and 93.71 $\mu g^2/m^6$ respectively.

In the OI scheme, the forecast of PM_{10} is calculated by summing the concentrations of all the species that form it (see Section 2). Once the measurements are assimilated, it is necessary to redistribute the analyzed PM_{10} over the model variables (keep in mind that in AURORA all the species that form PM_{10} are modeled separately). This redistribution was done as follows:

$$\left(P_i^j\right)^a = \frac{(PM_{10})_i^a}{(PM_{10})_i^f} \times \left(P_i^j\right)^f \quad (14)$$

where $\left(P_i^j\right)^f$ and $\left(P_i^j\right)^a$ are the concentrations of the chemical species j at the ith grid-cell for the forecast and analysis respectively, and $(PM_{10})_i^a$ and $(PM_{10})_i^f$ are the PM_{10} concentrations at the ith grid-cell for the forecast and analysis respectively. This implies that we assume that the relative chemical composition is well represented in AURORA. It would have been better if observed values for individual PM component concentrations could have been used in the data assimilation process but these are unfortunately lacking.

In Table 1 we can see the statistical indicators for the free AURORA run and the results of using the optimal interpolation technique. This table also presents the results of applying a well-known algorithm such as the Ensemble Kalman Filter (EnKF), which was configured to work with 60 ensembles using a similar setting to the one presented in Agudelo et al. (2011). It is clear that a large error reduction is obtained by using the EnKF and OI algorithms. For the case of the EnKF, the error of the PM_{10} estimates is decreased by 76.37% in the assimilation stations and by 60.85% in the validation stations. For the OI technique the error in the assimilation and validation stations is reduced by 79.31% and 61.92% respectively. Furthermore, from this table we can see that the assimilation of PM_{10} observations leads to a notably error reduction in the $PM_{2.5}$ estimations. Compared to the free-run of AURORA, the data assimilation algorithms increase the correlation coefficient by a factor not smaller than 2 and significantly decrease the magnitude of the mean bias for all the cases.

Figs. 3 and 4 present the average of the PM_{10} concentration over the assimilation and validation stations respectively. Notice that the free-run of AURORA is not able to properly follow the data. In spite of the big gap between AURORA and the observations, the EnKF and the optimal interpolation technique manage to notably reduce this difference.

The impact of the assimilation process on each PM_{10} validation station (keep in mind that the measurements coming from these stations were not used by the assimilation algorithms) and on each $PM_{2.5}$ station can be evaluated from Figs. 5 and 6 respectively. In all the cases the data assimilation techniques lead to a significant reduction of the estimation errors.

Fig. 7 shows the daily average PM_{10} concentration of Belgium from January 24th until January 28th, 2010. Results are shown for both the free run of AURORA and the run corrected using data assimilation as well as an interpolated map of the observed values which was obtained using the RIO interpolation method (Janssen et al. (2008)). The interpolated map was taken from the IRCEL website (www.irceline.be). Notice that in general, the spatial distributions of the concentrations obtained with the optimal interpolation technique are more or less in line with the ones published by IRCEL. According to expert judgment, the PM_{10} episode described in this paper was due to the import of PM originating in Eastern Europe.

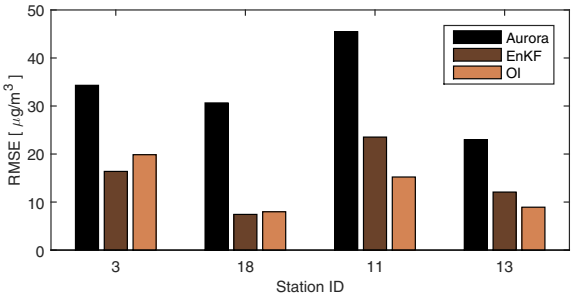

Fig. 5. RMSE of AURORA and the data assimilation techniques at the PM_{10} validation stations.

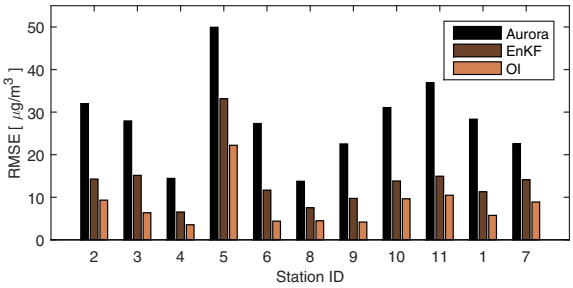

Fig. 6. RMSE of AURORA and the data assimilation techniques at the $PM_{2.5}$ stations.

5. CONCLUSIONS

In this paper we have presented the results of using a data assimilation technique known as Optimal Interpolation, for improving the quality of the PM_{10} estimates of the air quality model AURORA. The characterization of the background error covariance matrix $\mathbf{B}$ and the measurement error covariance matrix $\mathbf{R}$, has been done by means of the Hollingsworth-Lönnberg method. The assimilation process has been driven by ground-based measurements provided by IRCEL. In spite of the big gap between the AURORA estimates and the measurements, the optimal interpolation technique managed to significantly reduce the error in the PM_{10} estimates and to improve the spatial distribution of the PM_{10} concentrations. Given that $PM_{2.5}$ is highly correlated with PM_{10}, the $PM_{2.5}$ field was indirectly improved by this data assimilation technique.

In this paper, we have also compared the OI technique with a well-known data assimilation technique, the Ensemble Kalman Filter. Although much simpler and less computationally demanding than the EnKF algorithm, the OI technique was the method that in general led to the largest error reduction in the scenario treated in this paper.

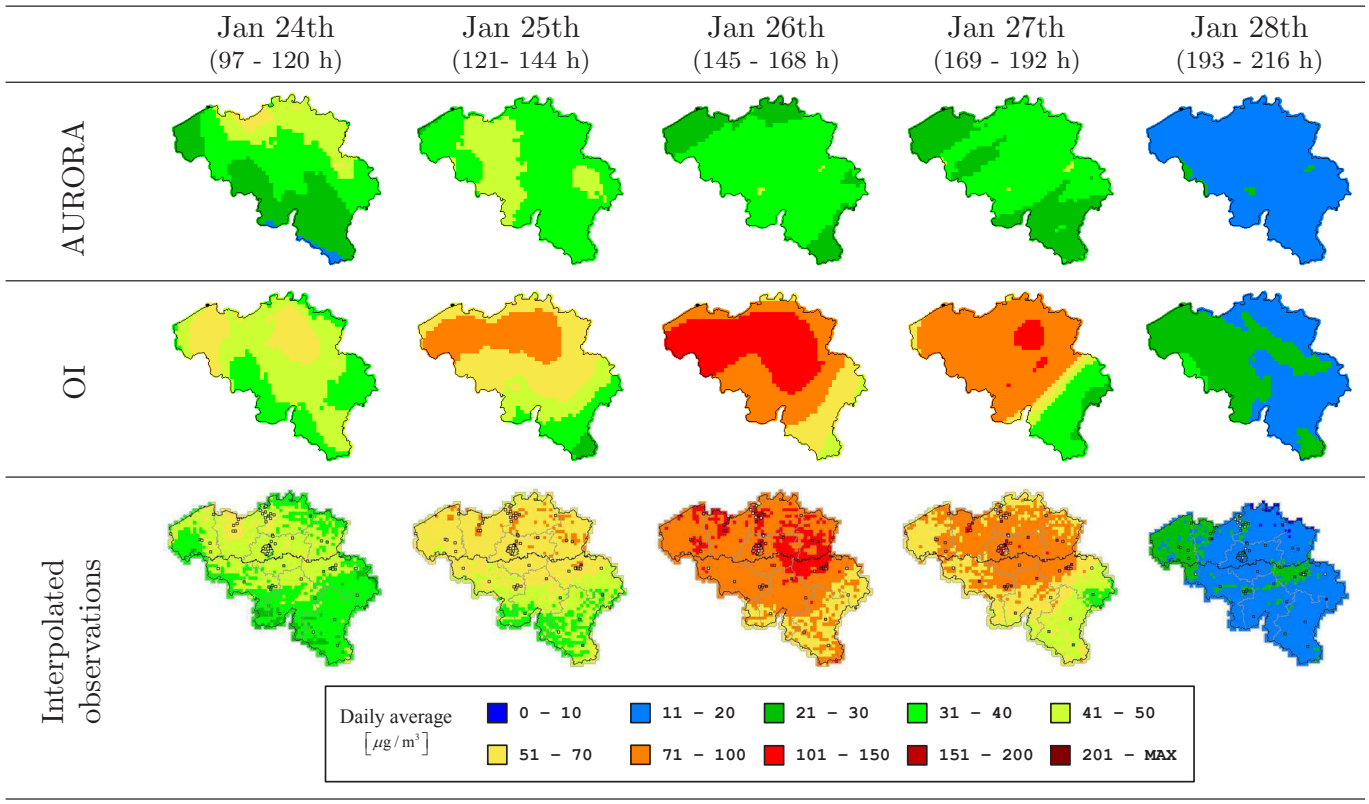

Fig. 7. Daily average PM_{10} concentration for January 24th -January 28th, 2010. The maps at the bottom of the figure were taken from the IRCEL website (www.irceline.be).

ACKNOWLEDGEMENTS

This work has been developed in the frame of the project IWT SBO Climaqs (Climate and Air Quality Modeling for Policy Support).

REFERENCES

Agudelo, O.M., Barrero, O., Viaene, P., and De Moor, B. (2011). Assimilation of ozone measurements in the air quality model AURORA by using the Ensemble Kalman Filter. In *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC 2011)*, 4430–4435. Orlando, FL, USA.

Candiani, G., Carnevale, C., Filisina, V., Finzi, G., Pisoni, E., and Volta, M. (2009). Optimal interpolation to re-analyse PM10 concentration modelling simulations. In *Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, 1794–1799. Shangai, P.R. China.

Daley, R. (1996). *Atmospheric Data Analysis*. Cambridge University Press.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143–10162.

Hollingsworth, A. and Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. part i: The wind field. *Tellus*, 38A, 111–136.

IRCEL (2011). Intergewestelijke cel voor het leefmilieu (Belgian interregional environment agency). [online] `http://www.irceline.be/`.

Janssen, S., Dumont, G., Fierens, F., and Mensink, C. (2008). Spatial interpolation of air pollution measurements using corine land cover data. *Atmospheric Environment*, 42(20), 4884 – 4903. doi: 10.1016/j.atmosenv.2008.02.043.

Kumar, U., De Ridder, K., Lefebvre, W., and Janssen, S. (2012). Data assimilation of surface air pollutants (O_3 and NO_2) in the regional-scale air quality model AURORA. *Atmospheric Environment*, 60, 99–108.

Mensink, C., De Ridder, K., Lewyckyj, N., Delobbe, L., Janssen, L., and Van Haver, P. (2001). Computational aspects of air quality modelling in urban regions using an optimal resolution approach (AURORA). In S. Margenov, J. Wasniewski, and P. Yamalov (eds.), *Large-Scale Scientific Computing, Lecture Notes in Computer Science*, volume 2179, 299–308. Springer Berlin / Heidelberg.

Nenes, A., Pandis, S.N., and Pilinis, C. (1998). ISORROPIA: A new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols. *Aquatic Geochemistry*, 4(1), 123–152.

Pope III, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., and Thurston, G.D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *J. Am. Med. Assoc. (JAMA)*, 287(9), 1132–1141.

Tilmes, S. (2001). Quantitative estimation of surface ozone observation and forecast errors. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 26, 759–762.

Tombette, M., Mallet, V., and Sportisse, B. (2009). PM_{10} data assimilation over Europe with the optimal interpolation method. *Atmos. Chem. Phys.*, 9, 57–70.

Vautard, R., Moran, M.D., Solazzo, E., Gilliam, R.C., Matthias, V., Bianconi, R., Chemel, C., Ferreira, J., Geyer, B., Hansen, A.B., Jericevic, A., Prank, M., Segers, A., Silver, J.D., Werhahn, J., Wolke, R., Rao, S., and Galmarini, S. (2012). Evaluation of the meteorological forcing used for the air quality model evaluation international initiative (AQMEII) air quality simulations. *Atmospheric Environment*, 53(0), 15 – 37.

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B. (2008). A comparison study of data assimilation algorithms for ozone forecasts. *Journal of geophysical research*, 113, D20310. doi: 10.1029/2008JD009991.