2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Assimilation of ozone measurements in the air quality model AURORA by using the Ensemble Kalman Filter

Oscar Mauricio Agudelo, Oscar Barrero, Viaene Peter
and Bart De Moor, *Fellow Member, IEEE*

Abstract— **This paper presents the results of using the Ensemble Kalman Filter (EnKF) for improving the ozone estimations of the air quality model AURORA. The EnKF is built around a stochastic formulation of the model, where some of its parameters are assumed to be uncertain. These uncertainties turn out to be the main reason behind the differences between the model predictions and the real measurements. The filter estimates these parameters as well as the ozone concentration field by using ground-based measurements from the Airbase database. The assimilation experiments are carried out over a region that consists of Belgium, Luxembourg, and some small parts of Germany, France and the Netherlands. The simulations results show that the EnKF significantly reduces the error of the ozone estimations.**

I. INTRODUCTION

Data assimilation is the common name given to several numerical techniques that combine the outputs of a numerical model with observational data in order to improve the quality of the model predictions. In data assimilation it is assumed that both the model and the measurements are subject to errors. These errors or uncertainties are defined in statistical terms, and their specification plays a very important role in the success or failure of the data assimilation.

Among the existing data assimilation techniques, the Ensemble Kalman Filter (EnKF) has gained a lot of popularity since its introduction by Evensen in 1994 [1], given its capability of handling nonlinear large-scale systems and its relatively easy implementation compared to other approaches where it is required to linearize the model and/or solve an optimization problem. The Ensemble Kalman Filter is a sequential data assimilation technique that uses Monte Carlo or ensemble integration. By integrating an ensemble of model states forward in time it is possible to calculate statistical moments like mean and error covariances whenever such information is required. Thus, all the statistical information about the predicted model state is contained in the ensemble [2].

The EnKF has been successfully used in meteorology [3], oceanography [4] and space weather forecast [5] [6]. In the air quality field, this technique has been applied for example to the Long Term Ozone Simulation (LOTUS) model [7] [8], to the regional chemistry transport model LOTOS-EUROS [9] [10] and to the TCAM (Transport Chemical Aerosol Model) model [11]. In all the cases the technique has led to better estimations of the pollutants under consideration.

Accurate estimation of tropospheric ozone is of great interest given its harmful effects on human health and on natural ecosystems. However this is not an easy task given that ozone formation and accumulation are nonlinear processes that depend on several chemical reactions. In addition, uncertainty in the boundary conditions, meteorological patterns, and emissions of its precursors (nitrogen oxides and volatile organic compounds) among other factors, make this task even harder. Here is where data assimilation techniques like the EnKF play a very important role.

In this paper, we apply the EnKF technique to the air quality model AURORA [12] in order to improve the ozone estimations over a prescribed spatial domain. The approach proposed not only estimates the ozone field but also some uncertain correction factors.

This paper is organized as follows. Section II presents a description of AURORA as well as the observation network used for assimilation and validation purposes. In Section III, a stochastic model of AURORA is derived and the EnKF algorithm is described. Section IV presents the details of the assimilation experiments as well as the results obtained. Finally, in Section V we present some concluding remarks and future research directions.

II. DESCRIPTION OF THE AURORA MODEL

AURORA (Air quality modelling 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 [12]. The model contains 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 such as ozone and (iii) the loss of these components by deposition.

O. M. Agudelo is with the Department of Electrical Engineering (ESAT), Research Group SCD-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Heverlee, Belgium. He is also with the Department of Automation and Electronics, Universidad Autónoma de Occidente, Calle 25 No 115-85, Cali, Colombia. E-mail: mauricio.agudelo@esat.kuleuven.be

O. Barrero is with the Department of Industrial Automation, Universidad de Ibagué, Carrera 22 Calle 67 B/Ambala, Ibagué, Colombia. E-mail: oscar.barrero@unibague.edu.co

Viaene Peter is with the Environmental Modelling Unit, VITO - Flemish Institute for Technological Research, Roderveldlaan 5, BE-2600 Berchem, Belgium. E-mail: peter.viaene@vito.be

B. De Moor is with the Department of Electrical Engineering (ESAT), Research Group SCD-SISTA, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Heverlee, Belgium. E-mail: bart.demoor@esat.kuleuven.be

978-1-61284-799-3/11/$26.00 ©2011 IEEE

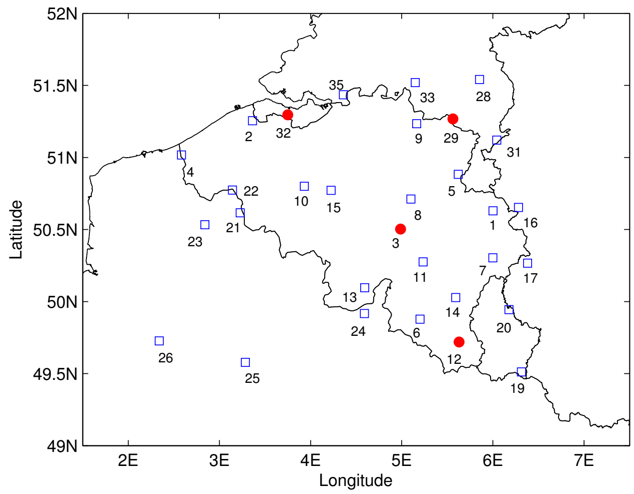

Fig. 1. Geographical location of the stations for ozone. The squares and circles represent assimilation and validation stations respectively.

In this work AURORA was set up for a domain covering Belgium, Luxemburg, and portions of Germany, France and the Netherlands (see Figure 1). The horizontal domain was divided into 11 x 15 grid-cells with a resolution of 25 km. For the vertical domain, 15 layers were used to span an altitude of 2000 m. The model was configured to simulate a two week period starting at May 28th, 2005. During this period elevated ozone concentrations were observed. The relatively small number of grid-cells, allows us to significantly reduce the computational burden during the simulation of the model, and therefore it makes easier testing different settings for the EnKF.

The air quality data for performing the assimilation experiments are obtained from the Airbase database [13]. Only measurements from rural-background stations are taken into account due to the spatial resolution of 25 km of the AURORA model setup. Figure 1 shows the location of the 31 air quality stations used along this study. From this figure we can see two groups of them, namely, the assimilation and 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 size of the assimilation and validation sets is 27 and 4 respectively. The validation set has been chosen such that the validation stations are properly distributed along the spatial domain and surrounded by nearby assimilation stations.

III. Sequential Data Assimilation Scheme

The deterministic model of AURORA can be compactly written in the following way,

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

where $\mathbf{c}(k)$ is the state vector containing the concentrations of the considered pollutants $(\mathrm{O_3}, \mathrm{NO_x}, \mathrm{SO_2}, \mathrm{NH_3}, \mathrm{PM_{10}},$ etc.) at every grid-cell, $\mathbf{u}(k)$ is the vector that comprises the inputs of the model such as the boundary conditions, the emissions of several components ($\mathrm{NO_x}$, $\mathrm{SO_2}$, ETHE, PAR, OLE, HCHO, etc.) and the meteorological conditions, and $\mathbf{M}_{\mathrm{A}}(\cdot)$ is the nonlinear state-space operator which computes the concentrations at time $k+1$ from the concentrations given at time k.

In order to apply the EnKF algorithm to AURORA, it is required first to define a stochastic model of the system that accounts for the model and measurement errors. It is well known that the knowledge of the uncertainties both in the model and in the measurements is crucial for a successful data assimilation. In the next section, we discuss the derivation of the stochastic representation of AURORA and afterwards we describe the EnKF algorithm.

A. Stochastic state-space representation of Aurora

In AURORA, process inputs such as boundary conditions and emissions, and model parameters such as deposition and cloudiness, are multiplied by correction factors that in principle can be estimated by the EnKF in order to reduce the differences between the predictions made by the model and the observations. The default value for these factors is 1, and they must be nonnegative. Since the idea of using these factors is to somehow account for the uncertainty of the boundary conditions, emissions, deposition, etc., each of them is modelled as follows:

$$f(k) = \max\left(0, 1 + \lambda(k)\right) \tag{2}$$

where $f(k)$ is the correction factor and $\lambda(k)$ is a colored noise process which has the following equation in scalar form:

$$\lambda(k+1) = \alpha\lambda(k) + \left(\sigma\sqrt{1-\alpha^2}\right) w(k), \tag{3}$$
$$w(k) \sim N(0,1).$$

Here, $\alpha \in [0\ 1]$ represents the time correlation parameter. If α is set to zero, then we obtain a white noise sequence with zero mean and standard deviation σ. When α is set to one, the colored noise process is reduced to a constant value. In order to ensure that a large number of samples of $\lambda(k)$ maintain the standard deviation σ, the initial sample, e.g. $\lambda(0)$, should be random distributed with the desired statistics. The temporal covariance $\mathrm{E}\left(\lambda(k+l)\lambda(k)\right)$ of the colored noise is equal to α^l. $\lambda(k)$ is a stationary Gaussian process with an exponential covariance function that is parameterized by $\alpha = e^{(-T_s/\tau)}$ [9], where τ is the time correlation length and T_s is the time between two consecutive samples, which in our case is equal to 1 hour. The fact of using colored noise for modeling the factors avoid the introduction of rapid fluctuations that occur when only white noise is used.

Similarly as it was done in [9] and [7], the stochastic model of AURORA is built by augmenting the state vector of the model (1) with the vector $\boldsymbol{\lambda}(k) = [\lambda_1(k), \lambda_2(k), \ldots, \lambda_n(k)]^T$ comprising the colored noise process of every uncertain factor. The stochastic model is

TABLE I

PERFORMANCE OF THE ENKF FOR DIFFERENT COMBINATIONS OF UNCERTAIN FACTORS WHEN OZONE MEASUREMENTS ARE ASSIMILATED

		Assimilation Stations	Validation Stations
RMSE of AURORA		29.1223 $\mu g/m^3$	27.472 $\mu g/m^3$
EnKF Case A	RMSE	15.9235 $\mu g/m^3$	21.404 $\mu g/m^3$
	Error reduction	45.32 %	22.09 %
EnKF Case B	RMSE	15.8485 $\mu g/m^3$	19.6083 $\mu g/m^3$
	Error reduction	45.58 %	28.62 %
EnKF Case C	RMSE	13.8361 $\mu g/m^3$	18.9379 $\mu g/m^3$
	Error reduction	52.49 %	31.06 %
EnKF Case D	RMSE	12.6167 $\mu g/m^3$	17.7398 $\mu g/m^3$
	Error reduction	56.68 %	35.43 %

RMSE calculated by using Equation (13)

then given by

$$\begin{bmatrix} \mathbf{c}(k+1) \\ \boldsymbol{\lambda}(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{\mathrm{A}}\left(\mathbf{c}(k), \boldsymbol{\lambda}(k), \mathbf{u}(k)\right) \\ \boldsymbol{\Psi}\boldsymbol{\lambda}(k) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma} \end{bmatrix} \mathbf{w}(k) \tag{4}$$

with

$$\begin{aligned} &\mathbf{w}(k) \in \mathbb{R}^n \sim N(\mathbf{0}, \mathbf{I}_n), \\ &\boldsymbol{\Psi} \in \mathbb{R}^{n \times n} = \mathrm{diag}\left([\alpha_1, \alpha_2, \ldots, \alpha_n]\right), \\ &\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n} = \mathrm{diag}\Big(\Big[\sigma_1\sqrt{1-\alpha_1^2}, \sigma_2\sqrt{1-\alpha_2^2}, \ldots \\ &\qquad\qquad \ldots, \sigma_n\sqrt{1-\alpha_n^2}\Big]\Big), \end{aligned}$$

where n is the number of uncertain factors, $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\sigma_1, \sigma_2, \ldots, \sigma_n$ are the time correlation parameters and the standard deviations of the colored noise processes associated to the factors.

If we define $\mathbf{x}(k) = \left[\mathbf{c}(k)^T, \boldsymbol{\lambda}(k)^T\right]^T$ as the new augmented state vector, Equation (4) can be written more compactly as

$$\mathbf{x}(k+1) = \mathbf{M}\left(\mathbf{x}(k), \mathbf{u}(k)\right) + \mathbf{G}\mathbf{w}(k). \tag{5}$$

The nonlinear operator $\mathbf{M}$ describes the time evolution of augmented state vector $\mathbf{x}(k)$ from time k to $k+1$, and $\mathbf{G}\mathbf{w}$(k) is the stochastic forcing term. The relation between the model state $\mathbf{x}(k)$ and the Airbase observations $\mathbf{y}(k)$ is explained by means of the following expression

$$\mathbf{y}(k) = \mathbf{C}(k)\mathbf{x}(k) + \mathbf{v}(k), \quad \mathbf{v}(k) \sim N\left(\mathbf{0}, \mathbf{R}(k)\right) \tag{6}$$

where $\mathbf{C}(k)$ is a linear observation operator that assigns the concentrations of some grid-cells (where the stations are located) at the bottom layer of AURORA to the corresponding measurements from Airbase, and $\mathbf{v}(k)$ is a Gaussian white noise vector with covariance $\mathbf{R}(k)$ that accounts for the uncertainties in the Airbase measurements. Given that the stations are not providing valid measurements all the time, $\mathbf{C}(k)$ and $\mathbf{v}(k)$ must be properly adjusted within the assimilation algorithm in order to cope with each particular case.

B. Ensemble Kalman Filter

The Ensemble Kalman Filter (EnKF) is a Monte Carlo method that is based on the representation of the probability density of the state estimate by a finite ensemble of N possible states, $\mathbf{x}_1(k), \mathbf{x}_2(k), \ldots, \mathbf{x}_N(k)$. This characteristic allow the filter to be applied to large-scale systems where the explicit storage and manipulation of the state error covariance are impossible or impractical. In the EnKF formulation, it is assumed that every ensemble member is a single sample taken from the probability distribution of the true state. Statistical information such as mean and error covariances can be approximated with sample statistics whenever such information is required.

The EnKF algorithm for the model (5)-(6) can be summarized as follows [2]:

1) Initialization
   First, generate an initial ensemble $\mathbf{x}_1^{\mathrm{a}}(k-1), \mathbf{x}_2^{\mathrm{a}}(k-1), \ldots, \mathbf{x}_N^{\mathrm{a}}(k-1)$ that properly represent the error statistics of the initial guess $\mathbf{x}(k=0)$ for the model state.

2) for $k = 1, \ldots$
   a) Forecast Step
      - Update every ensemble member using (5),
      $$\begin{gathered} \mathbf{x}_i^{\mathrm{f}}(k) = \mathbf{M}\left(\mathbf{x}_i^{\mathrm{a}}(k-1), \mathbf{u}(k-1)\right) + \mathbf{G}\mathbf{w}_i(k-1) \\ \mathbf{w}_i(k-1) \sim N(\mathbf{0}, \mathbf{I}), \quad \forall i = 1, 2, \ldots, N \end{gathered} \tag{7}$$
      - Calculate the forecast ensemble mean,
      $$\bar{\mathbf{x}}^{\mathrm{f}}(k) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i^{\mathrm{f}}(k) \tag{8}$$
      - Compute the forecast error covariance matrix
      $$\mathbf{P}^{\mathrm{f}}(k) = \frac{1}{N-1}\sum_{i=1}^{N} \boldsymbol{\xi}_i(k)\boldsymbol{\xi}_i(k)^T \tag{9}$$
      $$\boldsymbol{\xi}_i(k) = \mathbf{x}_i^{\mathrm{f}}(k) - \bar{\mathbf{x}}^{\mathrm{f}}(k)$$
   b) Analisys Step
      - Compute the Kalman Gain
      $$\mathbf{K}(k) = \mathbf{P}^{\mathrm{f}}(k)\mathbf{C}(k)^T\left(\mathbf{C}(k)\mathbf{P}^{\mathrm{f}}(k)\mathbf{C}(k)^T + \mathbf{R}(k)\right)^{-1} \tag{10}$$
      - Update $\mathbf{x}_i^{\mathrm{f}}(k)$ to $\mathbf{x}_i^{\mathrm{a}}(k)$, for $i = 1, \ldots, N$
      $$\begin{gathered} \mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{f}} + \mathbf{K}(k)\left(\mathbf{y}(k) + \mathbf{v}_i(k) - \mathbf{C}(k)\mathbf{x}_i^{\mathrm{f}}(k)\right) \\ \mathbf{v}_i(k) \sim N\left(\mathbf{0}, \mathbf{R}(k)\right) \end{gathered} \tag{11}$$
      - Calculate the state estimation (the analysis ensemble mean)
      $$\hat{\mathbf{x}}(k) = \bar{\mathbf{x}}^{\mathrm{a}}(k) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i^{\mathrm{a}}(k) \tag{12}$$

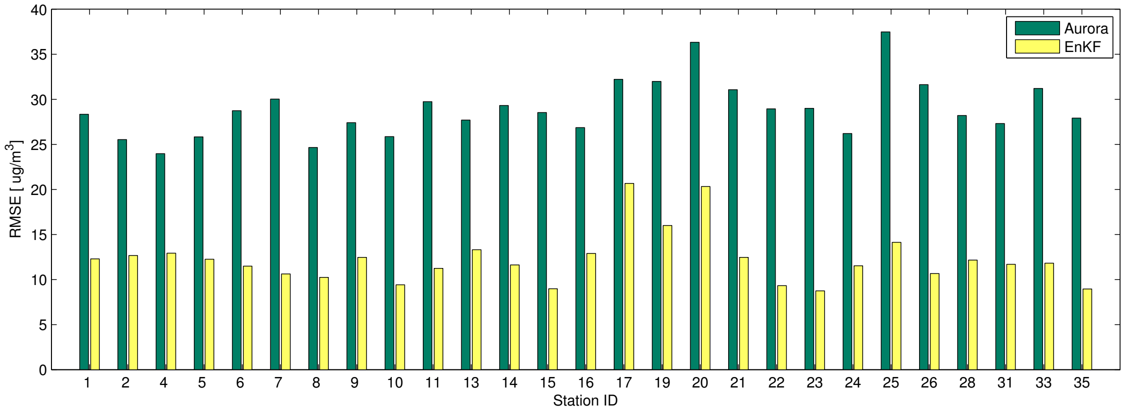

Fig. 2. Root mean square error of AURORA and the EnKF (case D) at the ozone assimilation stations.

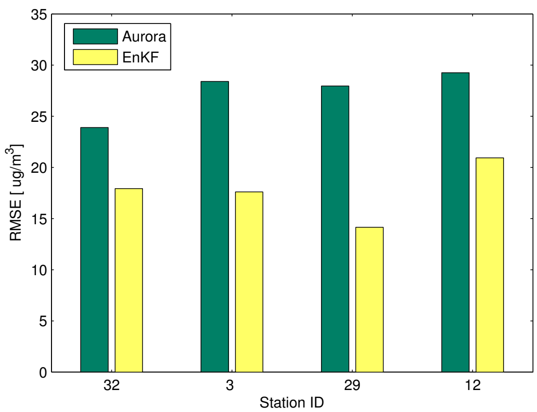

Fig. 3. Root mean square error of AURORA and the EnKF (case D) at the ozone validation stations.

where $\mathbf{y}(\mathrm{k})$ is a vector with the available measurements at time k. One of the advantages of this algorithm is that it is not required to linearize the model, since the ensembles are propagated by using the original nonlinear model operator. It is important to remark that in the final implementation of the EnKF, it is not necessary to explicitly compute the forecast error covariance matrix $\mathbf{P}^{\mathrm{f}}(k)$ for calculating the Kalman gain (e.g. we can compute the matrix product $\mathbf{P}^{\mathrm{f}}\mathbf{C}(k)^T$ in a recursive manner). This fact by the way, makes this algorithm appropriate for being used with large-scale systems.

For most practical applications, the most time consuming part of the EnKF algorithm is the evaluation of the ensembles in (7). Therefore, the computational load is approximately N model evaluations. The errors in the state estimate are of statistical nature and decrease slowly with the number of ensemble members. This diminution is proportional to $1/\sqrt{N}$ [6]. This is one of the very few drawbacks of this Monte Carlo approach.

IV. Assimilation results

As it was explained in Section III-A, the stochastic model of AURORA is constructed by considering some correction factors uncertain, and consequently the parameters and/or inputs that are multiplied by them. In this study, we have worked with the factors associated to the boundary conditions and deposition of the ozone, the cloudiness, and the emissions of nitrogen oxides ($\mathrm{NO_x}$) and Volatile Organic Compounds (VOCs). All of them play an important role in the formation and/or destruction of the ozone. In the current setup, the north, south, west and east boundary conditions are multiplied by the same correction factor. For the $\mathrm{NO_x}$ (NO and $\mathrm{NO_2}$) and VOCs (ETHE, PAR, OLE, XYL, HCHO, RCHO, TOL, CRES, and ISOP) emissions, a single factor is used in each case. After playing with different combinations of the aforementioned factors, we found that the most representative results were achieved with the following combinations:

- Case A
  - Correction factor 1: Boundary conditions of the ozone
- Case B
  - Correction factor 1: Boundary conditions of the ozone
  - Correction factor 2: Deposition of the ozone
- Case C
  - Correction factor 1: Boundary conditions of the ozone
  - Correction factor 2: Deposition of the ozone
  - Correction factor 3: Cloudiness
- Case D
  - Correction factor 1: Boundary conditions of the ozone
  - Correction factor 2: Deposition of the ozone
  - Correction factor 3: Cloudiness
  - Correction factor 4: $\mathrm{NO_x}$ emissions
  - Correction factor 5: VOCs emissions.

As it is stated in [7], the specification of the model error statistics, that is, the uncertainty in the colored-noise processes λ's, should be chosen on the basis of expert opinions. In our case, a standard deviation of 0.8 was set to the factors that multiply the boundary conditions of the ozone and the cloudiness parameter. For the factors associated to the deposition of the ozone and the $\mathrm{NO_x}$ and VOCs emissions, a standard deviation of 0.5 was considered. A time correlation length τ of 6 hours was used for all the factors. Now, for the ozone observations provided by Airbase, it was assumed an error of 7% of the measured concentration with a minimum value of $1\mu\mathrm{g/m^3}$.

A 404 hours period starting on May 28th, 2005 at midnight was chosen to carry out the assimilation experiments. This period was selected because during that time a serious ozone episode took place.

The assimilation experiments were carried out in a single workstation, a dual Opteron 250 with 4GB of RAM memory. The number of ensemble members has been set to 80, and in average every assimilation experiment took about two hours and a half.

Table I shows the global errors between the observations and the estimations made by both the model and the EnKF. The statistical measure used in this table for comparing the different estimations is the Root Mean Square Error (RMSE), which is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_\mathrm{s} N_\mathrm{h}} \sum_{i=1}^{N_\mathrm{s}} \sum_{k=1}^{N_\mathrm{h}} \left(y_i^\mathrm{M}(k) - y_i^\mathrm{O}(k) \right)^2} \tag{13}$$

where N_s and N_h are the number of stations and the number of hours respectively, and $y_i^\mathrm{M}(k)$ and $y_i^\mathrm{O}(k)$ are the estimated and measured concentrations of the ith station at time k.

Giving that AURORA is not covering an isolated region, the boundary conditions play a very important role in the quality of the model predictions. This can be corroborated by the results obtained in the case A (see Table I), where the error in the assimilation and validation stations is reduced by 45.32 % and 22.09 % respectively. Bear in mind that this significant improvement was achieved just by considereing uncertain the boundary conditions of the ozone. When the deposition of the ozone is also considered (case B), the error in the validation stations is smaller than in the case A, in spite of the fact that in both cases the error reduction in the assimilation stations is practically the same. From Table I it is clear that adding the cloudiness (case C) to the list of uncertain parameters improves the EnKF estimations for all the stations. However, the best results are obtained in the case D, where not only the boundary conditions, the deposition and the cloudiness are assumed to be uncertain but also the $\mathrm{NO_x}$ and VOCs emissions. In this case, the error is decreased by 56.68 % for the assimilation stations, while for the validation set an error reduction of 35.43 % is achieved.

In the remainder of this section we will present some of the results obtained in the case D. Figures 2 and 3 show the RMSE of every assimilation and validation station along the period of interest for AURORA and the EnKF. This RMSE

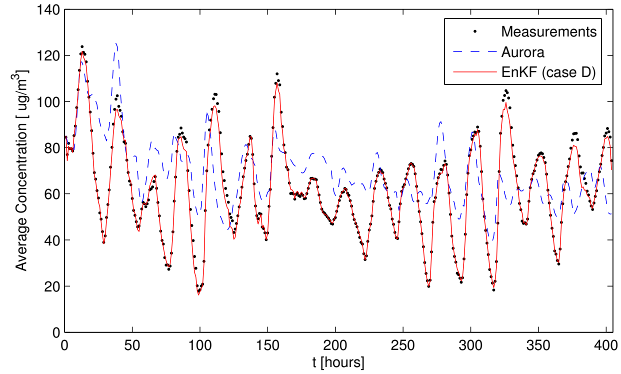

Fig. 4. Average of the ozone concentration over the assimilation stations. Starting date: May 28th, 2005.

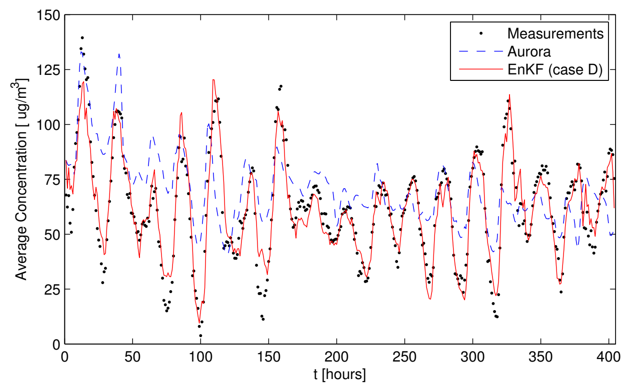

Fig. 5. Average of the ozone concentration over the validation stations. Starting date: May 28th, 2005.

has been computed in the following manner:

$$\mathrm{RMSE_{station}} = \sqrt{\frac{1}{N_\mathrm{h}} \sum_{k=1}^{N_\mathrm{h}} \left(y^\mathrm{M}(k) - y^\mathrm{O}(k) \right)^2} \tag{14}$$

where N_h is the number of hours, and $y_i^\mathrm{M}(k)$ and $y_i^\mathrm{O}(k)$ are the estimated and measured concentrations of a given station at time k.

As a result of using the EnKF, the errors in the assimilation stations have been remarkably decreased. The largest error reduction takes place in the station 23 (69.84 %), and the smallest one occurs in the station 17 (35.83 %). As it was mentioned in Section II, we use the validation stations for verifying the assimilation results and evaluating the impact of the data assimilation on the covered region. From Figure 3, it is clear that the error in every validation station has been lessened after using the EnKF. The largest (49.39 %) and the smallest (24.97 %) error reduction arise in the stations 29 and 32 respectively.

Figure 4 shows the average of the ozone concentration over the assimilation stations. Likewise in Figure 5, we can observe the time evolution of the average of the ozone concentration over the validation stations. From these figures, it is evident that the estimations made by the EnKF are notably more accurate than the ones made by AURORA. Keep in mind that for the case of the validation stations, no local measurements were used, and therefore the improvement observed is the consequence of the estimation of the correction factors by using observations from nearby stations.

Although in the current setup the EnKF increases the quality of the ozone estimations, this does not imply that for other pollutants the estimation error is decreased. In fact it might be increased. We have for example the NO_x case. It was observed that the estimation error was reduced at some monitoring stations, but increased at others. The EnKF corrects the NO_x emissions (case D) to improve the ozone estimations regardless the effect on the estimation of the NO_x field. A straightforward way of addressing this issue is by using not only ozone observations but also NO_x measurements. Nevertheless, in this case it would be expected to get a smaller improvement in the ozone estimations since the EnKF would have to deal with two closely related pollutants at the same time.

V. Conclusions and Future work

In this paper we have presented the results of applying a sequential data assimilation scheme, the EnKF, to the air quality model AURORA. Ground-based measurements from the Airbase database have been used to drive the assimilation process. In this study, the stochastic formulation of AURORA required by the EnKF was derived by considering uncertain some model parameters that have a key role in the formation/destruction of the ozone. It was observed that one of the most critical parameters is the factor associated to the boundary conditions. This is not a surprising result given that the region under consideration is not isolated and consequently the pollution of the adjacent regions have a significant impact on the local dynamics. In addition, the small size of the considered area and the coarse grid, raise the importance of the boundary conditions. The EnKF managed to significantly reduce the error in the ozone estimations. The best results were obtained when the correction factors associated to the boundary conditions and the deposition of the ozone, the cloudiness and the NO_x and VOCs emissions were estimated by the filter.

Future work will be focused on incorporating more pollutants in the current assimilation scheme, and on applying the EnKF algorithm to AURORA model setups with higher spatial resolution and for longer periods.

VI. ACKNOWLEDGMENTS

This work has been developed in the frame of the project IWT SBO Climaqs (Climate and Air Quality Modelling for Policy Support). In addition, this research has been supported by: • Research Council KUL: GOA/11/05 Ambiorics, GOA/10/09 MaNet , CoE EF/05/006 Optimization in Engineering (OPTEC) and PFV/10/002 (OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants; • Flemish Government: ◦ FWO: PhD/postdoc grants, projects: G0226.06 (cooperative systems and optimization), G0321.06 (Tensors), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine) research communities (WOG: ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC); ◦ IWT: PhD Grants, Eureka-Flite+, SBO LeCo-Pro, SBO POM, O&O-Dsquare; • Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011); • IBBT • EU: ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940) • Contract Research: AMINAL • Other: ◦ Helmholtz: viCERP; ◦ ACCM. Dr. Bart De Moor is a full professor at the Katholieke Universiteit Leuven, Belgium. The scientific responsibility is assumed by its authors.

References

[1] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, vol. 99(C5), pp. 10 143–10 162, 1994.

[2] G. Burgers, P. Jan van Leeuwen, and G. Evensen, "Analysis scheme in the Ensemble Kalman Filter," *Monthly Weather Review*, vol. 126, no. 6, pp. 1719–1724, 1998.

[3] P. L. Houtekamer and H. L. Mitchell, "A sequential Ensemble Kalman Filter for atmospheric data assimilation," *Monthly Weather Review*, vol. 129, pp. 123–137, 2001.

[4] G. Ueno, T. Higuchi, T. Kagimoto, and N. Hirose, "Application of the Ensemble Kalman Filter to atmosphere-ocean coupled model," in *Nonlinear Statistical Signal Processing Workshop, 2006 IEEE*, September 2006, pp. 125 –128.

[5] O. Barrero, B. De Moor, and D. Bernstein, "Data assimilation for magnetohydrodynamics systems," *Journal of Computational and Applied Mathematics*, vol. 189, pp. 242–259, 2006.

[6] O. Barrero, "Data assimilation in Magnetohydrodynamics systems using Kalman filtering," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, May 2005.

[7] A. Heemink and A. Segers, "Modelling and prediction of environmental data in space and time using Kalman filtering," *Stochastic Environmental Research and Risk Assessment*, vol. 16, no. 3, pp. 225–240, 2002.

[8] M. van Loon, P. Builtjes, and A. Segers, "Data assimilation of ozone in the atmospheric transport chemistry model Lotus," *Environmental Modelling & Software*, vol. 15, pp. 603–609, 2000.

[9] A. Barbu, A. Segers, M. Schaap, A. Heemink, and P. Builtjes, "A multi-component data assimilation experiment directed to sulphur dioxide and sulphate over Europe," *Atmospheric Enviroment*, vol. 43, pp. 1622–1631, 2009.

[10] B. Denby, M. Schaap, A. Segers, P. Builtjes, and J. Horálek, "Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale," *Atmospheric Environment*, vol. 42, pp. 7122–7134, 2008.

[11] G. Candiani, C. Carnevale, G. Finzi, E. Pisoni, and M. Volta, "Off-line data assimilation to provide the best estimate of tropospheric ozone concentrations by means of EnKF," in *49th IEEE Conference on Decision and Control (CDC 2010)*, USA, December 2010, pp. 3493–3498.

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

[13] Eionet, "European topic centre on air and climate change," 2010, [online] http://air-climate.eionet.europa.eu/databases/airbase.