DATA ASSIMILATION FOR TROPOSPHERIC OZONE PREDICTION PROBLEMS USING KALMAN FILTERING

A.W. Heemink[1], A.J. Segers[1], M. Verlaan[1]

[1]Delft University of Technology,
Faculty of Information Technology and Systems,
Department of Technical Mathematics and Informatics
Mekelweg 4, 2628 CD Delft, The Netherlands

P.H.J. Builtjes[2], M. van Loon[2]

[2]TNO, Institute of Environmental Sciences,
Postbus 342, 7300 AH Apeldoorn, The Netherlands

ABSTRACT

Large scale numerical air pollution models are of virtual importance for predicting air pollution concentrations and for reconstructing pollution emissions. Since these models are far from perfect, accurate results can only be obtained by integrating the model results with the concentration measurements that are available, both from satellites as from ground stations. Assimilating data into a numerical air pollution model is however, a procedure that required a huge amount of computer resources. Recently a number of efficient data assimilation algorithms based on Kalman filtering have been developed: the Ensemble Kalman filter and the Reduced Rank Square Root filter. The later algorithms has now been implemented in the LOTOS system for tropospheric ozone simulation studies of TNO. The resulting data assimilation system has been applied to identify the ozone generating mechanisms in the European region.

Key words: tropospheric ozone, data assimilation, Kalman filtering.

1. INTRODUCTION

For densely populated, industrialized countries, environmental modeling and simulation of pollution reduction scenarios is becoming more and more important in view of the growing awareness of damaging effects. Reduction and control of pollution is in general an expensive procedure, leading to high costs for society and industry. It is therefore necessary to determine, as accurately as possible, critical levels and to reduce and control pollution optimally so as to minimize costs.

To accomplish predictions of transport and exchange of chemical and biochemical constituents, accurate, three-dimensional (3D) mathematical simulation models must be used. These large-scale numerical models are based on the advection-diffusion equation. They are, however, far from perfect. Errors are introduced by fluctuations in the meteorological input or by poorly known parameters in the model. Furthermore, considerable uncertainty is associated with the open boundary conditions. Since measurements generally are more accurate then model predictions but limited to a small spatial scale only, it is appealing to combine them with the model results. This is a typical form of *data assimilation*: the incorporation of concentration measurements into a numerical model to improve the forecasts and to reconstruct pollution generating mechanisms. In this project we concentrate our attention to the prediction of tropospheric ozone.

The mechanisms which lead to the formation of tropospheric ozone differ considerably over Europe. At least five clearly different areas can be distinguished: Northern Europe (Scandinavia). Northwest and Central Europe, Southern Europe, and two areas over the sea, the North Sea with parts of the Atlantic, and the Mediterranean basin. The difference in ozone patterns and behavior are caused by a number of often interrelated phenomena. Different dynamical and physical properties, like the differences in temperature, cloud cover and land-sea circulation, differences in surfaces leading to differences in dry deposition, differences in anthropogenic precursor emissions and especially in biogene VOC-emissions, all this leads to differences in ozone patterns over Europe. Not only ground level ozone values will be different, also vertical profiles will be different and subsequently also vertical fluxes of ozone between the planetary boundary layer and the free troposphere, and the free troposphere and the stratosphere. The detailed investigation into these patterns and differences is severely hampered by the limited amount of observations available. Although the amount of observation of especially ozone has increased over the last years, the coverage is still limited to mostly the north-western part of the European continent. There is a severe lack of experimental data of ozone at ground level outside the north-western part of Europe, and hardly any data over the North Sea and Atlantic, and the Mediterranean Sea. Although at some loca-

tions vertical soundings are performed on a daily basis, our knowledge about the vertical profiles of ozone in the troposphere over Europe is very scarce. These available experimental data is insufficient to create a full 3-D data set of ozone in the troposphere over Europe. The three-dimensional Eulerian grid model LOTOS which calculate ozone patterns over Europe is available at TNO and has been used to study the controlling phenomena of ozone over the last decade. Combining these models with the ozone observations from ground level, vertical soundings and satellites by using data assimilation will lead to a coherent and complete full 3-D ozone data set over Europe. Such a data set will enable the differences of ozone patterns over Europe to be studied. Especially by performing budget studies the distinction between the influence of long range transport and of local ozone production can be revealed. Also vertical ozone fluxes over the troposphere and from the stratosphere can be analyzed and determined in detail.

2. DATA ASSIMILATION

Existing data assimilation schemes were developed mainly for numerical weather prediction. The most commonly used data assimilation technique in numerical weather prediction is optimal interpolation. This however, is not an accurate method because the correction produced by optimal interpolation is produced independently from the underlying numerical model, and is therefore not consistent with it.

Date assimilation schemes can also be developed by employing Kalman filtering (Ghil et al. 1981). In order to use a Kalman filter for assimilating data into a numerical transport model, this model is embedded into a stochastic environment by introducing a system noise process. In that way it is possible to account for the inaccuracies of the underlying deterministic system. By using a Kalman filter, the information provided by the resulting stochastic, dynamic model and the (noisy) state of the system. With a Kalman filter, unlike optimal interpolation, the statistics of the introduced noise are determined by using the stochastic extension of the model. Therefore the correction produced by this filter is consistent with the stochastic model.

In the last decenium Kalman filtering has gained acceptance as a powerful tool for data assimilation (Ghil et al. 1981), especially for linear or weakly nonlinear problems. The first applications of Kalman filtering for predicting air pollution were reported by Desalu et al. 1974, Koda & Seinfield 1978 and Fronza et al. 1979. The standard Kalman filter implementation imposes a very large burden on the computer for both memory and computation times. In order to obtain a computationally efficient filter, simplifications have to be introduced. In recent literature a number of new sub-optimal scheme's for solving large-scale filtering problems has been proposed (Cohn & Todling 1995, Verlaan & Heemink 1995, Evensen 1994, Fukumori & Malanotte-Rizzoli 1995).

Another approach to data assimilation which possesses many of the desirable features of Kalman filtering is the adjoint method, based on optimal control theory (see Courtier & Talagrand 1990). Here an unknown control function is introduced into the numerical model. Using the data available, this control function is identified by minimizing a cost function that compares the difference between the model results and the data. In order to obtain a computationally efficient procedure, the minimization is performed by using a gradient based algorithm where the gradient is determined by solving the adjoint problem. The adjoint method is more suitable for nonlinear data assimilation problems than Kalman filtering. However, a disadvantage of the adjoint approach is that it requires the implementation of the adjoint model. This is often very large programming effort compared to the implementation of the Kalman filter.

3. THE RIFTOZ PROJECT

One of the main contributions of Delft University of Technology and TNO to the RIFTOZ project is to develop a data assimilation scheme for the large scale numerical transport chemistry model LOTOS of TNO. Since the LOTOS model is still under development, the data assimilation method has to be flexible and model independent. Therefore the data assimilation is based on Kalman filtering. The Reduced Rank Square Root (RRSQRT) filter implementation does not require a tangent linear model nor an adjoint model. As a result it is relatively easy to implement and completely model independent. (Verlaan & Heemink 1995)

In the first year of the project the (first order) RRSQRT filter has been implemented for a test problem. Experiments were performed with simulated data to gain insight into the values of the various parameters of the data assimilation scheme. Attention has been concentrated on the performance of the filter with respect to the strong nonlinearity of the chemistry model. Results have been published by Van Loon and Heemink (1997).

In the second year of the project the RRSQRT filter has been improved by including second order terms in the algorithm. This new approach was evaluated in detail by using the test model with simulated data (see Segers et al. 1998). Furthermore the filter implementation was also coupled with the LOTOS model.

In the final phase of the project the filter technique has been applied on the LOTOS model. First, similar experiments as which were done with the test model has be applied with LOTOS in order to judge the system performance. These tests gave an indication of the required computing power. Second, the actual assimilation of real data with LOTOS calculations has been done by using the CRAY T3E of Delft University of Technology.

4. RESULTS WITH THE LOTOS SYSTEM

An important aspect of the RRSQRT filter is the computational efficiency. Here the dominating parameter is the number of modes used. The computational burden increase approximately linearly with this number. At the other hand also the accuracy of
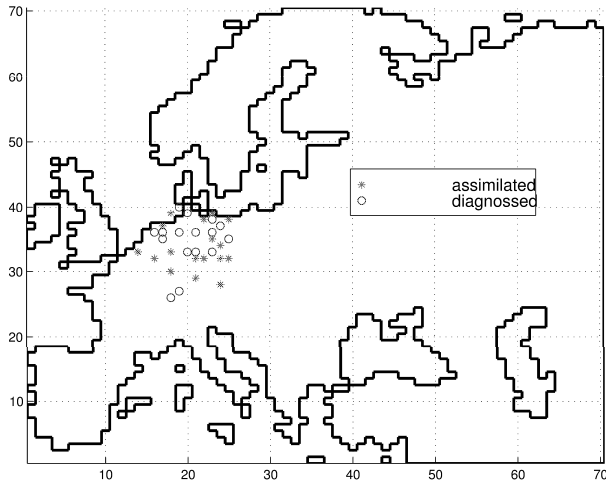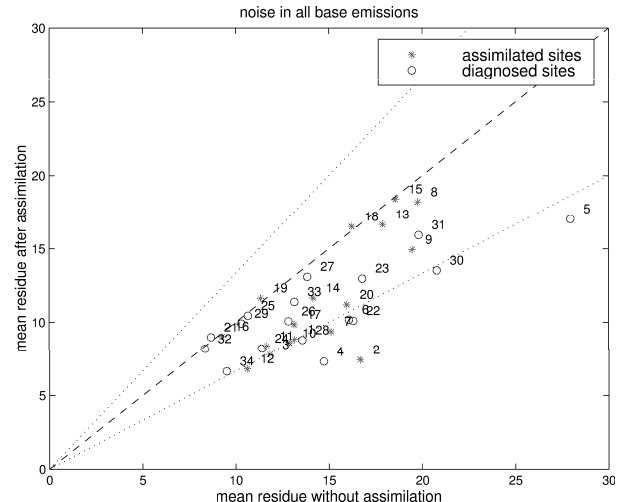
Figure 1. LOTOS grid and measurement sites.

Figure 2. Mean residues during the assimilation period versus the mean residues from a background run (a single LOTOS run without assimilation). The dashed line denotes equality, while the dotted lines denote a difference of ±30%.
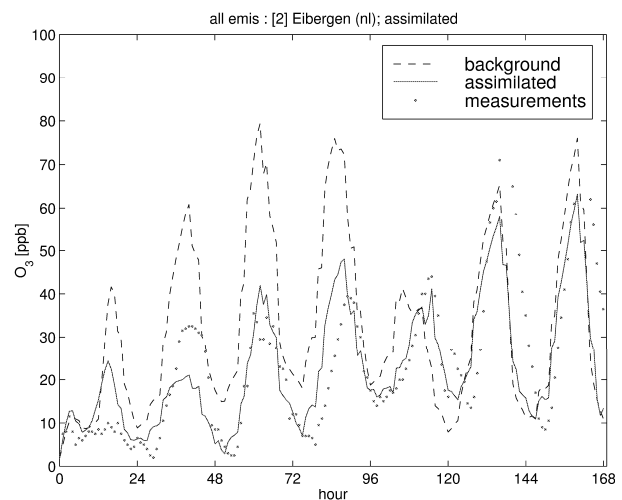

Figure 3. Ozone concentrations measured and calculated at the assimilated site Eibergen.

the data assimilation scheme increases with the number of modes. The experiments have shown that approximately 50 modes seems to be a good choice for atmospheric chemistry models. As a result the total computational burden will be 50-60 times the computational afford required for the underlying LOTOS model. Here we note that the filter implementation contains a lot of parallelism that can be exploited to improve the performance of the filter implementation on the CRAY T3E of Delft University of Technology considerably.

To find a useful specification of uncertainties in the LOTOS model, a sequence of assimilation experiments has been performed. The available set of measurements consists of hourly measured ground level values of ozone, measured at 34 sites in Germany and The Netherlands. The data from 17 of these sites are assimilated, while the other are only diagnosed (figure 1). The first week of august 1997 was taken as assimilation period.

One of the assimilation experiments made use of uncertainties specified for different emissions. The LOTOS model recognizes four types of emitted pollutants (VOC,NO_x,SO_x,and CO), emitted from 5 different antropogenic source categories; VOC is also emitted from 3 biogene sources. These emissions were supposed to vary from hour to hour with a standard deviation of 25% around their deterministic value. Spatial fluctuations in the variations are neglected. Because the measurement sites used for assimilation are located in in a rather small area, the later is not problematic. In fact, the chosen specification reflects the case of systematic error in the (modeling of the) emission data bases.

If all 23 emissions are specified to be uncertain in the way described, the filter is able to decrease the residue (difference between calculations and measurement data) with a maximum of about 30% in comparison with a non filtered simulation (figure 2). The improvement of the results is best shown by the decreased residue of the diagnosed sites, which have not been used in the assimilation process, but are only influenced by the assimilation of other data.

Figure 3 shows the ozone concentrations as calculated

for the assimilated site Eibergen in The Netherlands. Without assimilation, the model overestimates the ozone concentrations, especially during the first days of the selected period. If however the available data is assimilated, the mean concentration almost perfectly follows the data after an initialization period of two days.

Investigation of the actual variations in the emissions used by the filter to reduce the residues, resulted in a selection of 8 emission data bases of the 23 in total, for which the chosen specification of uncertainty is most useful in an assimilation procedure. All of these selected emissions had an antropogenic source; this reflects the fact that biogene sources have a minor impact on the ozone concentrations in Western Europe where the measurements sites involved in this study are located.

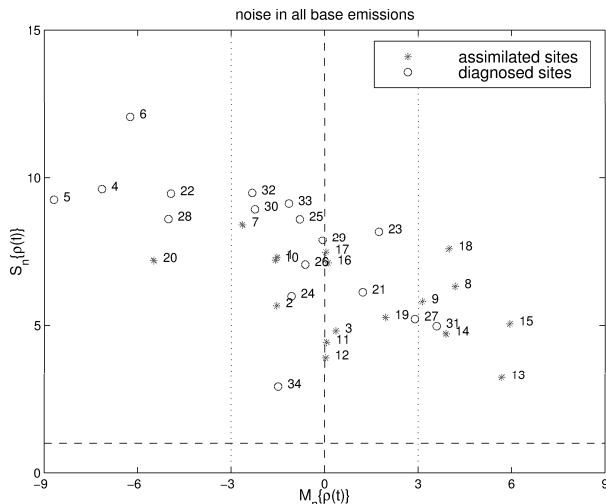A suitable way to judge the performance of the

Figure 4. Sample mean M_n and sample standard deviation S_n of the ratio ρ between the actual residue and its expected standard deviation taken over the assimilation period, for each available site. The ratio has a theoretical distribution of $N(0,1)$. The actual sample means are sometimes rather close to zero, but a standard deviation of one is never reached.

Kalman filter, is to compare the actual residues with their expected standard deviation. The later can be extracted from the covariance matrix of the state, and is in fact an impression of how accurate the filter guesses that the residue is. The ratio ρ between the residue and its guessed standard deviation is in theory $N(0,1)$ distributed. In practice however, the spread of the ratios has a standard deviation larger than one, because the actual residues are often much larger than a few times their expected standard deviation. This is also the case for the previous described filter experiment (figure 4). The underestimation of the standard deviation of the residues indicates that the noise specification chosen here does not account for all uncertainties in the model, and that other specifications of uncertainty should be examined.

In order to extend the noise specification and to increase the filter performance, the impact of uncertainties in the upper boundary conditions was examined. Instead of using deterministic aloft concentrations, a new boundary condition was formed using total ozone columns measured with the GOME satellite instrument. Because of the large height of the columns in comparison with the height of the LOTOS grid (about 2 km), they could not serve as measurements directly. Debruyn et al. (1998) calculated daily values for the total ozone column in the first 9 km of the troposphere from the raw GOME data, with a standard deviation of 40%-50%. These columns served to calculate mean aloft concentration; a grid of stochastic variations (forced by 16 parameters) was added to the new aloft concentrations in order to represent the error.

Assimilation of data using a selected number of uncertain emissions in combination with the new boundary condition, did however not resulted in an additional decrease of the residues. Comparison of the new model output with the output of a model using the original boundary conditions showed that there is hardly any difference in calculated ground level concentrations. Similar, specification of uncertainties in the aloft concentrations did not have a significant impact on the results in case of an assimilation experiment. Even the ratio between residue and expected standard deviation was not improved through the increased amount of available uncertainties. These results show that the impact of the upper boundary on ground level concentrations is minor.

In spite of the minor impact of the upper boundary, the large amount of GOME data which is available makes the idea of using it still interesting. Use of the GOME data as measurements will be complicated through the limited height of the LOTOS grid, and the large error present in the profiles. The operational Kalman filter is however a suitable tool to decide what the minimum accuracy of the profiles should be in order to have a significant impact in a data assimilation procedure, and this will be subject of further research. Besides, the increased availability of accurate measurements in the vertical by means of balloon soundings makes it possible to examine the impact of assimilation of satellite data on the accuracy of LOTOS calculations at higher grid layers.

Systematic research of the impact of other (groups of) stochastic parameters on the assimilation results should result in a minimal stochastic extension of the LOTOS model for use in data assimilation. Increase of the number of uncertain parameters improves the results, but also implies an increased demand on computation capacity. Continued experiment should therefore point out uncertainties responsible for the majority of the residue.

5. CONCLUSIONS

The results up to now indicate that the assimilation of data into the LOTOS model with a Kalman filter procedure is feasible. The nonlinearities do not seem to cause serious problems and the computational burden of the algorithm is large but not too large. For the final computations, however, a very powerful computer was required (CRAY T3E). The filter results with the real data show a good performance of the system. In a number of measurement locations that have not been used in the assimilation procedure, the model results with assimilated data are 10%-30% more accurate then the original model results. The use of GOME data does not seem to have have a significant impact on the results in the current setup.

ACKNOWLEDGMENTS

REFERENCES

Cohn, S.E., Todling, R., Approximate Kalman filters for unstable dynamics, In *Second Int. symposium*

on Assimilation of observations in Meteorology and Oceanography, pages 241–246, Tokyo, WMO, 1995.

Courtier, P., Talagrand, O., Variational assimilation of meteorological observations with direct and adjoint shallow water equations, *Tellus*, 42A:531–549, 1990.

Desalu, A.A., Gould, L.A., Schweppe, F.C., Dynamic estimation of air pollution, *IEEE Trans. Autom. Control*, AC-19:904–910, 1974.

Debruyn, W., Brems, E., Lissens, G., RIFTOZ: Regional Differences in Tropospheric Ozone in Europe, *technical report*, 1998/TAP/R/049, 1998

Evensen, E., Sequential data assimilation with a nonlinear QG model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99:10143–10162, 1994.

Fronza, G., Spirito, A., Tonielli, A., Real-time forecast of air pollution episodes in the Venetian region. Part 2: the Kalman predictor, *Appl. Math. Modelling*, 3:409–415, 1979.

Fukumori, I., Malanotte-Rizzoli, P., An approximate Kalman filter for ocean data assimilation; an example with an idealized Gulf Stream model, *J Geophys, Res*, 100:6777–6793, 1995.

Ghil, M., Cohn, S.E., Tavantzis, J., Bube, K., Isaacson, E., Application of estimation theory to numerical weather prediction, In M. Ghil L. Bengtsson and E. Källén, editors, *Dynamic Meteorology: Data Assimilation Methods*, New York, Springer-Verlag, 1981.

Koda, M., Seinfield J.H., Estimation of urban air pollution, *Automatic*, 14:583–595, 1978.

Segers, A.J., Heemink, A.W., Verlaan, M., Van Loon, M.,. Nonlinear Kalman Filters for Atmospheric Chemistry Models, Submitted to *American Geophysical Union*, 1998.

Verlaan, M., Heemink, A.W.,. Data assimilation schemes for non-linear shallow water flow models, In *Proceedings of the Second International Symposium on Assimilation of Observations*, pages 247–252, Tokyo, Japan, WMO, 1995.