DATA ASSIMILATION WITHIN THE AIR4EU PROJECT: THE ATHENS CASE

Theodora Slini[1], *Sam Erik Walker*[2] *and Nicolas Moussiopoulos*[1]
[1] Laboratory of Heat Transfer and Environmental Engineering, Aristotle University Thessaloniki, Box 483, GR-54124, Thessaloniki, Greece
[2] Norwegian Institute for Air Research, P.O.Box 100, 2027 Kjeller, Norway

INTRODUCTION

Complex models include a large number of poorly known parameters. While data are useful to validate and improve the numerical models, the assimilation of these is now recognized as the most efficient way to improve consistency between data sets and model simulations (*Triantafyllou et al.,* 2003). State estimation through data assimilation is a key to the development of appropriate forecasting systems.

In the framework of the Air4EU project, a simplified data assimilation (DA) methodology was developed, aiming to improve the air quality models performance at local, urban and regional scales. This work was focused on the enhancement of the urban scale model OFIS, by means of the developed DA tool. The DA tool is based on the Sequential Importance Resampling (SIR) method, a technique that makes no assumptions of linearity in the model equations, nor that the model or observation errors should be Gaussian, in contrast with most other well-known methods of data assimilation, such as the optimal or statistical interpolation methods, the three-dimensional or four-dimensional variational data assimilation method, or different variants of the Kalman filter.

METHODOLOGY

DA has been widely applied in various research fields for several modules in the last two decades, especially in the field of atmospheric dispersion, deposition, numerical weather prediction or meteorological pre-processing which produces meteorological data for the emergency response systems, food chain and hydrological modules and oceanic sciences (*Kovalets et al.,* 2003). *van Loon et al.* (2000) introduced DA in an atmospheric transport chemistry model to improve modelled ozone concentration.

A Bayesian view is taken in which the prior probability density of the model and the probability density of the observations are combined to form a posterior density. The mean and the covariance of this density give the variance-minimizing model evolution and its errors. At the same time, observational error can be reduced and information about model errors can be generated. Using Bayesian statistics one can consider the probability density of the model forecast as prior information, which is ‘‘updated’’ by the observations. This results in a new probability density of the model, given the observations.

The SIR method (*van Leeuwen,* 2003) is a relatively new data assimilation method that is truly variance minimizing, as no matrix inversions are needed, the observations can be distributed non-Gaussian, and the measurements can be nonlinear. Furthermore, it preserves prior model constraints such as positive definiteness, unlike Kalman filter–like methods that mix states at analysis time, and provides error estimates, unlike 4DVAR-like methods. Finally, it is easy to implement and parallel, by its very nature. Another interesting feature of the method is that the variance of the posterior density can be larger than that of the prior density.

APPLICATION

A tool was developed based on the SIR method (*Walker et al.*, 2006) and applied using the OFIS model results. The tool initially calculates the probability density functions (pdfs) of both observed and modelled concentration values which are coupled and combined using the Bayes Theorem to derive an assimilated value. Several types of distributions were included in the tool such as normal, log-normal, Cauchy. However, for the current study only the normal distribution was applied.

The SIR DA tool was applied to a number of air pollution concentrations, including NO_2, NO, O_3 and PM_{10}, for the Greater Athens Basin, Greece and the OFIS model results for the same time period. A one year long data set (2002) is being used, consisting of hourly air pollutant concentrations, as resulting from the operation of the corresponding monitoring networks at the stations of Agia Paraskevi, Liosia, Lykovrisi, Marousi, Patision, Pireaus and Thrakomakedones. This time period was selected on the basis of data homogeneity availability and completeness. Information on the air pollution monitoring network of Athens can be found in *Directorate of Air and Noise Pollution*, 2005.

The OFIS model is a photochemical dispersion model for calculating ground level of several air pollutants (like NO, NO_2, O_3, PM_{10} and $PM_{2.5}$) concentrations in and around urban areas (*Moussiopoulos and Sahm,* 2001). It belongs to the European Zooming Model system (EZM), a comprehensive model system for simulations of wind flow and pollutant transport and transformation (*Moussiopoulos,* 1995).

RESULTS AND DISCUSSION

The observed, modelled and "assimilated" concentrations of NO_2, NO, O_3 and PM_{10} were compared and evaluated using certain statistical performance measures (*Wilmott*, 1982*; Wilmott et al.*, 1985) like the mean value, bias, correlation coefficient (CC) between observed and modelled values, normalised mean square error (NMSE) and the index of agreement (IA) that is sensitive to differences between the observed and model means as well as to certain changes in proportionality.

The results are presented in Tables 1-4 for the NO_2, NO, O_3 and PM_{10} concentrations (in $\mu g/m^3$) respectively at the monitoring stations under investigation for the year 2002.

Table 1. Performance statistics for NO_2 observed (obs), modelled (OFIS) and assimilated (ass) values (in $\mu g/m^3$) at the monitoring stations of Agia Paraskevi, Liosia, Likovrisi, Marousi, Patision, Pireaus and Thrakomakedones for the year 2002.

	MEAN			BIAS		NMSE		CC		IA	
	obs	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**
AG.PAR	18	19	18	0.84	-0.01	0.97	0.32	0.141	0.871	0.59	0.89
LIOS	41	36	34	-5.24	-6.33	1.51	0.42	0.056	0.620	0.51	0.79
LYK	37	38	34	1.15	-0.86	0.59	0.20	0.208	0.813	0.55	0.86
MAR	42	44	41	1.91	-0.74	0.52	0.16	0.214	0.805	0.53	0.86
PAT	92	91	85	-0.07	-5.80	0.33	0.10	0.058	0.584	0.43	0.77
PIR	65	58	54	-7.91	-6.85	0.38	0.12	0.175	0.696	0.58	0.83
THRA	11	13	11	2.32	0.78	2.27	0.75	0.027	0.702	0.49	0.84

Table 2. Performance statistics for NO observed (obs), modelled (OFIS) and assimilated (ass) values (in µg/m^3) at the monitoring stations of Agia Paraskevi, Liosia, Likovrisi, Marousi, Patision, Pireaus and Thrakomakedones for the year 2002.

	MEAN			BIAS		NMSE		CC		IA	
	obs	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**
AG.PAR	3	43	13	39	10	51.56	5.56	0.132	0.454	0.06	0.44
LIOS	27	17	14	-10	-4	11.57	1.07	0.204	0.845	0.48	0.92
LYKO	23	42	21	19	-1	6.61	2.27	0.327	0.706	0.49	0.82
MAR	35	46	28	11	0.02	5.37	0.96	0.349	0.889	0.59	0.93
PAT	132	56	68	-75	-41	2.53	0.49	0.317	0.815	0.55	0.86
PIR	54	29	30	-26	-14	3.32	0.73	0.376	0.834	0.62	0.87
THRA	6	2	5	-4	-0.59	4.33	1.04	0.134	0.534	0.63	0.79

Table 3. Performance statistics for O_3 observed (obs), modelled (OFIS) and assimilated (ass) values (in µg/m^3) at the monitoring stations of Agia Paraskevi, Liosia, Likovrisi, Marousi, Patision, Pireaus and Thrakomakedones for the year 2002.

	MEAN			*BIAS*		*NMSE*		*CC*		*IA*	
	obs	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**
AG.PAR	93	70	72	-21	-18	0.67	0.19	0.499	0.721	0.58	0.79
LIOS	64	69	62	5	-2	0.52	0.15	0.440	0.868	0.78	0.94
LYK	58	47	49	-10	-9	0.68	0.19	0.574	0.860	0.80	0.93
MAR	51	37	41	-134	-10	0.82	0.24	0.541	0.856	0.76	0.91
PAT	18	23	19	4	1	1.69	0.47	0.383	0.817	0.69	0.91
PIR	43	42	38	0,1	-3	0.82	0.23	0.489	0.843	0.69	0.91
THRA	94	73	78	-20	-13	0.23	0.08	0.492	0.773	0.91	0.96

Table 4. Performance statistics for PM_{10} observed (obs), modelled (OFIS) and assimilated (ass) values (in µg/m^3) at the monitoring stations of Agia Paraskevi, Liosia, Likovrisi, Marousi, Patision, Pireaus and Thrakomakedones for the year 2002.

	MEAN			BIAS		NMSE		CC		IA	
	obs	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**	**OFIS**	**ass**
AG.PAR	38	26	25	-11	-8	2.38	0.51	-0.340	0.667	0.57	0.85
LYK	62	34	36	-31	-19	1.72	0.45	0.088	0.669	0.42	0.70
MAR	69	38	26	-27	-19	2.55	0.74	0.091	0.749	0.69	0.89
PIR	62	40	35	-22	-18	1.98	0.49	-0.002	0.592	0.56	0.82
THRA	34	22	20	-13	-9	2.31	0.56	0.041	0.699	0.62	0.86

As it can be pointed out, the use of the developed DA tool in the OFIS model resulted in major improvements almost in all indexes for each pollutant, even in cases where the model has a poor performance, like at the Liosia station for NO concentrations, PM_{10} concentrations at stations with available data. The NMSE was significantly reduced, expressing a decrease of the systematic error accounted to the model. At the same time, the correlation coefficient states an improved ability to capture the linear connection between observations and assimilated values. Last but not least, the IA, which allows for sensitivity toward difference in observed and predicted values as well as proportionality changes, indicates a satisfactory enhancement of the model results when the DA is applied.

Figures 1-2 present a random time series of the observed, modelled and assimilated hourly average values (all in µg/m^3) of all studied pollutants indicatively for the monitoring stations of Thrakomakedones and Agia Paraskevi.

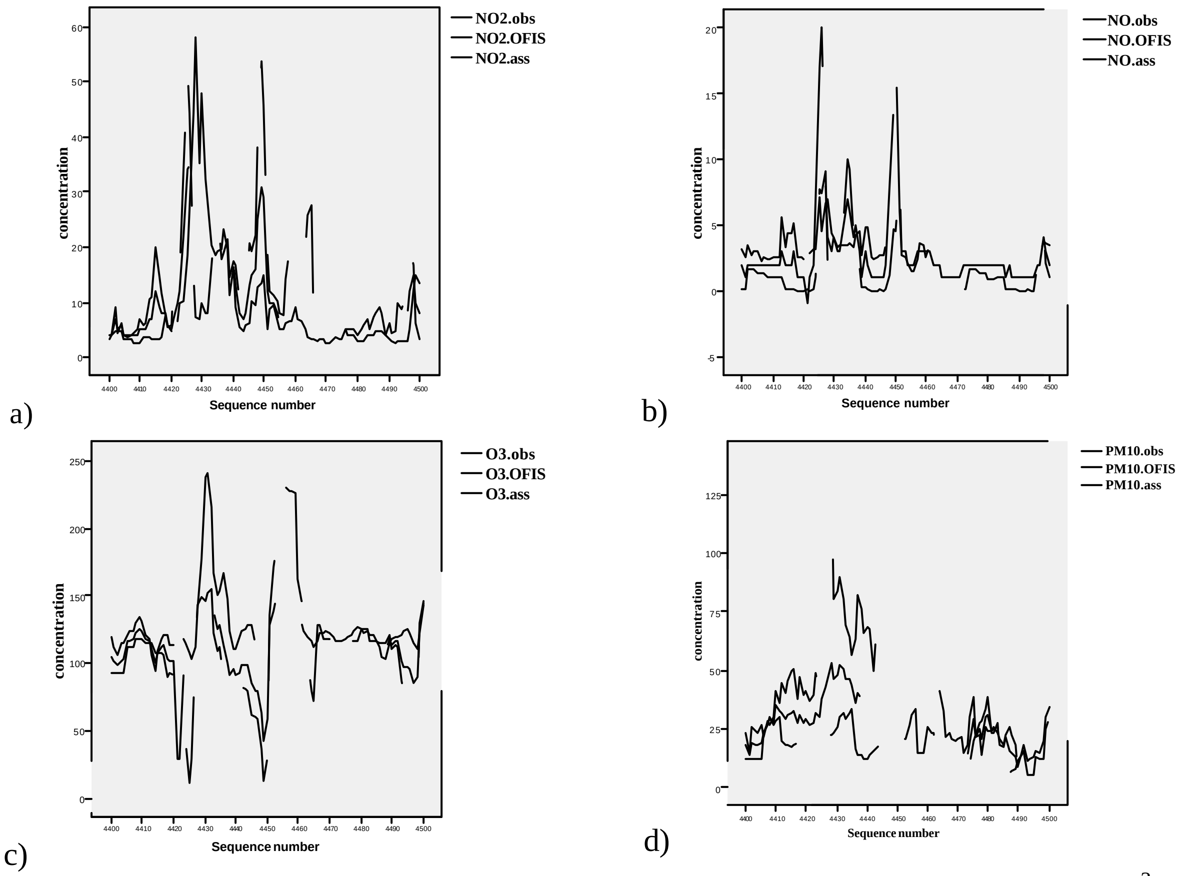

Fig. 1; Time series of the observed and predicted hourly average concentrations (in μg/m^3) of a) NO_2, b) NO, c) O_3 and d) PM_{10} at the Thrakomakedones monitoring station.

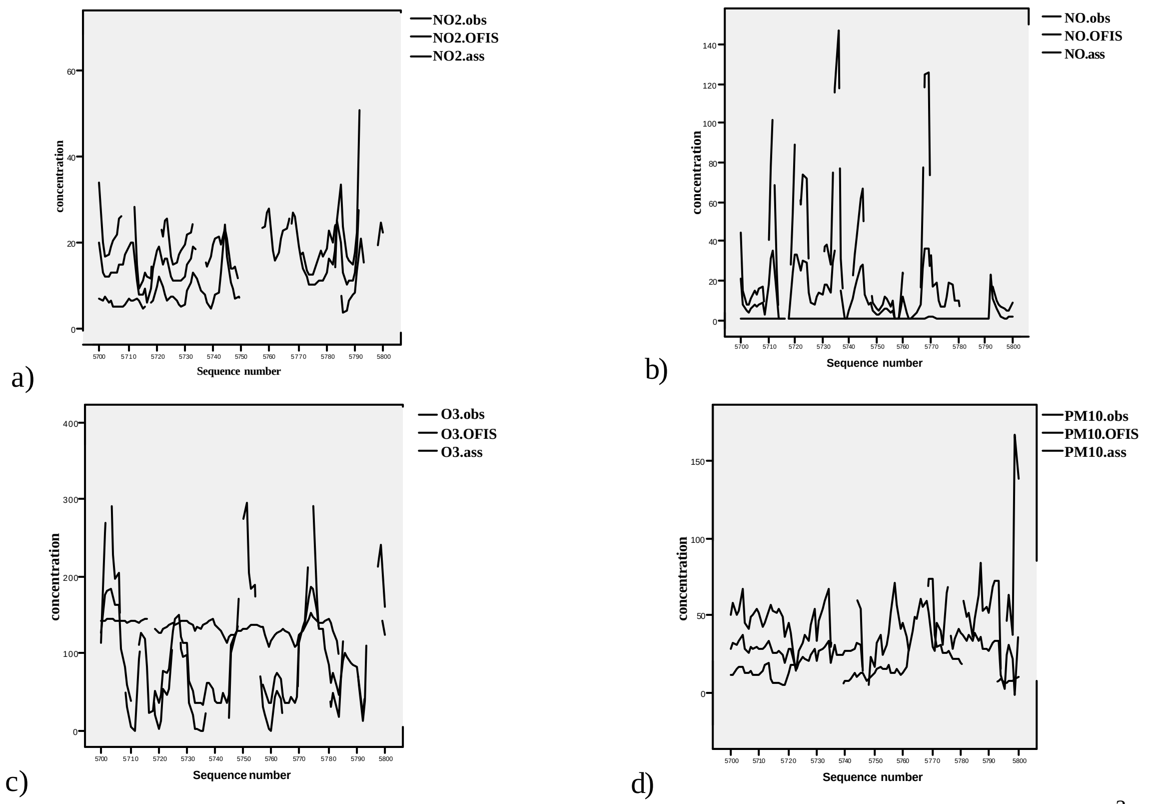

Fig. 2; Time series of the observed and predicted hourly average concentrations (in μg/m^3) of a) NO_2, b) NO, c) O_3 and d) PM_{10} at the Agia Paraskevi monitoring station.

However, the DA tool performance is not very sufficient in cases when the pollutant does not follow the normal distribution, reflecting some abnormal situations and distinct episodes, such as the presence of uncontrolled factors with dispersion from urban environments or huge variability due to emission sources and extreme or unstable meteorological conditions. Thus, the implementation of different kind of distribution should be performed in a following study. Nevertheless, the quality and usefulness of the application of the DA technique is reflected in abnormal situations and distinct episodes.

CONCLUSIONS

The suggested approach seems to be attractive because it can handle complex phenomena that cannot be appropriately described by the numerical equations of a model. The results indicate that overall the model without data assimilation performs well in simulating the trend and magnitude of the observed concentrations, with a small bias towards under prediction. Air quality episodes are unique and the relative processes and variables are seldom linear. Data Assimilation is the ensemble of techniques combining in an optimal way the mathematical information provided by the model equations and the physical information given by the observation in order to retrieve the natural state space of a process. The goal of DA is to link together these heterogeneous (in nature, quality and density) sources of information in order to retrieve a coherent state of the environment at given conditions.

REFERENCES

Directorate of Air and Noise Pollution, 2006: Air pollution in Athens-2005. Hellenic Ministry of Environment, Planning and Public Works. http://www.minenv.gr/4/41/00/ekthesi.2005.pdf

Kovalets, I., S. Andronopoulos, J. G. Bartzis, N. Gounaris and A. Kushchan, 2003: Introduction of data assimilation procedures in the meteorological pre-processor of atmospheric dispersion models used in emergency response systems. Atmospheric Environment 38, 457–467.

Moussiopoulos, N., 1995: The EUMAC Zooming Model, A tool for local-to-regional air quality studies. *Meteorology and Atmospheric Physics*, **57**, 115-133.

Moussiopoulos, N. and P. Sahm, 2001: The OFIS model: an efficient tool for assessing ozone exposure and evaluating air pollution abatement strategies. *International Journal of Environment and Pollution*, **14**, 597-606.

Triantafyllou, G., I. Hoteit and G. Petihakis, 2003: A singular evolutive interpolated Kalman filter for efficient data assimilation in a 3-d complex physical biogeochemical model of the Cretan Sea. *Journal of Marine Systems,* **40-41**, 213–231.

van Leeuwen, P. J., 2003: A variance minimizing filter for large scale applications. *Monthly Weather Review,* **131**, 2071-2084.

van Loon, M., P. J. H. Builtjes and A. J. Segers, 2000: Data assimilation of ozone in the atmospheric transport chemistry model LOTOS. *Environmental Modelling and Software,* **15**, 603–609.

Walker, S.-E., M. Schaap and T. Slini, 2006: Air4EU Milestone Report 6.8. Cross-Cutting Issue 5: Data Assimilation. Norwegian Institute for Air Research.

Wilmott, C. J., 1982: Some comments on the evaluation of model performance. *Bulletin of American Meteorological Society,* **63**, 1309-1313.

Wilmott, C. J., S .G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Donnell and C. M. Rowe, 1985: Statistics for the evaluation and comparison of models. *Journal of Geophysics Research,* **90**, 8995-9005.