THE COMBINED USE OF MODELS AND MONITORING FOR APPLICATIONS RELATED TO THE EUROPEAN AIR QUALITY DIRECTIVE: A WORKING SUB-GROUP OF FAIRMODE

Bruce Denby[1] and Wolfgang Spangl[2]

[1]The Norwegian Institute for Air Research (NILU). PO BOX 100, Kjeller 2027, Norway. bde@nilu.no
[2] Federal Environment Agency, Austria

Abstract: The Forum for Air Quality Modelling in Europe (FAIRMODE) has recently established a number of sub-groups under the working group for model quality assurance (WG2). These sub-groups are intended to discuss, promote and develop recommendations on harmonised quality assurance approaches when using models for applications related to the European AQ Directive. One of these sub-groups has been established to deal with the combined use of monitoring and monitoring data and the spatial representativeness of monitoring data used for assessment and validation purposes. In this paper an overview of the various methods currently used to combine monitoring and modelling data is provided along with the relevant institutes and projects that apply these methods. In this regard their use in AQ Directive related applications such as the determination of exceedances, forecasting and providing near real time public information is addressed.

Key words: *Air quality modelling, air quality assessment, Directive, data assimilation, FAIRMODE*

INTRODUCTION

The Forum for Air Quality Modelling in Europe (FAIRMODE) has recently established a number of sub-groups under the working group for model quality assurance (WG2). These sub-groups are intended to discuss, promote and develop recommendations on harmonised quality assurance approaches when using models for applications related to the European AQ Directive. One of these sub-groups has been established to deal with the combined use of monitoring and monitoring data and the spatial representativeness of monitoring data used for assessment and validation purposes.

Traditionally monitoring data has been used to assess air quality. However, the limited spatial representativeness of these data does not allow for the complete spatial coverage required by the EU AQ Directive or to properly assess human and eco-system exposure. More complete spatial coverage is available through the use of air quality models but these are generally considered to have a higher uncertainty than monitoring. By combining these two sources of data it is possible to provide more optimal estimates of the spatial distribution of air quality.

There are a variety of methods available to achieve this combination, ranging from geometric combinations of the data sources, through statistically based methods of interpolation and 'data fusion' to the use of 'data assimilation' methods. It is the aim of this FAIRMODE sub-group to bring together the various groups applying these methods in order to promote good practise and to develop and apply quality assurance practices when combining models and monitoring.

In this paper an overview of the various methods currently used to combine monitoring and modelling data is provided. In addition a table of institutes and projects currently implementing data assimilation or fusion methods are provided. This paper is a first step in developing an overview of activities in data assimilation and fusion in Europe, with the intention of improving methodologies and developing harmonised quality assurance practises.

OVERVIEW OF METHODS FOR COMBINING MODELS AND MONITORING

There is a range of methods available for combining models and monitoring data. An overview of some of these methods can be found in Denby et al. (2005) and Denby et al. (2009). We use the general term 'combination of modelling and monitoring' to describe any method that makes use of both models and monitoring to provide improved information on air quality. Other terms are also often used to describe these methods or to more specifically describe any particular method. For example 'data integration' or 'data fusion' are terms used when combining different data types but without any indication of how they are actually combined. 'Data assimilation' is most commonly used to describe the use of monitoring data to provide improved modelling results, often based in a Bayesian framework. These terms are often loosely applied, dependent on the application and motivation. However, it is worth making a distinction between the different methods based mostly on their accessibility.

Data integration: Refers to the 'bringing together' of various data sources (e.g. monitoring, modelling, satellite, meteorology, emission) in a common form or in a common system to enable their use in that system. It does not necessarily refer to any combined use of the same type of data for improved modelling. Though integration is important for air quality modelling it is not the subject of this paper or working group.

Data fusion: Is a general term that refers to any method that combines, in either a statistical or geometric way, various data sources to create a new data set. E.g. the weighted combination of satellite and modelling data to provide new maps of air quality (Sarigiannis et al., 2004) or the use of regression or other least square methods for adjusting modelling data using monitoring data (Denby and Pochmann, 2007). Data fusion may also make use of supplementary (proxy) data, such as land use or meteorological data, which can provide relevant spatial information for air quality assessment. What mostly distinguishes these methods from 'data assimilation' is that they do not take into account any physical laws or limitations but are generally 'statistical' in nature. These methods can also be seen as post processing methods for modelling results, i.e. the result does not interact with the model, and have also be termed 'passive data assimilation'.

Data assimilation: Refers to any method where monitoring data is used to 'guide' models towards monitoring results during the model integration. These methods ensure that the physical and chemical character of the problem, as described by the model, is followed. Such methods are widely employed in meteorology (e.g. 4D-var) providing continuously updated initial conditions for forecasting. These methods are generally based in a Bayesian framework. They are also sometimes termed 'active data assimilation'.

It is useful to separate the methods into data fusion and data assimilation as described above for practical reasons. Data fusion does not require complete access and understanding of the model but requires only the model results. This makes data fusion a much more accessible and less complicated method to implement than data assimilation that requires expert knowledge of both the model and the data assimilation method for it to be implemented. Data fusion also allows the use of simpler models, e.g. statistically based multi-source Gaussian models, that could not be implemented in a data assimilation framework. One of the important drawbacks with data fusion methods however is that they do not adhere to the physics of the problem and they can result in quite impossible realisations of air quality, e.g. NO_2 concentrations that are far from equilibrium with ozone concentrations. Though the above differentiation is made, many of the techniques used in data fusion and assimilation are actually based on the same principles, i.e. minimisation of some specified error. In the following sections a brief review of data fusion and data assimilation methods, as defined here, is provided.

Data fusion

As previously described data fusion methods take a variety of data sources such as ground based monitoring, air quality models, satellite retrieved data or any other spatially distributed data relevant to air quality (such as altitude, land use or emissions) and combines these data in some (hopefully optimised) way to produce an air quality assessment. One of the most straightforward methods is single or multiple linear regression, where model concentrations, and perhaps other supplementary data, are fitted to the available observations using least squares optimization (e.g. Horálek et al., 2007; Denby and Pochmann, 2007). Regression methods may also be used to estimate the background concentration (EC working group, 2000) or multiple linear regression may also be applied on the individual model source contributions in order to adjust the modelled field (Laupsa et al., 2009).

Linear regression methods will provide an unbiased model field, in regard to the observations, but there may still be significant deviation from the observations. To account for these deviations residual interpolation methods may be applied. With such methods the residual, the difference between the modelled and observed concentrations, may be interpolated using either geometric methods, e.g. Inverse Distance Weighting (Hogrefe et al., 2009) or Radial Basis Functions (Tarrason et al., 1998), or using geostatistical methods such as kriging. When kriging methods are employed then the interpolation is made in order to minimise the spatial variance, i.e. to provide the statistically most likely concentration at any point in space. In this way the model field provides the basis for the concentration map and the residual deviations are accounted for by using interpolation methods (e.g. Horálek et al., 2007; Kassteele et al., 2006; 2007; Hogrefe et al., 2009). These methods may also be called, or are similar to, 'universal kriging', 'detrended kriging' or 'kriging with external drift' methods.

In addition to the regression and residual interpolation methods outlined above there are also a number of more complex statistically based methods for achieving data fusion. Such methods include those described by Fuentes and Raftery (2005), Gelfand and Sahu (2009) and McMillan et al. (2009). These methods combine Bayesian approaches with a range of statistical methods. Optimal interpolation, which can also be used as a data assimilation method similar to 2-D var, may also be used as a data fusion method (Flemming et al., 2002).

Data assimilation

As outlined above we define data assimilation methods as those actively implemented in a model to provide optimal solutions during run time. A number of the methods used for data fusion, e.g. optimal interpolation and residual kriging are also applied for data assimilation (e.g. Flemming et al., 2002; Blond et al., 2003).

The most common type of data assimilation methods applied are the variational methods. These methods are termed either 2-, 3- or 4D var dependent on the dimensions used for the assimilation. Variational methods in chemical transport models were first implemented by Elbern et al. (1999), though these methods have been extensively used in meteorological forecasting previously. The variational methods are based on the minimization of a cost function for the difference between model concentrations and observations (Lorenc, 1986). These techniques require the development and implementation of a so-called adjoint version of the model. There are currently few models in operational use that have implemented 4D var methods and these are region or global scale CTMs including EURAD (Elbern, 1999), Polyphemus (Malet et al., 2007) and MOCAGE (Geer et al., 2006).

Another method that is applied to regional scale CTMs is the Ensemble Kalman filter (Evensen, 1994; van Loon et al., 2000). This methodology requires an ensemble of model runs, perturbing each ensemble member in some way, so that the model error covariance matrix may be estimated with this ensemble. This method avoids the complications of implementing an adjunct model but requires a significant number of ensemble members for its implementation. This method, and its variations, are currently implemented in LOTOS-EUROS (Schaap et al., 2005), Polyphemus (Malet et al., 2007) and are currently being implemented in CHIMERE.

Data assimilation is now used operationally in air quality forecasting (e.g. Sahu et al., 2009 and MACC, www.gmes-atmosphere.eu/) and it is also applied for air quality assessment purposes (Denby et al., 2008), see Figure 1. In addition it is being applied for source apportionment studies by using the data assimilation methods as inverse modelling techniques to derive emission estimates (Elbern et al., 2007). Data assimilation is most often applied on the regional scale and is rarely applied on the urban scale.

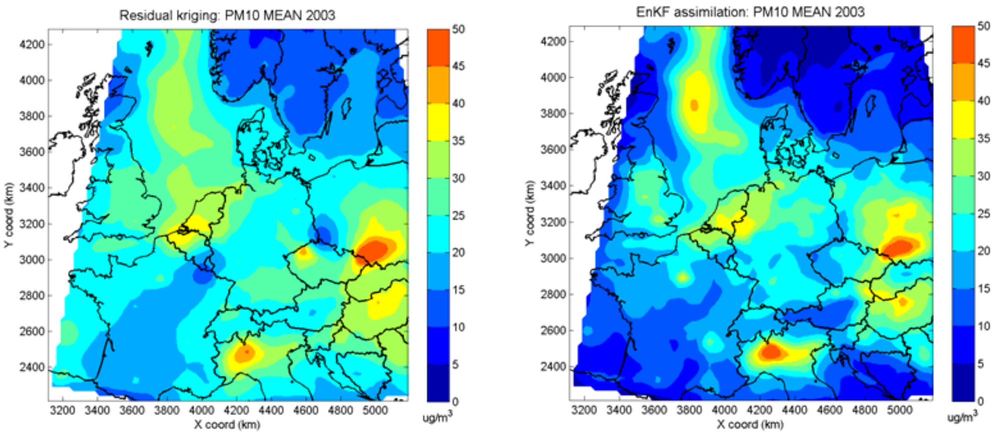

Figure 1. Maps showing the annual mean concentration of PM_{10} in 2003 using residual kriging and regression (left) and Ensemble Kalman Filter (right). Taken from Denby et al. (2008).

Activities within Europe

There are a number of projects that are dealing with data assimilation or fusion applications. In particular the GMES projects of MACC (www.gmes-atmosphere.eu/) and PASADOBLE (no URL available) will provide air quality assessment and forecasting using a number of different data assimilation techniques. Apart from these two operationally oriented projects there are also a number of institutes in Europe engaged in data assimilation activities. Table 1 provides an overview of these institutes. The table is not currently complete but provides an indication of a number current activities.

Table 1. List of contacts and institutes actively engaged in data assimilation/ fusion applications or in studies concerning representativeness in Europe.

Person	**Institute**/project	**Contact**	**Model**	**Method**	**Application (resolution)**
Hendrik Elbern	**RIU**/MACC/PASADOBLE	he@eurad.Uni-Koeln.DE	EURAD-IM	3-4D var	European forecasts (45 – 1 km)
Martijn Schaap	**TNO**/MACC	martijn.schaap@tno.nl	LOTOS_EUROS	Ensemble Kalman filter	European assessments and forecasting (25km)
L. Menut	**INERIS**/MACC	menut@lmd.polytechnique.fr	CHIMERE	Optimal interpolation , residual kriging and EnKF (in development)	European and Urban scale forecasts and assessments (25 km)
Hilde Fagerli	**Met.no**/MACC	hilde.fagerli@met.no	EMEP	3 – 4D var (in development)	European scale forecasts and assessment (25km)
Valentin Foltescu	**SMHI**/MACC	Valentin.Foltescu@smhi.se	MATCH	2 – 4D (in development)	European to Urban scale (25 - ? km)
Sébastien Massart	**CERFACS**/MACC	massart@cerfacs.fr	MOCAGE/PALM	3 -4D var	Global to European
Bruno Sportisse	**INRIA,CEREA**	Bruno.Sportisse@inria.fr	Polyphemus	3 -4D var, OI, EnKF	European
John Stedman	**AEAT**	John.stedman@aeat.co.uk	ADMS	Statistical interpolation, residual kriging	UK wide assessment of air quality
Bruce Denby	**NILU**/ETC-ACC	bde@nilu.no	EMEP, LOTOS-EUROS	Statistical interpolation, residual kriging	European wide assessments at 10 km
Jan Horálek	**CHMI**/ETC	horalek@chmi.cz	EMEP	Statistical interpolation, residual kriging	European wide assessments at 10 km
Dennis	**JRC Ispra**	Dimosthenis.SARIGIAN	CTDM+ (model	Data fusion	Urban scale

Sarigiannis		NIS@ec.europa.eu	not important, platform more relevant) ICAROS NET	(unknown methodology)	
Marta Garcia Vivanco Palomino Marquez Inmaculada Fernando Martín	**CIEMAT**	m.garcia@ciemat.es inma.palomino@ciemat.es fernando.martin@ciemat.es	MELPUFF CHIMERE	Anisotropic inverse distance weighting Regression and residual kriging.	Assessment Spain
Clemens Mensink Stijn Janssen	**VITO**	stijn.janssen@vito.be Clemens.mensink@vito.be	RIO and BelEUROS	Detrended kriging. Land use regression model used for downscaling CTM	Belgium (3km)
J.A. van Jaarsveld	**RIVM**	hans.van.jaarsveld@rivm.nl	OPS	Kriging with external drift	Nederland (5km)
Florian Pfäfflin (Goetz Wiegand Volker Diegmann)	**IVU Umwelt GmbH**	fpf@ivu-umwelt.de	FLADIS/ IMMISnet/ EURAD	Optimal interpolation	Ruhr, Germany (5km)
Arno Graff	**Umwelt Bundes Amt, UBA II**	arno.graff@uba.de	REM-CALGRID	Optimal interpolation	Germany
Wolfgang Spangl	Umweltbundesamt	Wolfgang.spangl@umweltbundesamt.at		Representativeness of monitoring data	
Sverre Solberg	**NILU**/EMEP	sso@nilu.no	EMEP	Representativeness of monitoring data	EMEP monitoring network

APPLICATIONS RELEVANT FOR THE AIR QUALITY DIRECTIVES

Within the reporting requirements of the AQ Directive three applications are most relevant. These are:

- Assessment of air quality for the reporting of current air quality and the determination of exceedance areas
- Near-real time assessment, including forecasting, for public information and alerts
- Source apportionment

The most relevant application of data fusion methods is in air quality assessment and exceedance estimation. Though monitoring is considered the most reliable method for assessing air quality, its spatial coverage is exceedingly limited and the application of models is extremely useful for extending the spatial coverage.

Near real-time assessment may also make use of both monitoring and modelling data. Currently near real time spatial mapping, e.g. EEA ozone web (www.eea.europa.eu/maps/ozone/), is based on monitoring data. However a number of models involved in the MACC project (www.gmes-atmosphere.eu/) also make use of data assimilation methods in providing regional scale forecasts. There are currently no examples where the combined use of monitoring and modelling is applied for near real time assessments and forecasts of air quality on the local (street) or urban scale.

Source apportionment is one of the major applications of data assimilation methods. This can be achieved through simpler linear regression methods (Laupsa et al., 2009; Denby, 2009) which optimally fit modelled source contributions to observed concentrations or through complete inverse modelling applications of 4D-variational methods (Elbern, 2007).

CONCLUSION

The activities of FAIRMODE, in particular SG1-WG2 on data assimilation, will continue. Currently the main activity is to develop an overview of methods used for combining modelling and monitoring data and to identify institutes and projects in Europe that are currently engaged in these activities. Having identified these, and a number of the challenges facing data assimilation, a discussion concerning the need for quality assurance and the 'fitness for purpose' is required. The aim here is to help disseminate expertise and provide a focus and framework for further development and discussion. This paper is the first step in this process.

REFERENCES

Blond, N., Bel L., Vautard R., 2003. Three-dimensional ozone data analysis with an air quality model over the Paris area. Journal Geophysical Research, 108(D23), 4744, doi:10.1029/2003JD003679

Denby, B., Horálek, J., Walker, S.E., Eben, K. and Fiala, J. (2005) Interpolation and assimilation methods for European scale air quality assessment and mapping. Part 1: Review and recommendations. ETC/ACC Technical Paper 2005/7. December 2005. Final draft. http://air-climate.eionet.europa.eu/reports/reports/ETCACC_TechnPaper_2005_7_spatial_AQ_interpol_Part_I

Denby, B. and Pochmann, M. (2007). Basic data assimilation methods for use in urban air quality assessment. In: Proceedings of the 6th International Conference on Urban Air Quality. Limassol, Cyprus, 27-29 March 2007. Ed. by R.S. Sokhi and M. Neophytou. Hatfield, University of Hertfordshire(CD-ROM).

Denby B., M. Schaap, A. Segers, P. Builtjes and J. Horálek (2008). Comparison of two data assimilation methods for assessing PM10 exceedances on the European scale. Atmos. Environ. 42, 7122-7134.

Denby B. (2008). Source apportionment of PM2.5 in urban areas using multiple linear regression as an inverse modelling technique. Proceedings of the 12th international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, Cavtat, Croatia, 6-10 October 2008. Croat. Met. J., 43, 31-36.

Denby, B., V. Garcia and D. Holland (2009). Integration of Air Quality Modeling and Monitoring Data for Enhanced Health Exposure Assessment. EM Magazine, October 2009.

EC Working group for ambient air quality directives (2000). Guidance on assessment under the EU air quality directives - final draft. (http://europa.eu.int/comm/environment/air/pdf/guidanceunderairquality.pdf)

Elbern, H. and H. Schmidt (1999): A four-dimensional variational chemistry data assimilation scheme for eulerian chemistry transport modeling. JGR, 104, 18583-18598.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O. (2007). Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys. Discuss., 7, 1725-1783.

Evensen, G. (1994) Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. J. Geophys. Res., 99 (C5), 10143-10162.

Flemming J., E. Reimer and R. Stern (2002) Data Assimilation for CT-Modelling Based on Optimum Interpolation. Air Pollution Modeling and Its Application XV. Eds. Carlos Borrego and Guy Schayes. Springer US. pp 255-263

Fuentes, M., and Rafter, A. (2005). Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models. Biometrics 61, 36-45.

Geer, A.J. , W.A. Lahoz, S. Bekki, N. Bormann, Q. Errera, H.J. Eskes, D. Fonteyn, D.R. Jackson, M.N. Juckes, S. Massart, V.-H. Peuch, S. Rharmili and A. Segers, 2006 : The ASSET intercomparison of ozone analyses : method and first results, Atmos.Chem.Phys., 6, 5445-5474.

Gelfand, A. E., and Sahu, S. K. (2008). Combining Monitoring Data and Computer Model Output in Assessing Environmental Exposure. In: The Handbook of Bayesian Analysis. (O'Hagan A,West M, eds). Oxford: Oxford University Press.

Hogrefe, C., B. Lynn, R. Goldberg, C. Rosenzweig, E. Zalewsky, W. Hao, P. Doraiswamy, K. Civerolo, J.Y. Ku, G. Sistla, and P.L. Kinney (2009) Combined Model-Observation Approach to Estimate Historic Gridded Fields of PM2.5 Mass and Species Concentrations. Atmospheric Environment. Volume 43, 2561-2570

Horálek J., P. Kurfürst, P. de Smet, F. de Leeuw, R. Swart, T. van Noije, B. Denby and J. Fiala (2007) Spatial mapping of air quality for European scale assessment. ETC/ACC Technical Paper 2006/6. http://air-climate.eionet.europa.eu/reports/ETCACC_TechnPaper_2006_6_Spat_AQ

Kassteele, van de, J., A. Stein, A. L. M. Dekkers and G. J. M. Velders (2007). External drift kriging of NOx concentrations with dispersion model output in a reduced air quality monitoring network. Environmental and Ecological Statistics, online publication. DOI: 10.1007/s10651-007-0052-x.

Kassteele, van de J., R. B. A. Koelemeijer, A. L. M. Dekkers, M. Schaap, C. D. Homan and A. Stein (2006) Statistical mapping of PM10 concentrations over Western Europe using secondary information from dispersion modelling and MODIS satellite observations. Stoch Environ Res Risk Assess, Vol 21, pp. 183–194

Laupsa, H., Denby, B., Larssen, S. and Schaug, J. (2009). Source apportionment of particulate matter (PM2.5) in an urban area using dispersion, receptor and inverse modelling. Atmospheric Environment. 43, 4733-4744. doi:10.1016/j.atmosenv.2008.07.010.

Lorenc, A. (1986) Analysis methods for numerical weather prediction. Quart. J. Roy. Meteor. Soc. 112, 1177-1194.

Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H. (2007). Technical Note: The air quality modeling system Polyphemus, Atmos. Chem. Phys. Discuss., 7, 6459-6486

McMillan, N., Holland, D. M., Morara, M., and Feng, J. (2009). Combining numerical model output and particulate data using Bayesian space-time modeling. Environmetrics, (21), 48 - 65

Sahu, S., Yip, S., and Holland, D. M. (2009). Improved space-time forecasting of next day ozone concentrations in the eastern U.S.. Atmospheric Environment, 43, 494–501.

Sarigiannis, D. A., N. A. Soulakellis, and N. I. Sifakis (2004). Information Fusion for Computational Assessment of Air Quality and Health Effects. Photogrammetric Engineering & Remote Sensing Vol. 70, No. 2, pp. 235–245.

Schaap, M., M. Roemer, F. Sauter, G. Boersen, R. Timmermans and P.J.H. Builtjes (2005). LOTOS-EUROS Documentation,

TNO report, B&O 2005/297, 2005.

Tarrasón, L., Semb, A., Hjellbrekke, A.-G., Tsyro, S., Schaug, J., Batnicki, J. and Solberg, S.: (1998), Geographical distribution of sulphur and nitrogen compounds in Europe derived both from modelled and observed concentrations. The Norwegian Meteorological Institute, Oslo, Norway (EMEP/MSC-W Note 4/98).

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