TECHNICAL MEMORANDUM

571

Aerosol analysis and forecast in the ECMWF Integrated Forecast System: Data assimilation

A. Benedetti[1], J.-J. Morcrette[1], O. Boucher[2], A. Dethof[1], R.J. Engelen[1], M. Fisher[1], H. Flentjes[3], N. Huneeus[4], L. Jones[1], J.W. Kaiser[1], S. Kinne[5], A. Mangold[6], M. Razinger[1], A.J. Simmons[1], M. Suttie[1], and the GEMS-AER team

[1]European Centre for Medium-Range Weather Forecasts, Reading, UK
[2]Met Office, Exeter, United Kingdom
[3] Deutscher Wetterdienst (DWD), Offenbach, Germany
[4]Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France
[5] Max-Planck-Institut für Mathematik, Bonn, Germany
[6] Royal Meteorological Institute of Belgium, Brussels, Belgium

Research Department

To be submitted to *Journal of Geophysical Research*

August 2008

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

This study presents the new aerosol assimilation system developed at the European Centre for Medium–Range Weather Forecasts for the Global and regional Earth-system Monitoring using Satellite and in-situ data (GEMS) project. The aerosol modelling and analysis system is fully integrated in the operational four–dimensional assimilation apparatus. Its purpose is to produce aerosol forecasts and reanalysis of aerosol fields using optical depth data from satellite sensors. This paper is the second of a series which describes the GEMS aerosol effort and focuses on the theoretical architecture and practical implementation of the aerosol assimilation system. It also provides a discussion of the background errors and observations errors for the aerosol fields, and presents a subset of results from the two–year reanalysis which has been run for 2003 and 2004 using data from the Moderate Resolution Imaging Spectroradiometer on the Aqua and Terra satellites. Independent datasets are used to show that, despite some compromises that have been made for feasibility reasons in regards to the choice of control variable and error characteristics, the analysis is very skillful in drawing to the observations and in improving the forecasts of aerosol optical depth.

1. Introduction

Environmental monitoring is a fundamental activity in current times of rapid transformations of the natural environment due to human activity. In particular, monitoring of greenhouse gases, reactive gases and atmospheric particulate plays an important role due to the open–ended debate about climate change and its long–term implications (Bellouin et al. 2008). Issues raised by the proven links between some atmospheric constituents, such as ozone and particulate, and human health have also raised the level of attention toward these activities (Thompson et al. 2006; Lewtas 2007).

One of the fore-front projects dedicated to this goal is the Global and regional Earth-system (Atmosphere) Monitoring using Satellite and in-situ data (GEMS) project, which counts thirty-two European partners with expertise in various aspects of atmospheric composition monitoring. GEMS is part of the Global Monitoring for Environment and Security (GMES) initiative and was established under European Commission funding in 2005 to create an assimilation and forecasting system for monitoring aerosols, greenhouse gases and reactive gases, at global and regional scales, through exploitation of satellite and in–situ data (Hollingsworth et al. 2008). An important component of GEMS is also monitoring of regional air quality at the European scale which is performed with an ensemble of models from the participating institutes. Boundary conditions for the high–resolution models are provided by the global model.

The forecast and analysis systems which include atmospheric constituents have been developed. The basis for these systems is the operational ECMWF Integrated Forecasting System (IFS[1]) and the incremental 4D-Var system, extended to include new prognostic variables for the atmospheric tracers (i.e. gases and aerosols). A coupled chemical transport model is also part of the GEMS system and provides tendencies for the chemically–active species which are present in the model. In this paper, we will focus on the development and the performance of the aerosol analysis system. Companion papers by Morcrette et al. (2008) and by Mangold et al. (2008) discuss in details the aerosol model and the validation of the forecast/analysis results respectively.

Modelling and prediction of aerosols is associated with a large degree of uncertainty due to uncertainties in the emissions, transport and nonlinear physical processes involving aerosols (for example, radiative effects, cloud and rain formation, etc.). Ground–based observing networks have been crucial in improving our knowledge

[1]The documentation for the ECMWF Integrated Forecasting System is available online at www.ecmwf.int/research/ifsdocs/.

of this atmospheric component, complemented in more recent years by satellite sensors which offer a more global view of the aerosol distribution. Harmonization of models and data is, however, required in order to tackle the model deficiencies and obtain a more accurate representation of aerosols and their interaction with the atmospheric system as a whole.

The current attempt at variational data assimilation of satellite aerosol data into a Numerical Weather Prediction (NWP) model for reanalyses and forecasts is an unprecedented effort. Previous applications in this relatively young field, aimed at assimilating aerosol in global models use the Optimal Interpolation approach with focus on regional studies (Collins et al. 2001; Rasch et al. 2001). A global assimilation OI system is described in Generoso et al. (2007) and used to better constrain the Arctic aerosol burden. Weaver et al. (2007) describe an off–line retrieval/assimilation system for the Goddard Chemistry and Aerosol Radiation Transport (GOCART) model using Moderate Resolution Imaging Spectroradiometer (MODIS) radiances based on the Kalman filter approach. Successful 4D–Var aerosol assimilation has been implemented in a chemical transport model by Fonteyn et al. (2000) and Errera and Fonteyn (2001). More recently, Zhang et al. (2008) described the first attempt at building a 3D–Var system for operational aerosol assimilation using the Naval Research Laboratory (NRL) model, while Niu et al. (2008) introduced a novel data assimilation system for dust aerosols in the Chinese Unified Atmospheric Chemistry Environment- Dust (CUACE/Dust) forecast system.

The scheme presented in this paper represents the European effort at building a pre–operational system for routine aerosol assimilation and forecasting at global level. The aerosol assimilation is fully integrated as an option within the general 4D–Var system employed operationally at ECMWF, and the aerosol model which underlies the assimilation is likewise integrated in the forecasting model. All aerosol species are model prognostic variables, advected by the semi-lagrangian scheme, consistently with other dynamical fields, and treated as tracers in the vertical diffusion and convection schemes. They are also subject to aerosol–specific physical processes such as sedimentation and wet/dry deposition. In the current model version, species included are sea salt, desert dust, black carbon, organic matter and sulphate. The emission sources for the various aerosol species are defined either using established emission inventories or through model–dependent parameterizations. The observational operator for aerosol optical depth uses the mass mixing ratios and pre-computed optical properties according to the aerosol species at the specific wavelength. The control variable is the total aerosol mixing ratio defined as the sum of the single contributing species. Background error statistics for this variable are computed with the NMC method and the background error covariance matrix is constructed using the "wavelet-J_b" approach. Errors for observations over ocean are defined using a parameterization which is function of the viewing geometry through a dependence on the scattering angle, while for observations over land an *ad hoc* percentage error is used. These errors are inflated with respect to the MODIS product specifications to account for representativeness errors deriving from the forward model and the observation processing.

The aerosol assimilation system has been recently completed and is currently being used to produce a two–year analysis for 2003–2004. Observations assimilated are the aerosol optical depths (AOD) retrieved from the MODIS instruments on board the Terra and Aqua satellites. The building blocks of the system are described in section 2. Section 3 presents the observations used in the reanalysis along with a discussion of biases and representativeness errors. The experimental set–up is also included in this section. Results for 2003 are examined in section 5 with focus on the month of May 2003. A validation of these results using independent observations from the AErosol RObotic NETwork (AERONET) is also presented. It is demonstrated that the GEMS aerosol assimilation system has good potential to provide high–quality analysis and forecasts of atmospheric particulate.

2. Technical description of the aerosol assimilation system

a. *The aerosol model*

The implementation of an aerosol module in the ECMWF model has involved the introduction of new prognostic variables (i.e. aerosol mass mixing ratios) and the definition of aerosol–specific physical parameterisations (Morcrette et al. 2006). The physical package for aerosols was partially taken from the Laboratoire d'Optique Atmosphérique (LOA) Laboratoire de Météorologie Dynamique (LMD) model (Boucher et al. 2002; Reddy et al. 2005). It includes sources for sea salt and desert dust and a representation of sedimentation, and wet and dry deposition processes. The sedimentation scheme has been modified following recent developments by Tompkins (2005) while the wet and dry deposition schemes were adapted directly from the LMD model. All aerosol species are treated as tracers in the IFS vertical diffusion and convection schemes and are advected by the semi-Lagrangian scheme, consistently with all other dynamical fields and tracers. Five types of tropospheric aerosols are included: sea salt, desert dust, organic matter, black carbon and sulphate aerosols. Stratospheric aerosols are not included in the current assimilation configuration.

Aerosols of natural origin are represented via a three-bin representation. Bin limits for sea salt are set at 0.03, 0.5, 5 and 20 μm and for desert dust at 0.03, 0.55 0.9 and 20 μm. This ensures that approximately 10, 20 and 70 % of the total mass is respectively included in the three bins. For organic matter and black carbon both the hydrophobic and the hydrophilic component are modelled. Sulphates are represented as one variable with no explicit chemistry.

State-of-the-art emission sources have been implemented (Morcrette et al. (2006, 2008)). For anthropogenic aerosol, the sources come from available emission inventories (GFED, Global Fire Emission Database; SPEW, Speciated Particulate Emission Wizard; EDGAR, Emission Database for Global Atmospheric Research). For natural aerosol, the sources are instead related to model parameters (i.e. 10-m wind for sea salt, soil moisture and wind for desert dust, among others). The aerosol model provides the background information to feed into the variational assimilation system described below.

b. *The ECMWF 4D–Var*

The variational method is a well–established approach to combine model background information and observations to obtain the "best" forecast possible by adjusting the initial conditions. This approach is widely used in many NWPs centres. The method is based on minimization of a cost function which measures the distance between observations and their model equivalent, subject to a background constraint usually provided by the model itself. Optimization of this cost function is performed with respect to selected control variables (e.g. the initial conditions). Adjustments to these control variables allow for the updated model trajectory to match the observations more closely. Assuming the update to the initial condition (also known as the increment) is small, an incremental formulation can be adopted to ensure a good compromise between operational feasibility and physical consistency in the analysis (Courtier et al. 1994). The cost function in the incremental approach can be written as:

$$J(\delta x_0) = \frac{1}{2}\delta x_0^T \mathbf{B}^{-1}\delta \mathbf{x}_0 + \frac{1}{2}\sum_{i=0}^{n}\left(H'_i\delta x_i - \mathbf{d}_i\right)^T \mathbf{R}_i^{-1}\left(H'_i\delta x_i - \mathbf{d}_i\right) + J_c(\delta x_0), \quad (1)$$

where $\delta \mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_i^b$ is the analysis increment and represents the departure of the model state ($\mathbf{x}$) with respect to the background ($\mathbf{x}^b$) at any time t_i. H' is the linearized observation operator and $\mathbf{d}_i = \mathbf{y}_i^o - H_i(\mathbf{x}^b{}_i)$ is the departure of the model background equivalent from the observation ($\mathbf{y}_i^o$). The matrix $\mathbf{R}_i$ is the observation error covariance matrix, which comprise both pure observation errors (instrumental, calibration, retrieval) and representativeness

errors due to forward model assumptions and to the interpolations needed to go from model to observation space. **B** represents the background error covariance matrix, formulated according to the "wavelet–J_b" method of Fisher (2003; 2004). The aerosol–specific background error covariance matrix is discussed briefly in section 2c and more extensively in Benedetti and Fisher (2007). The **R** and **B** matrices represent respectively the relative weight assigned to observations and background fields in the analysis. The background at $t = 0$, $\mathbf{x}_0^b$, is obtained from a short-range forecast valid at the initial time of the assimilation period. A penalty cost function, $J_c(\delta\mathbf{x}_0)$, is used to impose all other physical constraints on the solution.

The flow of the 4D–Var minimization is as follows. A nonlinear integration provides the linearization state (trajectory) in the vicinity of which the model is linearized. The departures are computed during the nonlinear integration at high resolution and then interpolated to the lower resolution. The gradient of the cost function required in the minimization is computed using the adjoint model.

The minimization is solved using an iterative algorithm, based on the Lanczos conjugate gradient algorithm with appropriate pre-conditioning. In order to reduce the computational costs and strong non-linearities in the operational 4D–Var system, the perturbations $\delta\mathbf{x}_i$ are computed with a tangent-linear model using simplified physics (Tompkins and Janisková 2004; Lopez and Moreau 2005) at a lower resolution than the trajectory.

After the minimization, the trajectory and the departures are recomputed and a second minimization at a higher horizontal resolution is run. For the analysis a resolution of T159 (corresponding to $\sim$ 120 km) is used in the nonlinear trajectory and the forecast, while the two minimizations are run at T95 ($\sim$ 215 km) and T159. An average of 50–70 iterations are required to reach a satisfactory convergence of the minimization. Convergence criteria and a detailed description of the incremental 4D–Var can be found in Fisher (1998) and Trémolet (2005).

The current assimilation window is 12 hours. MODIS observations of aerosol optical depth are ingested over the window and sub–divided into time slots of half hour, together with all other meteorological data. A thinning is applied to better match the spatial resolution of the observations to that of the analysis.

The model fields, including aerosol mass mixing ratios, required by the operators are interpolated at the observation locations and the model equivalent of the observations is computed using the specific observation operator. The operator for AOD is described in section 2d.

c. *The background error covariance matrix for aerosols*

The aerosol **B** matrix used for the GEMS aerosol reanalysis was derived using the Parrish and Derber method (also known as NMC method, Parrish and Derber (1992)) as detailed in Benedetti and Fisher (2007). The difference with respect to the results presented in that paper lies in the use of updated error statistics derived from forecast differences computed with the current aerosol model described in section 2a. The method is the following: six months of 2-day forecasts at T159 are run and the differences between the 48-h and the 24-h forecasts are used as statistics for the estimate of the background errors. These are in turn used to construct a **B** matrix using the wavelet technique devised by Fisher (2003, 2006). In section 4a the sets of statistics from the run of Benedetti and Fisher (2007) and the current configuration are presented. Their impact on the aerosol analysis is also discussed.

Benedetti and Fisher (2007) showed that the NMC method leads to a satisfactory background error covariance matrix without the need to prescribe the vertical and horizontal correlation lengths. The NMC method applied to the definition of background error statistics for aerosol has also been revisited and compared with other methods by Kahnert (2008) who concluded that it is the most appropriate for assimilation over the time windows typically used in NWP applications (6-12 hours).

d. *Observation operator*

The observation operator for aerosol optical depth is based on pre-computed optical properties (mass extinction coefficient, α_e, single scattering albedo, ω, and asymmetry parameter, g) for the relevant aerosol species included in the model. The aerosols are assumed to be externally mixed, i.e. the individual species are assumed to co-exist in the volume of air considered and to retain their individual optical and chemical characteristics. These characteristics are computed at the MODIS wavelengths using Mie theory, e.g. particles are assumed to be spherical in shape, and integrated over the physical size range using a prescribed lognormal distribution (Reddy et al. 2005). Optical properties of hygroscopic aerosols such as sulphate, hydrophilic organic matter and sea salt, are parameterized as a function of relative humidity (RH). Table 1 summarizes the optical properties at 550 nm and 50% RH.

Table 1: Optical properties at 550 nm and 50% Relative Humidity.

Aerosol type	α_e (m^2g^{-1})	ω	g
Sulphate	6.609	1.000	0.673
Black Carbon	9.412	0.206	0.335
Organic matter	5.502	0.982	0.655
Dust			
(0.03-0.55 μm)	2.6321	0.9896	0.7300
(0.55-0.9 μm)	0.8679	0.9672	0.5912
(0.9-20.0 μm)	0.4274	0.9441	0.7788
Sea salt			
(0.03-0.5 μm)	3.0471	0.9996	0.7394
(0.5-5.0 μm)	0.3279	0.9961	0.7703
(5.0-20.0 μm)	0.0924	0.9916	0.8224

For the calculation of the model equivalent optical depth, the relative humidity is computed from model temperature, pressure and specific humidity. The appropriate mass extinction coefficients are then retrieved from the look–up table for the wavelength of interest (here, 550 nm), multiplied by the aerosol mass which has been previously interpolated at the observation locations, and then integrated vertically. The total optical depth is the sum of the single–species optical depths as given by

$$\tau_\lambda = \sum_{i=1}^{N} \int_{p_{surf}}^{0} \alpha_{ei}(\lambda, RH(p)) r_i(p) \frac{dp}{g}, \tag{2}$$

where N is the total number of aerosol species, r is the mass mixing ratio, dp is the pressure of the model layer and g is the constant of gravity. p_{surf} represents the surface pressure.

e. *Total aerosol mixing ratio as control variable*

The aerosol model comprises a mixed bin and bulk representation of the aerosol species. This was deemed to be a necessary compromise between a full–blown bin representation of all species which would have introduced many more tracers in the IFS, and a modal representation of the aerosols which would have possibly over–simplified the aerosol model. However, the eleven additional tracers that are currently used in the forward model, can constitute a heavy burden for the analysis if they are all included in the control vector. The reason for this is three–fold: (i) background error statistics would have to be generated for all species separately, (ii) the control vector would be much larger in size which would, in turn, increase the cost of the iterative minimiza-

tion, and most importantly (iii) the aerosol analysis would be severely under–constrained as one observation of total aerosol optical depth would be used to constrain eleven profiles of aerosol species. As a way to alleviate these problems, the total aerosol mixing ratio, defined as the sum of the eleven aerosol species, is used instead as control variable. The increments in the total mixing ratio deriving from the assimilation of MODIS optical depths have to be re–distributed into the mixing ratios of the single species. Even with this expedient, the problem remains under–constrained with respect to the observations, and the redistribution of the increments relies heavily on the background. However, the size of the control vector is more manageable. Some assumptions are needed in order to implement this control variable correctly:

1. the sum of the single species have to be equal to the total mixing ratio at all times and for all locations, i.e. the aerosol mass needs to be conserved over the 12–hour assimilation window;

2. the relative contribution of a single species/bin to the total mixing ratio has to be kept constant over the assimilation window.

Assumption (1) implies that processes which do not conserve the aerosol mass, such as deposition and sedimentation, should not be activated during the trajectory run. Assumption (2) follows from (1), and effectively implies that perturbations from species with higher specific density contribute more to the perturbation in total mixing ratio even if their contribution to the optical thickness at a given wavelength might be smaller than that of species with lower specific density. In practice these assumptions are relaxed and the trajectory run is performed with all aerosol processes switched on. This still provides a meaningful analysis since most of the dominant physical processes happen over time–scales longer than 12 hours. For example, the typical residence time for the largest bin of desert dust and sea salt is approximately one day, whereas anthropogenic species have a typical residence time of a week.

The way this control variable works in practice is the following. In the nonlinear trajectory run the total aerosol mixing ratio is computed by summing all other species/bins, i.e. $r_T = \sum_{i=1}^{N} r_i$, where r is the mixing ratio, and the subscript T indicates the total mixing ratio. All aerosol variables, including the total aerosol mixing ratio, are subject to advection, vertical diffusion and convection. The mixing ratios of the individual species are used to compute the total optical depth using the tabulated optical properties as outlined in section 2d. The tangent linear run is then started with zero perturbations for the single species to compute the perturbation in optical depth. The latter is passed to the adjoint routine to compute the gradient with respect to the individual species. The gradient with respect to the total mixing ratio is obtained as

$$r_T^{'} = \sum_{i=1}^{N} f_i r_i^{'} \tag{3}$$

where $r^{'}$ is gradient of the mixing ratio and $f_i = \left(\frac{r_i}{r_T}\right)$ is the fractional contribution of the single species to the total mass. The gradient with respect to the total mixing ratio is then used in the minimization and the resulting increment in r_T is used in the following iteration of the tangent linear run to compute updated perturbations on the individual species/bin mixing ratios as follows

$$r_i^{'} = f_i r_T^{'}. \tag{4}$$

These, in turn, are used to compute perturbations in optical depth to be fed to the adjoint, and so on until the convergence criteria is met. To avoid the analyzed total aerosol mixing ratio becoming negative as a result of adding a negative increment, the total aerosol mixing ratio is screened for values less than zero and reset to zero when those happen.

3. Data description and experimental setup

a. *MODIS aerosol retrievals*

Of the available satellite data sources for aerosol optical depth data, the MODIS instrument on board of Terra and Aqua was selected for its accuracy and reliability. The availability of data in near real–time was a further factor. These are important aspects in view of an operational application. The retrievals of aerosol optical depth from MODIS are described in Remer et al. (2005). Two separate retrievals with different accuracies are applied over land and ocean. The retrievals over land suffers from higher uncertainties due to the impact of the surface reflectance, notoriously harder to model over land than ocean. Over highly reflective surfaces such as deserts and snow-covered areas, there is no sufficient contrast to discern the signal coming from the aerosols over that coming from the surface. For this reason, the land retrieval is only possible over "dark" surfaces. Several other factors affect the accuracy of the retrievals both over land and ocean: cloud contamination, assumptions about the aerosol types and size distribution, near–surface wind speed, radiative transfer biases, and instrumental uncertainties. These factors are reviewed in detail in Zhang and Reid (2006).

For our purposes the most recent MODIS release (collection 5) was used since it has been proven to be more accurate, particularly over land. MODIS retrieved optical depths are provided at different wavelengths These are 470 nm, 550 nm, 660 nm, 870 nm, 1240 nm, 1630 nm and 2130 nm. However, as a first step, only the optical depth at 550 nm is assimilated in the analysis. In what follows it is understood that aerosol optical depth refers to the aerosol optical depth at a wavelength of 550 nm, unless otherwise stated.

The original data have a resolution of 10x10 km. Since the analysis is run at T159 which is approximately 120x120 km, the MODIS optical depths are thinned to this coarser resolution. Observations are taken at the original location and model aerosol fields are interpolated to this location before applying the observation operator described in section 2d.

b. *Discussion of observations and model biases*

Observations and model biases are very important to characterize for a successful analysis as the analysis itself does not remove biases, but only aims at minimizing the error between model and observations in a least square sense. Zhang and Reid (2006) propose a method to remove biases from the MODIS ocean aerosol product before assimilation in the NRL system through quality assurance procedures and selective data screening. They indicate that the reduction in error between the corrected MODIS and the AERONET verifying data can be 10-20%, mainly due to the elimination of the cloud–contaminated retrievals.

While recognizing the validity of this effort we did not apply a similar rigorous procedure to the MODIS data. Our approach was instead to devise a correction dependent on the model optical depth as described in Benedetti and Janisková (2008) for the assimilation of MODIS cloud optical depths. In that study, the authors divide the range of possible optical depths into eighteen bins and for each bin they calculate the average of the corresponding first guess departures. The averages are then stored and subsequently subtracted during the assimilation run from the model optical depths falling in the specific bin. One of the shortcomings of this method is that model biases can be aliased into observation biases. We applied this procedure to the aerosol analysis and it was found that for aerosol optical depth this bias correction does not improve the analysis. This is possibly due to the issue highlighted also in Zhang et al. (2008) that checks based on first guess departures do not flag cases in which both the observations and the first guess have large biases but the difference between the two is small. It was hence decided not to implement any bias correction. All results presented here are from an analysis with the raw MODIS optical depth data. The issue of a bias correction for the MODIS aerosol retrievals is still open and will be addressed in the future.

c. *Observation and representativeness errors*

The overall accuracy of the MODIS aerosol optical depth product is given in Remer et al. (2005). The ocean retrievals are more accurate with an estimated uncertainty $\Delta\tau = \pm 0.03 \pm 0.05\tau$. The land retrievals are assigned an estimated $\Delta\tau = \pm 0.05 \pm 0.15\tau$. The authors also quote a relative error between the MODIS land retrievals and AERONET observations of 41% at 0.55 μm where MODIS shows a positive bias and overestimates τ.

In this study the error on over–ocean retrievals of aerosol optical depth at 0.55 μm was re-assigned following Crouzille et al. (2007). In their study the authors analyze the MODIS retrievals to devise a multi–regression formula for assigning errors as a function of the scattering angle at a pixel level. They make use of the quality flags provided as part of the standard MODIS product to choose and include only "good" retrievals in the regression. Following their analysis, the standard deviation on the aerosol product over ocean, ε_{τ_o} can then be parameterized as

$$\varepsilon_{\tau_o} = max(0.05, \tau_o(a + b * \Theta) + c) \tag{5}$$

where $a = 0.007$, $b = 0.0012$ and $c = 0.001$ are the regression coefficients, τ_o is the aerosol optical depth over ocean, and Θ is the scattering angle. In the current study a slightly different formulation was used. As it was noticed that the free-running forecast tends to overpredict optical depth over the oceans, the minimum error for the MODIS product was taken to be 0.02 according to the following formula:

$$\varepsilon_{\tau_o} = max(0.02, \tau_o(a + b * \Theta) + c) \tag{6}$$

in order to bind the analysis more to the observations. An extra five percent error was arbitrarily added to account for errors due to the interpolation of the aerosol fields to the observation location (representativeness error).

For the land retrievals, it was decided to assign an arbitrarily inflated error to account for the discrepancies with the AERONET product also mentioned by Remer et al. (2005) and representativeness. The error for land retrievals, ε_{τ_l}, was hence assigned as

$$\varepsilon_{\tau_l} = max(0.02, 0.5\tau_l) \tag{7}$$

where τ_l is the aerosol optical depth over land. The impact of these choices for the errors are discussed in section 4b.

Other sources of representativess error for aerosol optical depth not included in the current formulation are discussed in Tsigaridis et al. (2008). Those are related to the assumptions made on the underlying aerosol model which is used to obtain the optical properties, for example the assumption of sphericity of the aerosol particles, the choice of the size distribution and its parameters (characteristic radius and variance) the preassigned hygroscopic behavior of the aerosols, and most importantly the assumption on the state of mixing of the aerosol particulate, most commonly treated as comprising individual non–interacting chemical species (external mixtures). All these can contribute to increase the error on the optical depth by up to 30%. In the future there will be an attempt to include these uncertainties into the assignment of the observation error, but in the current study these error contributions were neglected.

d. *Experimental setup*

All reanalysis tests and the long reanalyses were run with the same configuration. Species included in the analysis are sea salt, desert dust, black carbon, organic matter and sulphate. It was decided not to include stratospheric aerosol due to the low concentrations for the years 2003–2004. The model resolution was set to T159 and 60 vertical levels. The background error covariance matrix was specified as detailed in section 2c,

while the observation covariance matrix was assumed to be diagonal with standard deviations prescribed by equations (6) and (7).

Initial tests covered a period of ten days in April 2003 and were used to look at the general behaviour of the analysis. The month of May 2003 was used for more extensive statistics on the analysis biases. Lessons learned from these investigations are discussed in section 4.

4. Lessons learned so far

a. *The importance of the background error covariance matrix*

Preliminary runs from test analyses showed anomalous increments in total aerosol mixing ratio, and consequently in aerosol optical depth, over the polar regions. The values produced in the analysis were clearly unrealistic, with optical depths reaching values as large as 3 over the North Pole.

Several experiments were conducted to understand the reason for these large increments, and identify the system sensitivities to changes in the analysis configuration, including data denial experiments, the implementation of aerosol loss parameterizations in the minimization, and the redefinition of the background error statistics. The latter proved to be the solution to the problem. The old background aerosol statistics had been calculated with a preliminary model version which included only sea salt, desert dust and a generic continental background aerosol type; hence they had become outdated for the current aerosol model which includes many more species. Comparisons with the old set of statistics show dramatic differences both in the standard deviation and in the vertical and horizontal correlations. Specifically the new standard deviation is one order of magnitude smaller than the old one, indicating that the model has a smaller degree of variability between the aerosol forecasts at 24 and 48-h (see figure 1).

This, in turn, translates into a smaller background error, which is more realistic given the improvements implemented in the current version of the aerosol model. With this new background error, the analysis, while still drawing to the observations, takes the model background constraint more into account. This is extremely important, especially in areas that are data–limited such as the polar regions where the aerosol analysis is severely under–constrained relatively to the observations and relies entirely on the background.

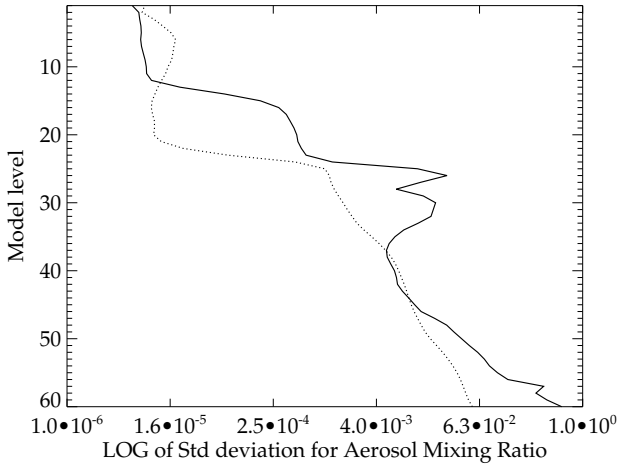


Figure 1: Comparison between the old (solid line) and the recomputed (dotted line) global background standard deviation for total aerosol mixing ratio at Lampedusa, Italy. Units are mg/kg. The x–axis is in logarithmic scale.

The correlations between the two sets of statistics also differ remarkably. Figure 2 shows a comparison between the old correlations and the new correlations for a specific location (Lampedusa, Italy, 35.5N–12.6E). It can be seen that the updated horizontal correlations (top panel on the right) are much broader that the old correlations (top panel on the left) in the stratosphere which is more consistent with observed long–range transport of atmospheric particulate. The lower panels show the vertical correlations. Again there is a large difference in the two sets with the new correlations being much broader than the old correlations in the boundary layer.

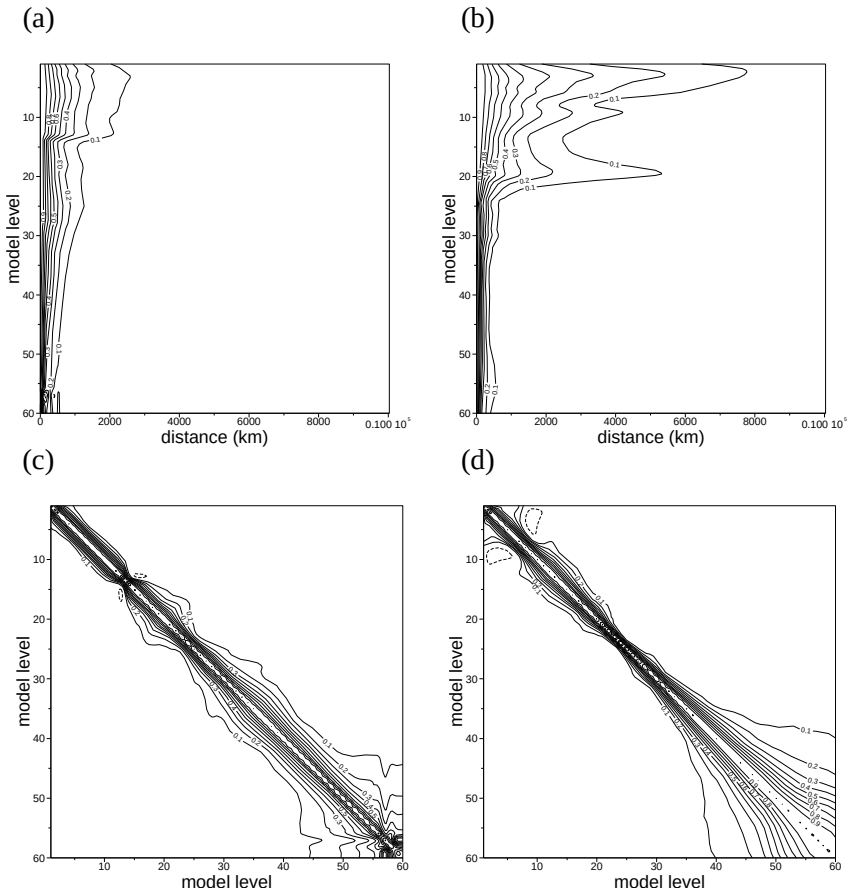


Figure 2: Sample background error correlations for total aerosol mixing ratio at Lampedusa, Italy: (a) "old" horizontal correlations, (b) recomputed horizontal correlations as functions of distance, (c) 'old" vertical correlations, and (d) recomputed vertical correlations as functions of model levels.

The impact of these different background error covariance matrices on the analysis is dramatic. Figure 3 shows the difference between a 10–day analysis for April 20–30 2003, with the old and the new background error covariance matrix. The current reanalysis presents a more realistic distribution of aerosol optical depth at high latitudes in both hemispheres and also over the Southern Hemisphere oceans. This experience shows how fundamental the correct definition of the background error covariance matrix is for the aerosol problem. While *ad hoc* formulated background error covariance matrices may work in specific cases, a system which is potentially pre–operational requires a more robust definition of this important component of variational assimilation.

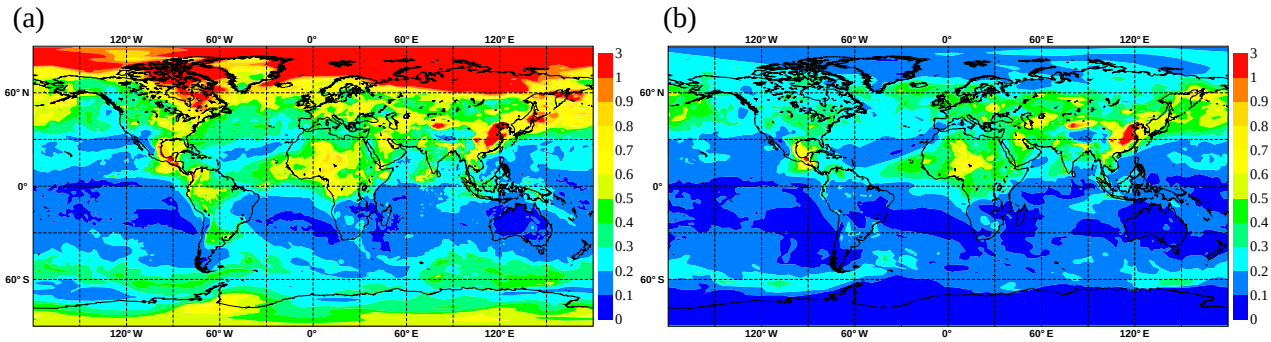


Figure 3: Comparisons of simulated aerosol optical depths for April 20-30, 2003: (a) analysis with old background error covariance matrix; (b) current analysis with updated background error covariance matrix.

b. *Analysis biases*

The aerosol optical depth from the analysis for the whole month of May 2003 was used to investigate the analysis performance with respect to the assimilated observations. This type of comparison is considered a sanity check as in a successful analysis the departures should be smaller than the first guess departures, and the analysis should better match the observations in a statistical sense. Figure 4 shows scatterplots of assimilated aerosol observations versus first guess (top left) and analysis (top right). By visual inspection, it is apparent that the scatter in the analysis is smaller than in the first guess. The root mean square error with respect to the MODIS data is lower for the analysis (0.122) than for the first guess (0.168) while the correlation coefficient is higher for the analysis (0.888) than for the first guess (0.757), indicating a good performance of the analysis. However, while we did not expect the analysis to improve on the first guess biases, it was surprising to notice that the analysis effectively has a larger bias than the first guess. The distribution appears to be skewed and it is evident from the shape of the scatterplot that the analysis is more efficient in increasing low values of optical depth rather than in reducing high values.

This "asymmetric" behaviour of the analysis merits further attention. As a first step we checked whether this bias could be caused by the choice of control variable presented in section 2e. By definition this variable cannot be negative. However it is possible that in areas where the first guess is larger than the observed values, the analysis increments can be large as well and negative, as the analysis attempts to bring the model values closer to the observations. When these negative increments are added to the trajectory values of total aerosol mixing ratio, it is possible that the updated value of mixing ratio becomes negative, and hence unphysical. In the implementation, this is avoided by resetting to zero all negative values of total aerosol mixing ratio. However, this could introduce a bias in the analysis. To further investigate this, a logarithmic total aerosol mixing ratio was implemented as control variable. By construction, this alternative control variable is positive definite and there is no need to reset the analysis total concentration to zero *a posteriori*. The bottom panels of figure 4 show scatterplots for an experiment with the logarithmic total aerosol mixing ratio. We can notice that in this case, the distribution of the points along the 1:1 line is more symmetric. The bias in the analysis is comparable to the first guess (0.018 versus 0.008) while the RMS is still lower (0.157 for the analysis versus 0.182 for the first guess) and the correlation coefficient is higher (0.799 versus 0.707). There is still, however, a tendency of the analysis to be more effective at increasing low values than decreasing large values.

This behavior can be further explained by looking back at the error assumptions for the MODIS optical depths discussed in section 2c. From the error formulations of equations (6) and (7) it appears that high values of optical depths are penalized more, since the error is prescribed as a percent of the optical depth. One possible solution is to implement a capping of the errors on the observations above a certain optical depth threshold,

hence giving more weight to these observations. This will be considered in future re-runs of the analysis along with more stringent quality checks and screening on the ingested data as those employed by Zhang et al. (2008).

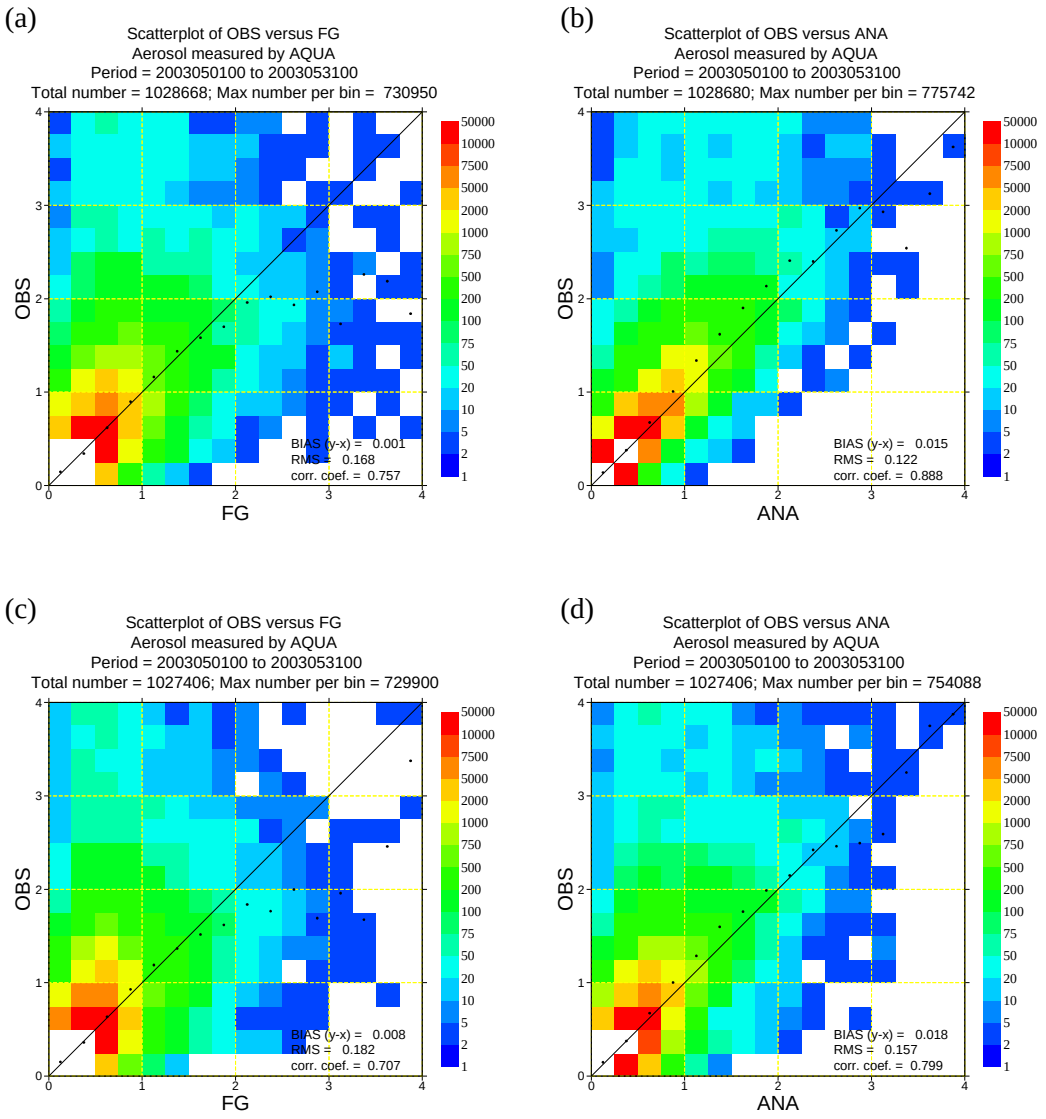


Figure 4: Scatter plots of MODIS Aqua observations of aerosol optical depth versus (a) first guess optical depth and (b) analysis optical depth for the standard reanalysis run for May 2003. (c) same as in (a) and (d) as in (b), but using a logarithmic total aerosol mixing ratio variable in the control vector. See text for explanations. Note that the scatterplots only include MODIS data from the Aqua satellite, whereas data from the Terra satellite were also included in the assimilation.

As a side note, the implementation of the logarithmic variable did not improve dramatically the analysis performance. On the contrary, the RMS is higher and the correlation lower in the analysis with the logarithmic control variable than in the analysis with total mass mixing ratio. The reason for this could lie in the fact that the logarithmic control variable is only used at the level of the minimization, whereas the rest of the model is formulated in terms of mass mixing ratio. A more effective way to handle tracers could be to formulate the whole forward model in terms of logarithmic, hence positive definite, variables. This however would involve an extensive effort in modifying and rewriting the model, and it is not a viable option at this point. The use of alternative normalized control variables with a more Gaussian error distribution can still be investigated for

future developments, following existing examples (Hólm et al. 2002, e.g.).

5. Reanalysis results

The performance of the long reanalysis is assessed globally with comparisons with other optical depth databases both from space–borne sensors (i.e. the Multiangle Imaging SpectroRadiometer, MISR) and from established ground–based sunphotometer networks (AERONET). Comparisons with MODIS data from the Aqua satellite are also shown as reference. A more in–depth validation of the analysis which includes both optical depth and physical property data (aerosol mass concentration) will be presented in Mangold et al. (2008).

a. *Comparisons with MODIS and MISR data*

The MISR instrument (Diner et al. 1998, 2005) measures 4 bands (blue, green, red and near-infrared) at different viewing angles (0., 26.1, 45.6, 60.0, and 70.5 degrees) using 9 cameras. The swath width is approximately 360 km. The global coverage time is 9 days, with repeat coverage between 2 and 9 days depending on latitude. Thanks to the unique viewing geometry, MISR can measure aerosol optical depth over different reflecting surfaces including bright surfaces as deserts. The aerosol optical depth product is quoted to have an accuracy of 20% or 0.05 (whichever is larger) with greater accuracy over dark surfaces (Kahn et al. 2007). Although MISR retrievals cannot be assumed as "ground truth" as they suffer from inaccuracies related to cloud contamination, wind–speed assumptions, etc, they offer an independent platform to assess the forecast and the analysis.

Figure 5 shows comparisons between the "free–running" forecast of aerosol optical depth without any assimilation, the analysis of optical depth from assimilated MODIS observations and the MISR aerosol optical depth for May 2003. MODIS AODs are also shown as reference. Optical depth retrievals are assimilated over both land and ocean. The model aerosol optical depths are averages of three–hourly forecasts started at 00UTC from the free–running model and from the analysis. Figure 6 shows differences between the MODIS and MISR monthly means with respect to the forecast and analysis optical depths. Despite some evident discrepancies, these figures show that the analysis is effective in bringing the model aerosol optical depth closer to the observations especially in areas where the free–running forecast underestimates the AOD.

The assimilation generally improves the aerosol distribution over areas with extensive biomass burning in equatorial West Africa. The aerosol amount in the Southern Ocean is lower in the analysis than in the free–running forecast, also in better agreement with the observations. Other areas such as the Indian Ocean, the Indian subcontinent and Eastern Asia dominated by anthropogenic emissions and not captured as well in the free–running simulation because of inadequate definition of the sources for these emissions, are also improved. Note, however, the overall skill of the forecast model in predicting the global distribution of the aerosol fields thus providing a good first–guess for the analysis. Of note is also the overall large positive bias of the analysis over Eurasia, and the inability of the analysis to constrain areas of large optical depths evident also in the free-running forecast but absent in the observations (e.g. Eastern United States). The possible reasons for this behaviour have already been discussed in section 4b. It is, however, instructive to see the geographical distribution of this bias to pinpoint in which areas the analysis can be improved both through the use of observations with better coverage and of higher quality and through refinements in the methodology.

The figures also highlight discrepancies between the two satellite products which can be as large as the largest departures in the free–running model and in the analysis.

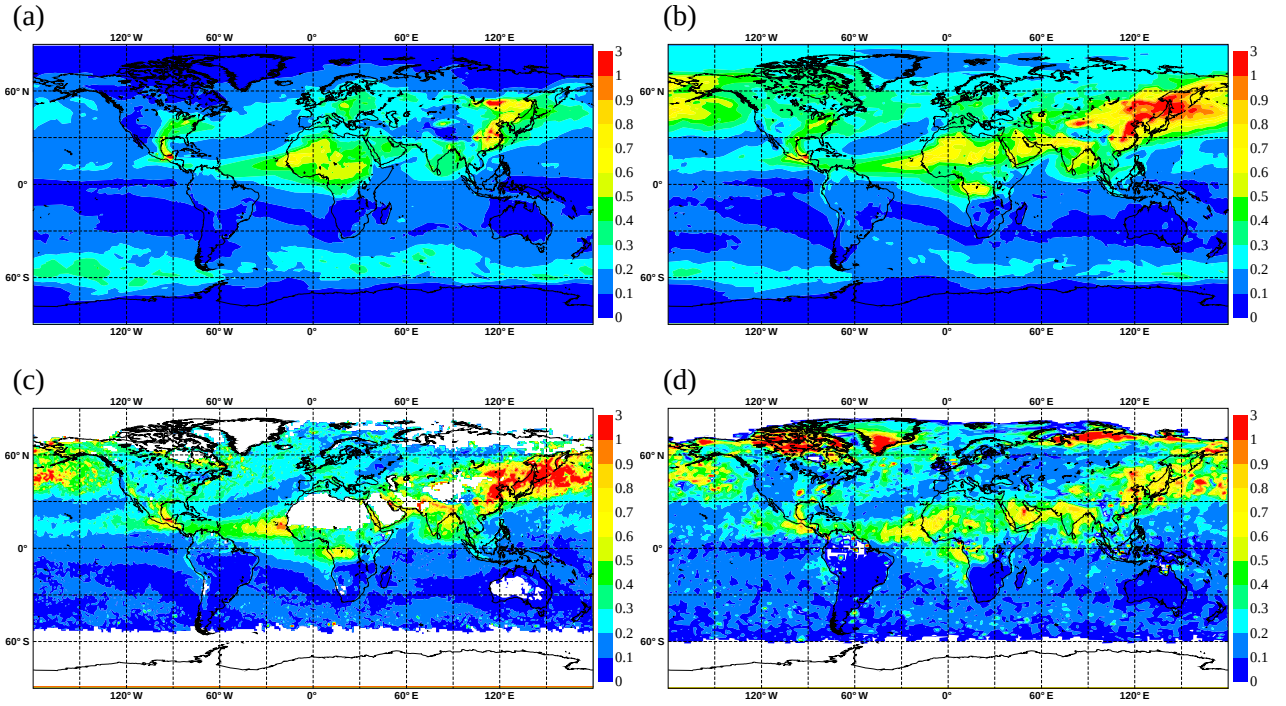


Figure 5: Comparisons of simulated aerosol optical depths from the new aerosol module implemented in the IFS model: (a) free–running forecast; (b) analysis using MODIS collection 5 observations; (c) MODIS AQUA and (d) MISR aerosol optical depth for May 2003. The high values of AOD over Greenland, and other northern high–latitude areas in the MISR data plot should be disregarded.

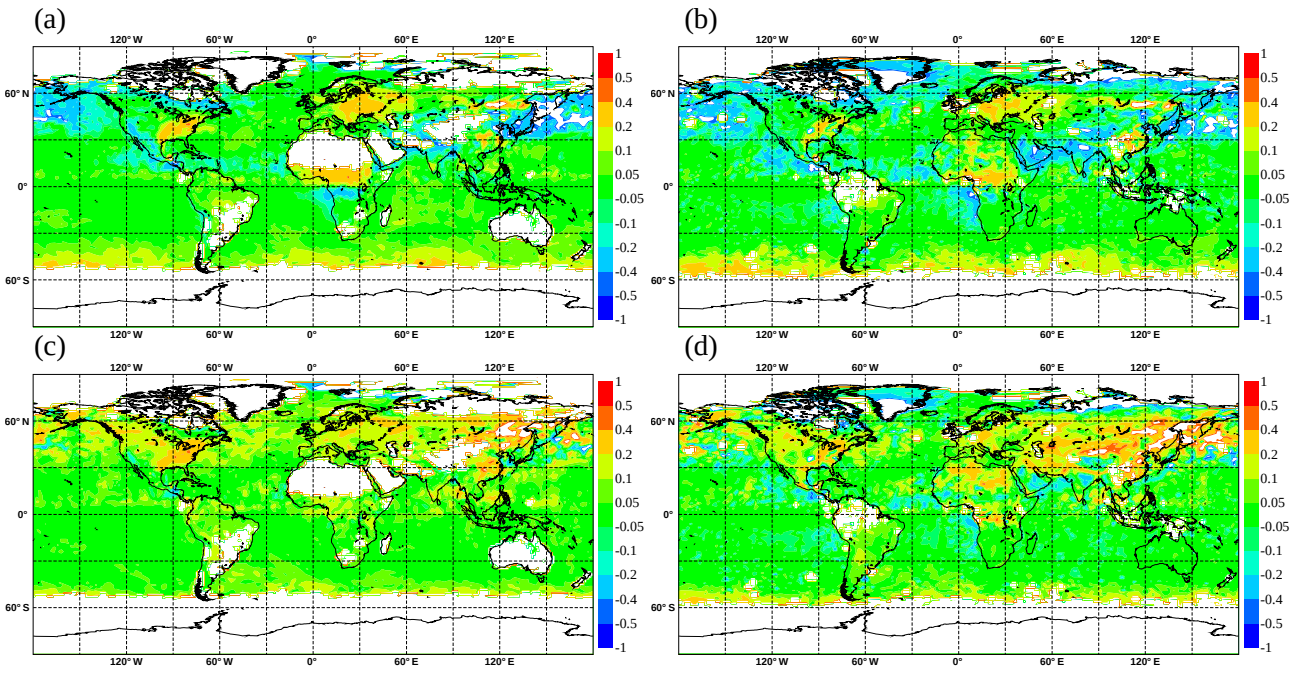


Figure 6: Global maps of AOD differences for May 2003: (a) free–running forecast minus MODIS Aqua observations, (b) free–running forecast minus MISR observations, (c) analysis minus MODIS Aqua observations, and (d) analysis minus MISR observations.

b. *Comparisons with AERONET observations*

The AERONET program (Holben et al. 1998) is a federation of ground-based remote sensing providing globally distributed observations of spectral aerosol optical depth, inversion products, and precipitable water in diverse aerosol regimes. The aerosol optical depth data used in this comparison are the Level 2.0 (cloud-screened and quality-assured) product.

Figure 7 shows some comparisons with AERONET independent data for the month of May 2003. In order to calculate a bias and a root mean square error (RMS) that are roughly indicative of global performance, data from a selected group of approximately evenly spaced, high-availability AERONET stations was used. The site selection was made using an algorithm which looped through all available sites, checking each for proximity to others. If two sites were found within 700 km of each other, then the site with greater availability (measured as the number of 6 hour periods with at least one observation at 500 nm during January 2003) was kept and the other discarded. This resulted in a selection of 41 stations.

The AODs from the model are averages over 6 hours, whereas the AERONET observations are instantaneous. To make them comparable, the AERONET observations are averaged over the same period. Because the observations are unevenly spaced in time, a weighted mean is computed in such a way that it is equal to the mean of the series of straight lines that join neighbouring observations over the period. Forecast AODs from the free–running experiment and the analysis are bilinearly interpolated to the observation location in space.

The analysis is shown in red and the free–running forecast in blue. Both plots show that the analysis is on average closer to the AERONET observations displaying a lower bias and RMS error than the forecast.

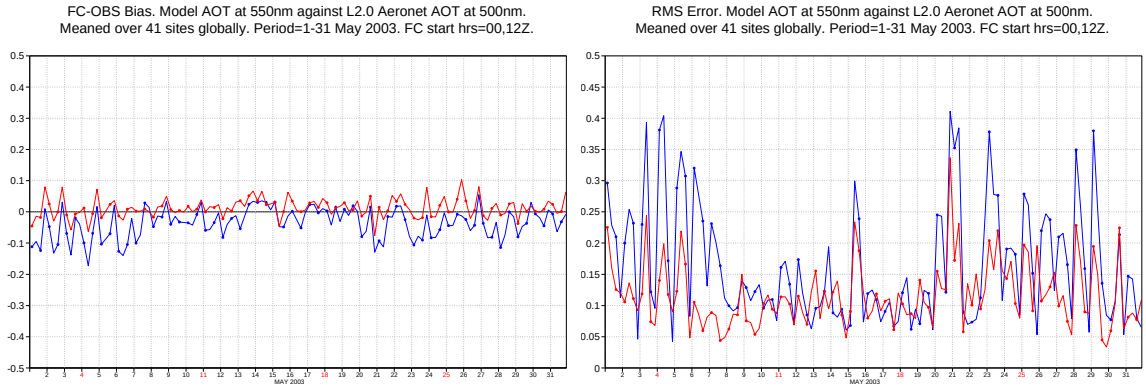
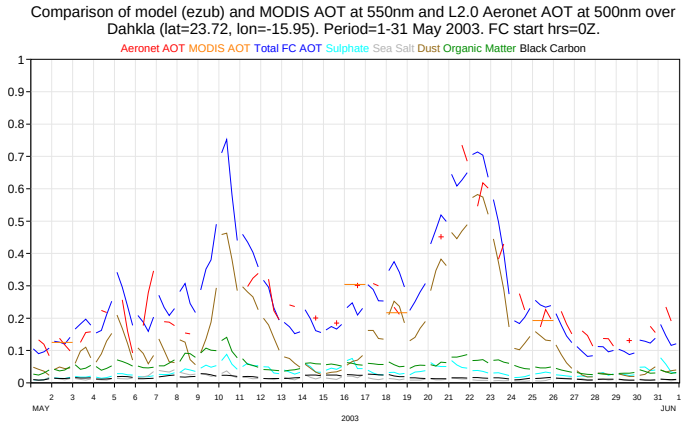


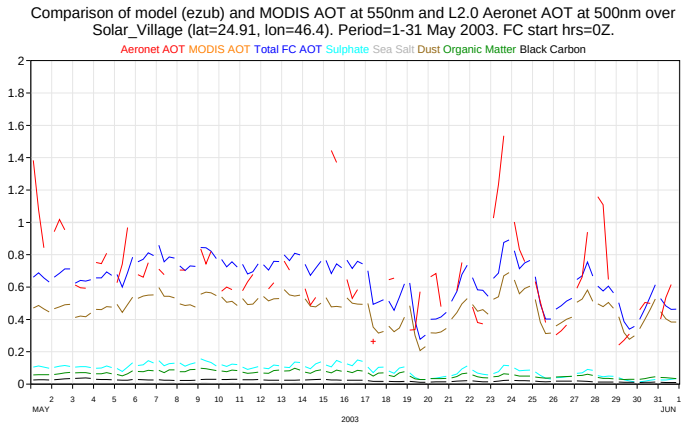
Figure 7: Bias (left) and RMS (right) of the AOD at 550 nm from the free–running forecast (blue) and analysis (red) with respect to AERONET ground–based observations at 500 nm for May 2003.

As an additional example, Figure 8 compares the analyzed optical depth at 550 nm with the observed aerosol optical depth over the AERONET sites of Dakhla (Morocco), Solar Village (Saudi Arabia) and Fresno (California, USA). The contributions to the total optical depth from the single constituents are also shown. Note how over the sites of Dakhla and Solar Village, the analysis is able to reproduce the observed variability and intensity of the dust episodes, despite the lack of MODIS data which are not available over highly–reflective surfaces.

(a)



(b)



(c)

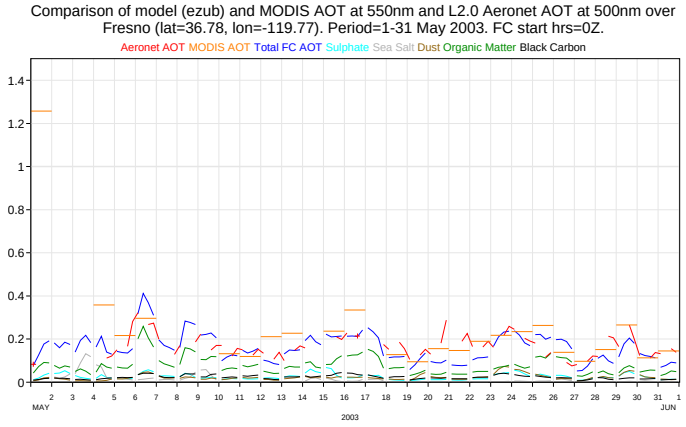


Figure 8: Comparisons of analysis aerosol optical depths with AERONET observations: (a) Dakhla (Morocco); (b) Solar Village (Saudi Arabia); and (c) Fresno (California, USA).

c. *A Saharan dust outbreak*

To further assess the performance of the analysis we looked at a case study relative to a major Saharan dust storm recorded in early March 2004. Cold air was advected from Europe to Western Africa, fanning out across the Sahara, highly diverging over subtropical regions and thus creating the dust storm. In the following days, the dust was blown out across the Atlantic Ocean and reached the coast of South America. The storm was detected by several satellite sensors and ground–based sites. Very large values of AOD were recorded. Figure 9 shows comparisons between AODs from the free-running model and the analysis compared to MODIS observations for 5-6 March 2004. The shape of the dust outflow is well represented in both free–running model and analysis, but the magnitude of the AODs is much larger in the latter in better agreement with the observations. This is also confirmed by looking at the AERONET data at key stations (see figure 10). The figure shows a comparison between AOD at 670 nm from the analysis and the free–running forecast and AERONET AODs at 675 nm for Agoufou (Mali), Dakar (Senegal) and Cape Verde. The peaks shown in the AERONET data are well captured by the analysis, with the exception of the 8th-9th of March AOD maximum over Agoufou. Level 2.0 AERONET data were used when possible. For Agoufou station only level 1.5 data were available for the dates of interest. The plot shows again a good degree of skill of the analysis in representing the observed values of aerosol optical depth. Details on this and other case studies will be presented in Mangold et al. (2008).

6. Conclusions and future outlook

This study presented the general architecture and the first results of the GEMS aerosol assimilation system developed at ECMWF. The aerosol species active in the model are sea salt, desert dust, organic matter, black carbon and sulphate. Appropriate parameterizations and inventories are used to describe emission of these species. Aerosol physical processes such as sedimentation and wet/dry deposition are also included. The assimilation uses the operational 4D–Var apparatus which has been extended to include atmospheric tracers among the control variables. At present, the total mixing ratio is used as control variable for the aerosol assimilation. Increments in this variable are redistributed into the different species according to their fractional contributions. The background error statistics have been computed for the total aerosol mixing ratio using six months of aerosol forecast differences at 48 and 24 hours (NMC method). The background error covariance matrix derived from this set of statistics has proven adequate to describe the error characteristics of the background aerosol fields, provided it is updated each time major model changes are implemented. The assimilation system uses retrievals of optical depth from the MODIS sensor on the Aqua and Terra satellites. All available observations over land (except bright surfaces) and ocean are used at their time and location over the 12-hour 4D–Var window. Results from the reanalysis for 2003 show a great degree of skill of the analysis to draw to the assimilated observations, although the analysis is more efficient in increasing rather than reducing the values of aerosol optical depth. Comparisons with independent measurements of AOD from the ground–based AERONET network show that the analysis has a lower bias and a lower RMS for most sites than a free–running forecast without assimilation. Of particular note is the ability of the analysis to improve the AOD forecast over sites where the MODIS observations are not available. This occurs thanks to the influence of observations in the neighboring areas and to the spreading out of information in the horizontal and vertical directions due to the use of the dynamical model in the 4D–Var minimization.

To make the analysis more effective, it will be necessary to assimilate other observations, for example AODs from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board of the Meteosat Second Generation satellites (Meteosat- 8 onward). Use of other sensors will also be investigated.

MODIS retrievals also provide general information on the breakdown between fine and coarse particle optical depth. One possibility is to assimilate this information directly into the 4D–Var system. Another possibility is

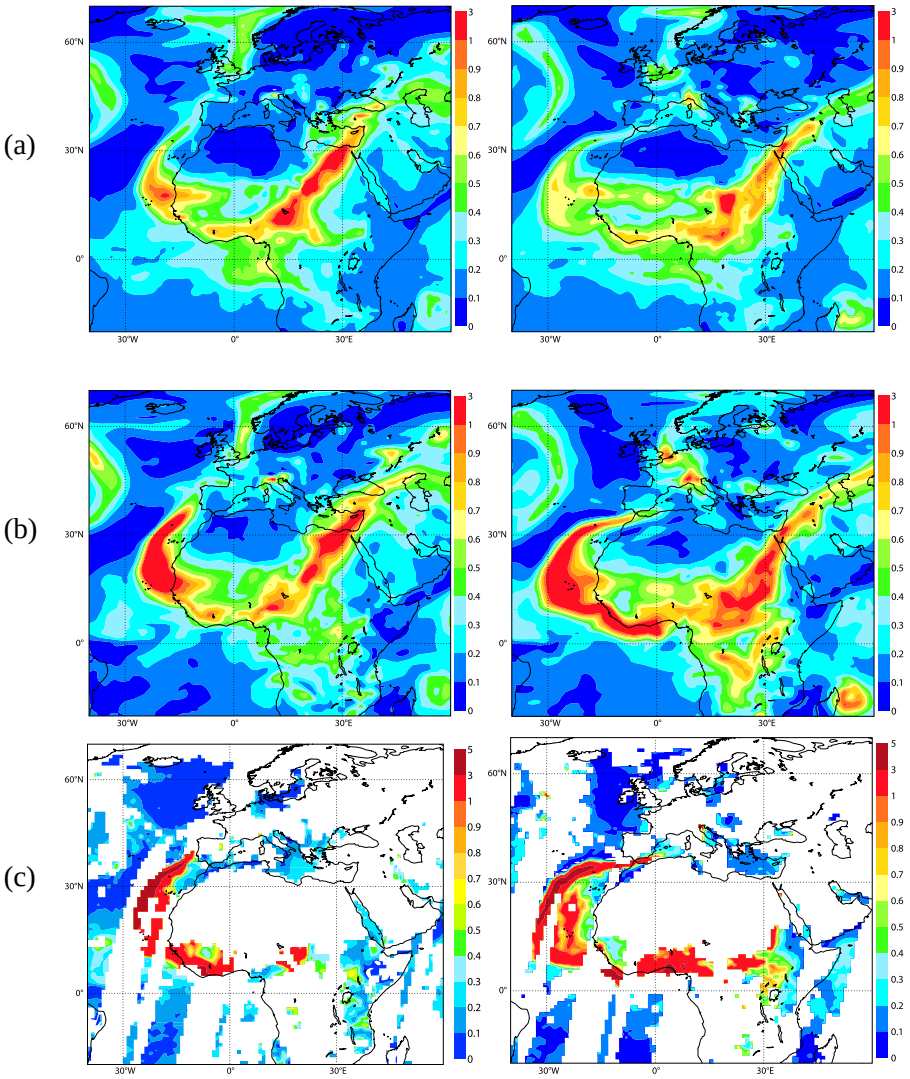
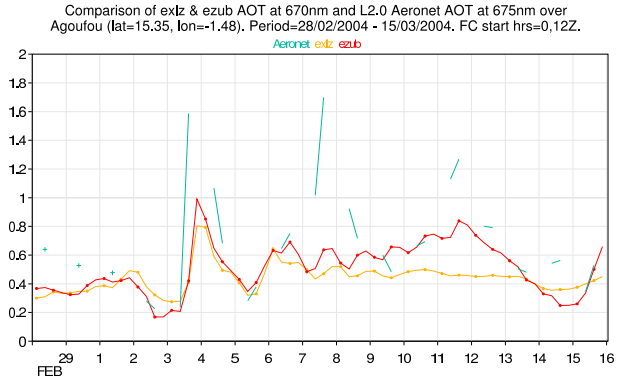
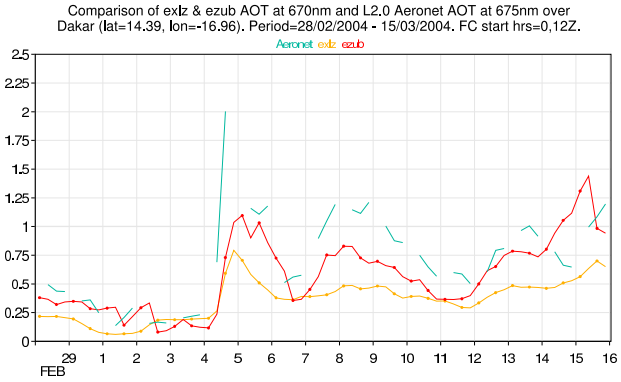


Figure 9: March 2004 Saharan dust outbreak: comparisons of free-running model and analysis 550 nm AODs with MODIS (assimilated) observations: (a) free running model ; (b) analysis ; and (c) MODIS observations. Panels on the left side show March 5 2004 at 1200UTC, while panels on the right side are for March 6 2004 at 1200UTC.

(a)



(b)



(c)

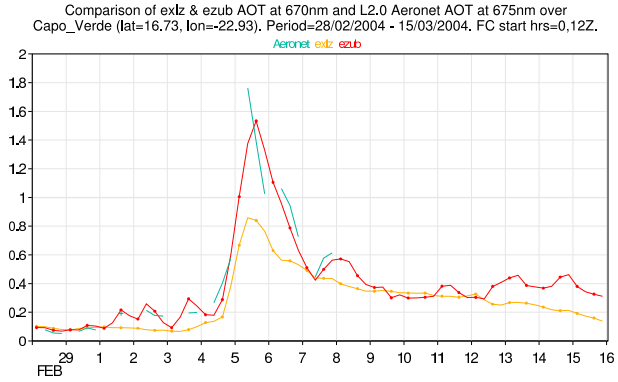


Figure 10: Comparisons of analysis aerosol optical depths with AERONET observations for the Saharan dust outbreak of March 2004: (a) Agoufou (Mali) ; (b) Dakar (Senegal); and (c) Cape Verde. AERONET data are shown here in light blue, the analysis in red and the free–running forecast in dark yellow.

to make use of the Angstrom parameter which also gives information on the size of the aerosol particulate from observations of optical depth at different wavelengths. This will require a re-thinking of the control variable and the possible introduction of more aerosol–related variables in the control vector.

A third possibility, which will be given priority, is direct assimilation of multi–wavelength reflectances. This development is already under way and the radiative transfer code for visible wavelengths has been already prepared to be plugged into the IFS. The tangent linear and adjoint operators for the radiative transfer code (6S, Vermote et al. (1997a,b)) , necessary for the incremental variational assimilation, are under development.

The GEMS aerosol reanalysis for 2003–2004 will be completed in August 2008. An in–depth review of the results and comparisons with yet more independent datasets is needed for a final assessment of the quality of the analysis. This will involve several of the partners in the GEMS project. First results are however encouraging and show the capability of the analysis to draw from the observations and provide optimal initial conditions for improved forecasts of atmospheric aerosol fields. A follow-on project, the Monitoring Atmospheric Composition & Climate (MACC) project, also funded by the European Commission, will explore the feasibility of pre-operational implementation of the GEMS assimilation system for reanalysis and near real–time forecasts of aerosols. In MACC, there are also plans to make the aerosol fully interactive with the radiation scheme thus allowing us to explore fully the impact of the improved aerosol fields on the whole atmospheric system.

7. Acknowledgements

We gratefully acknowledge the developers of the GES-DISC (Goddard Earth Sciences Data and Information Services Center) Interactive Online Visualization ANd aNalysis Infrastructure for providing an invaluable service of visualization and data processing for the MODIS and MISR data (see http://g0dup05u.ecs.nasa.gov/Giovanni/). AERONET data were obtained from the AERONET web site (http://aeronet.gsfc.nasa.gov/).

References

Bellouin, N., A. Jones, J. Haywood, and S. A. Christopher, 2008: Updated estimate of aerosol direct radiative forcing from satellite observations and comparison against the Hadley Centre climate model, *J. Geophys. Res.*, **113**, D10205, doi:10.10292007JD009385.

Benedetti, A. and M. Fisher, 2007: Background error statistics for aerosols, *Q. J. R. Meteorol. Soc.*, **133**, 391–405.

Benedetti, A. and M. Janisková, 2008: Assimilation of MODIS cloud optical depths in the ECMWF model, *Mon. Weather Rev.*, **136**, 1727–1746.

Boucher, O., M. Pham, and C. Venkataraman, 2002: Simulation of the atmospheric sulphur cycle in the LMD GCM. Model description, model evaluation and global and European budgets, Technical Report 23, IPSL, 26 pp.

Collins, W. D., P. J. Rasch, B. E. Eaton, B. V. Khattatov, and J.-F. Lamarque, 2001: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, *J. Geophys. Res.*, **106**, 7313–7336.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D–Var, using an incremental approach, *Q. J. R. Meteor.Soc.*, **120**, 1367–1387.

Crouzille, B., B. Gérard, J.-F. Léon, and D. Tanré, 2007: Methodology for quality assurance MODIS Aerosol Products, GEMS Technical Report, available at http://gems.ecmwf.int.

Diner, D. J., J. Beckert, T. Reilly, C. Bruegge, J. Conel, R. Kahn, J. Martonchik, T. Ackerman, R. Davies, S. Gerstl, H. Gordon, J.-P. Muller, R. Myneni, R. Sellers, B. Pinty, and M. Verstraete, 1998: Multi-angle Imaging Spectroradiometer (MISR) description and experiment overview, *IEEE Trans. Geosci. Rem. Sens.*, **36**, 1072–1087.

Diner, D. J., B. H. Braswell, R. Davies, N. Gobronc, J. Hud, Y. Jine, R. A. Kahn, Y. Knyazikhin, N. Loeb, J.-P. Muller, A. W. Nolin, B. Pinty, C. B. Schaaf, G. Seizi, and J. Stroeve, 2005: The value of multiangle measurements for retrieving structurally and radiatively consistent properties of clouds, aerosols, and surfaces, *Rem. Sens. Environ.*, **97**, 495–518.

Errera, Q. and D. Fonteyn, 2001: Four–dimensional variational chemical assimilation of CRISTA stratospheric measurements, *Mon. Weather Rev.*, **102**, 12253–12265.

Fisher, M., 1998: Minimization algorithms for variational data assimilation, Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling, 7-11 September 1998. Proceedings, ECMWF, pp. 364-385.

Fisher, M., 2003: Background error covariance modelling, Seminar on Recent developments in data assimilation for atmosphere and ocean, 8-12 September 2003. Proceedings, ECMWF, pp. 45-64.

Fisher, M., 2004: Generalized frames on the sphere, with application to background error covariance modelling, Seminar on Recent developments in numerical methods for atmospheric and ocean modelling, 6-10 September 2004. Proceedings, ECMWF, pp. 87-101.

Fisher, M., 2006: "Wavelet" J_b – A new way to model the statistics of background errors, *ECMWF Newsletter*, **106**, 23–28.

Fonteyn, D., Q. Errera, M. De Maziére, G. Franssens, and D. Fussen, 2000: 4D–Var assimilation of stratospheric aerosol satellite data, *Adv. Space Res.*, **26**, 2049–2052.

Generoso, S., F.-M. Bréon, F. Chevallier, Y. Balkanski, M. Schulz, and I. Bey, 2007: Assimilation of POLDER aerosol optical thickiness into the LMDz–INCA model: Implications for the Arctic aerosol burden, *J. Geophys. Res.*, **112**, D02311, doi:10.1029/2005JD006954.

Holben, B. N., T. Eck, I. Slutsker, D. Tanré, J. Buis, A. Setzer, E. Vermote, J. Reagan, Y. Kaufman, T. Nakajima, F. Lavenue, I. Jankowiak, and A. Smirnov, 1998: AERONET: A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, **66**, 1–16.

Hollingsworth, A., R. J. Engelen, C. Textor, A. Benedetti, O. Boucher, F. Chevallier, A. Dethof, H. Elbern, H. Eskes, J. Flemming, C. Granier, J. W. Kaiser, J.-J. Morcrette, P. Rayner, V.-H. Peuch, L. Rouil, M. G. Schultz, A. J. Simmons, and the GEMS consortium, 2008: The global earth–system monitoring using satellite and in–situ data (GEMS) project: Towards a monitoring and forecasting system for atmospheric composition, *Bull. Amer. Meteor. Soc.*, To appear.

Hólm, E., E. Andersson, A. Beljaars, P. Lopez, J.-F. Mahfouf, A. J. Simmons, and J.-N. Thépaut, 2002: Assimilation and Modelling of the Hydrological Cycle: ECMWF's Status and Plans, Technical Memorandum 383, ECMWF, 55 pp.

Kahn, R. A., M. J. Garay, D. L. Nelson, K. K. Yau, M. A. Bull, B. J. Gaitley, J. V. Martonchik, and R. C. Levy, 2007: Satellite-derived aerosol optical depth over dark water from MISR and MODIS: Comparisons with AERONET and implications for climatological studies, *J. Geophys. Res.*, **112**, D18205,doi:10.1029/2006JD008175.

Kahnert, M., 2008: Variational data analysis of aerosol species in a regional CTM: Background error covariance constraint and aerosol optical observation operators, *Tellus B*.

Lewtas, J., 2007: Air pollution combustion emissions: Characterization of causative agents and mechanisms associated with cancer, reproductive, and cardiovascular effects, *Mutat. Res. - Rev. Mut. Res.*, **636**, 95–133.

Lopez, P. and E. Moreau, 2005: A convection scheme for data assimilation: Description and initial tests, *Q. J. R. Meteorol. Soc.*, **131**, 409–436.

Mangold, A., H. De Backer, B. De Paepe, S. Dewitte, M. Schulz, Y. Balkanski, N. Huneeus, I. Chiapello, D. Ceburnis, C. O'Dowd, H. Flentje, S. Kinne, O. Boucher, J.-J. Morcrette, A. Benedetti, J. W. Kaiser, L. Jones, and the GEMS–AER team, 2008: Aerosol analysis and forecast in the ECMWF Integrated Forecast System. Part III: Evaluation, In preparation.

Morcrette, J.-J., A. Benedetti, O. Boucher, P. Bechtold, A. Beljaars, M. Rodwell, S. Serrar, M. Suttie, A. Tompkins, and A. Untch, 2006: GEMS–Aerosols at ECMWF, Seminar on Global Monitoring of the Earth-System, 5-9 September 2005. Proceedings, ECMWF, pp. 255-270.

Morcrette, J.-J., O. Boucher, D. Salmond, L. Jones, P. Bechtold, A. Beljaars, A. Benedetti, A. Bonet, A. Hollingsworth, J. W. Kaiser, M. Razinger, S. Serrar, A. J. Simmons, M. Suttie, A. Tompkins, A. Untch, and the GEMS–AER team, 2008: Aerosol analysis and forecast in the ECMWF Integrated Forecast System: Forward modelling, Technical report, ECMWF Technical Memorandum No. XXX, in preparation.

Niu, T., S. L. Gong, G. F. Zhu, H. L. Liu, X. Q. Hu, C. H. Zhou, and Y. Q. Wang, 2008: Data assimilation of dust aerosol observations for the CUACE/dust forecasting system, *Atmos. Chem. Phys.*, **8**, 3473–3482.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical–interpolation analysis system, *Mon. Weather Rev.*, **120**, 1747–1763.

Rasch, P. J., W. D. Collins, and B. E. Eaton, 2001: Understanding the Indian Ocean Experiment (INDOEX) aerosol distributions with an aerosol assimilation, *J. Geophys. Res.*, **106**, 7337–7355.

Reddy, M. S., O. Boucher, N. Bellouin, M. Schulz, Y. Balkanski, J.-L. Dufresne, and M. Pham, 2005: Estimates of global multicomponent aerosol optical depth and direct radiative perturbation in the Laboratoire de Météorologie Dynamique General Circulation Model, *J. Geophys. Res.*, **110**, D10S16, doi:10.1029/2004JD004757.

Remer, L. A., Y. J. Kaufman, D. Tanré, S. Mattoo, D. A. Chu, J. V. Martins, R.-R. Li, C. Ichoku, R. C. Levy, R. G. Kleidman, T. F. Eck, E. Vermote, and B. N. Holben, 2005: The MODIS aerosol algorithm, products and validation, *J. Atmos. Sci.*, **62**, 947–973.

Thompson, M. C., A. M. Molesworth, M. H. Djingarey, K. R. Yameogo, F. Belanger, and L. E. Cuevas, 2006: Potential of environmental models to predict meningitis epidemics in africa, *Trop. Med. Int. Health*, **11**, 781–788.

Tompkins, A. M., 2005: A revised cloud scheme to reduce the sensitivity to vertical resolution, Technical memorandum, ECMWF.

Tompkins, A. M. and M. Janisková, 2004: A cloud scheme for data assimilation: Description and initial tests, *Q. J. R. Meteorol. Soc.*, **130**, 2495–2517.

Trémolet, Y., 2005: Incremental 4D–Var convergence study, Technical Memorandum 469, ECMWF, 34 pp.

Tsigaridis, K., Y. Balkanski, M. Schulz, and A. Benedetti, 2008: Global error maps of aerosol optical properties: an error propagation analysis, *Atmos. Chem. Phys.*, Submitted.

Vermote, E. D., D. Tanré, J. L. Deuzé, M. Herman, and J.-J. Morcrette, 1997a: Second Simulation of the Satellite Signal in the Solar Spectrum: an overview, *IEEE Trans. Geosci. Remote Sens.*, **35**, 675–686.

Vermote, E. D., D. Tanré, J. L. Deuzé, M. Herman, and J.-J. Morcrette, 1997b: Second Simulation of the Satellite Signal in the Solar Spectrum: an overview, 6S, User guide version 2, 53 pp.

Weaver, C., A. da Silva, M. Chin, P. Ginoux, O. Dubovik, D. Flittner, A. Zia, L. Remer, B. Holben, and W. Gregg, 2007: Direct insertion of MODIS radiances in a global aerosol transport model, *J. Atmos. Sci.*, **64**, 808–826.

Zhang, J. and J. S. Reid, 2006: MODIS aerosol product analysis for data assimilation: Assessment of over–ocean level 2 aerosol optical thickness retrievals, *J. Geophys. Res.*, **111**, D22207, doi:10.1029/2005JD006898.

Zhang, J., J. S. Reid, D. L. Westphal, N. L. Baker, and E. J. Hyer, 2008: A system for operational aerosol optical depth assimilation over global oceans, *J. Geophys. Res.*, **113**, D10208, doi:10.1029/2007JD009065.