Implementation of aerosol assimilation in Gridpoint Statistical Interpolation v. 3.2 and WRF-Chem v. 4.3.1

M. Pagowski\(^1\), Z. Liu\(^2\), G. A. Grell\(^1\), M. Hu\(^1\), H.-C. Lin\(^2\), and C. S. Schwartz\(^2\)

\(^1\)NOAA Earth System Research Laboratory (ESRL), Boulder, Colorado, USA
\(^2\)National Center for Atmospheric Research, Boulder, Colorado, USA

Received: 17 March 2014 – Accepted: 9 April 2014 – Published: 24 April 2014

Correspondence to: M. Pagowski (mariusz.pagowski@noaa.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Gridpoint Statistical Interpolation (GSI) is an assimilation tool that is used at the National Centers for Environmental Prediction in operational weather forecasting. In this article we describe implementation of an extension to the GSI for assimilating surface measurements of PM$_{2.5}$, PM$_{10}$, and MODIS Aerosol Optical Depth at 550 nm with WRF-Chem. We also present illustrations of the results. In the past the aerosol assimilation system has been employed to issue daily PM$_{2.5}$ forecasts at NOAA/ESRL and, in our belief, is well tested and mature enough to make available for wider use. We provide a package that, in addition to augmented GSI, consists of software for calculating background error covariance statistics and for converting in-situ and satellite data to BUFR format, plus sample input files for an assimilation exercise. Thanks to flexibility in the GSI and coupled meteorology-chemistry of WRF-Chem, assimilating aerosol observations can be carried out simultaneously with meteorological data assimilation. Both GSI and WRF-Chem are well documented with user guides available on-line. This article is primarily intended as a technical note on the implementation of the aerosol assimilation. Its purpose is also to provide guidance for prospective users of the computer code. Limited space is devoted to scientific aspects of aerosol assimilation.

1 Introduction

Data assimilation plays an increasingly important role in forecasting concentrations of chemical species, replacing a somewhat outdated procedure where meteorological assimilation was only applied to meteorology, while chemical species were continuously recycled without referring to observations. Interest in chemical data assimilation dates back to Elbern and collaborators (Elbern et al., 1997, 2000, 2001, 2007; Elbern and Schmidt, 1999) and Carmichael and collaborators (Daescu and Carmichael, 2003; Carmichael et al., 2003; Sandu et al., 2005; Chai et al., 2007; and Constanti-
Assimilation methods described in these publications include 3-D- and 4-D-Var and ensemble Kalman filters.

Realisation that aerosols affect weather and climate and are a pivotal contributor to air pollution has led to recent developments in assimilating aerosols. Examples of such developments include Benedetti and Fisher (2007), Kahnert (2008), Morcrette et al. (2009), Benedetti et al. (2009), Schutgens et al. (2010a, b), Pagowski et al. (2010), Liu et al. (2011), Pagowski and Grell (2012), Schwartz et al. (2012), Saide et al. (2013), and Schwartz et al. (2014).

The above publications have shown that initial conditions play an important but not a dominant role in chemical forecasting. Especially for predicting air quality, i.e. chemical composition in the boundary layer, inaccurate source emissions and deficient physical and chemical parameterizations result in deteriorating forecasts soon after the assimilation. In this context, applying 3-D-Var assimilation methods that aim to exclusively ameliorate initial conditions constitutes only a first step towards improving chemical forecasts.

Below we describe aerosol observations that can be currently assimilated with our extension of the Gridpoint Statistical Interpolation (GSI, Wu et al., 2002; Purser et al., 2003a, b). We also briefly describe the assimilation system and show some example applications.

## 2 Observations and measurement errors

In our implementation, assimilated observations include surface measurements of PM$_{2.5}$ and PM$_{10}$, plus Aerosol Optical Depth (AOD, alternatively, Aerosol Optical Thickness, AOT) retrievals at 550 nm from Moderate Resolution Imaging Spectroradiometer (MODIS) satellites Aqua and Terra.

In North America, continuous measurements of surface aerosol concentrations at hourly resolution are made available thanks to monitoring stations participating in the US EPA AIRNow program. The observations are processed with minimal delay, making
them suitable for real-time assimilation. A free subscription to the real-time data feed is possible through the AIRNow gateway. Details on accessing the data can be found at http://www.airnowgateway.org/GatewayDataUseGuidelines.pdf. There are over 600 sites measuring PM$_{2.5}$ and over 150 sites measuring PM$_{10}$. From the gateway, a file with information on measurements can be obtained. The file contains nine-digit IDs of the measurement sites and their latitudes and longitudes recorded at an accuracy of four decimal places. For the majority of sites, information is also available on their elevations and rural, suburban, or urban character of their locations.

In our processing, we create a formatted input file that for each measurement contains date and time (with minute accuracy), latitude, longitude, concentration in $\mu$gm$^{-3}$, site character (urban, suburban, or rural), elevation, and ID. Since AIRNow measurements are one-hour averages, we assign the time of their validity at 30 min past the hour. Computer code is made available to convert this text-formatted file to BUFR (Binary Universal Form for the Representation of Meteorological Data) as required by the GSI.

AIRNow PM$_{2.5}$ and PM$_{10}$ concentrations are measured using Tapered Element Oscillating Microbalance instruments (TEOM, Thermo Fisher, Continuous particulate TEOM monitor, Series 1400ab, product detail, 2007, available at http://www.thermo.com/com/cda/product/detail/1,10122682,00.html). The uncertainty of both aerosol measurements $\epsilon_m$ is $1.5 \, \mu$gm$^{-3}$ plus an inaccuracy of 0.75% times the species concentration.

AOD data come from MODIS sensors on board the Terra and Aqua satellites. Retrievals over land and sea are derived from the dark target product (Remer et al., 2005) and deep blue product over bright land surface (Hsu et al., 2004, 2006). Currently, the dark target ocean and land AOD products are available from both Terra and Aqua, but deep blue retrievals are only available from Aqua. MODIS retrieved AOD is provided at seven wavelengths: 470, 550, 660, 870, 1240, 1630, and 2130 nm. In our implementation, only Level 2 (L2) AOD retrievals at 550 nm are used. The AOD observation error is specified after Remer et al. (2005) as $\epsilon_{AOD} = 0.03 + 0.05\tau$ over water.
and $\varepsilon_{\text{AOD}} = 0.05 + 0.15\tau$ over land, where $\tau$ is an AOD observation. Only AOD retrievals marked with the highest quality flag are retained for the assimilation.

L2 retrievals from Aqua are available at ftp://ladsweb.nascom.nasa.gov/allData/51/MYD04_L2 and L2 retrievals from Terra are available at ftp://ladsweb.nascom.nasa.gov/allData/51/MOD04_L2. These data come in HDF-EOS format at 5 min segments of the satellite’s orbit that correspond to 10 km $\times$ 10 km resolution at the surface. Computer code (W. Wolf, personal communication, 2013) is available in the package to convert HDF to BUFR for the GSI.

3 Aerosol assimilation within the Gridpoint Statistical Interpolation

GSI includes a 3-D-VAR assimilation tool from which an analysis is obtained by minimization of a cost function given by

$$J(x) \equiv (x - x_b)^T B^{-1} (x - x_b) + (y - H(x))^T R^{-1} (y - H(x)) \quad (1)$$

In Eq. (1), $x$ is a vector of analysis, $x_b$ is the forecast or background vector, $y$ is an observation vector, $H$ is an observation operator, and $B$ is the background error covariance matrix. $R$ is the observation error covariance matrix, and is usually assumed to be diagonal. Each element of the matrix combines measurement and representativeness errors. The matrix $B$ is separated into vertical and horizontal components and is represented as a product of error variances and spatial correlation matrices which are modelled with recursive filters (Purser et al., 2003a, b). Application of the recursive filters requires specification of the background error correlation length scales. Incremental approach (Courtier et al., 1994) is employed to minimize the cost function.

Our extension to the GSI includes separate options for Goddard Chemistry Aerosol Radiation and Transport (GOCART, Chin et al., 2000, 2002; Ginoux et al., 2001) and all other aerosol modules in WRF-Chem (Grell et al., 2005). Since the Community Radiative Transfer Model (CRTM, Han et al., 2006; Liu and Weng, 2006), which is coupled
to the GSI, is currently only available for GOCART, AOD can only be assimilated with the GOCART model background.

3.1 Forward models and observation processing in GSI

Forward models for GOCART differ from other aerosol parameterizations and are described first.

PM\textsubscript{2.5}, PM\textsubscript{10}, and AOD are all integrated measurements that require a summation of individual aerosol species. In WRF-Chem implementation, GOCART aerosol species include unspecified P\textsubscript{25}, sulphate S, hydrophobic and hydrophilic black carbon (BC\textsubscript{1}, BC\textsubscript{2}), hydrophobic and hydrophilic organic carbon (OC\textsubscript{1}, and OC\textsubscript{2}), five dust bins (D\textsubscript{1}, D\textsubscript{2}, D\textsubscript{3}, D\textsubscript{4}, and D\textsubscript{5}), and four sea salt bins (SS\textsubscript{1}, SS\textsubscript{2}, SS\textsubscript{3}, and SS\textsubscript{4}).

PM\textsubscript{2.5} concentration is calculated as

\[
PM_{2.5} = \rho_d [P_{2.5} + 1.375S + BC_1 + BC_2 + 1.8(OC_1 + OC_2) + D_1 + 0.286D_2 \\
+ SS_1 + 0.942SS_2],
\]

where \(\rho_d\), dry air density, is multiplied by mixing ratios of aerosol species. Factors for sulphate and organic carbon account for increasing the mass of the compounds due to the presence of ammonium ion and oxygen, respectively. Factors for dust and seasalt account for a size cut-off at the 2.5 \(\mu\)m diameter. An expression for PM\textsubscript{10} concentration is

\[
PM_{10} = \rho_d [P_{2.5} + 1.375S + BC_1 + BC_2 + 1.8(OC_1 + OC_2) + D_1 + D_2 + D_3 + 0.87D_4 \\
+ SS_1 + SS_2 + SS_3].
\]

Only a brief description of the observation operator for AOD is given here and we refer the reader to Liu et al. (2011) and Schwartz et al. (2012) for full details. We assume that the size distribution of each aerosol mode is logarithmic and that the particles are spherical and externally mixed. CRTM contains profiles of GOCART aerosol species
that include their effective radii, standard deviations, and refractive indices. The extinction coefficient of each aerosol species is computed for a given wavelength based on Mie scattering theory and accounting for hygroscopic size growth of hydrophilic species. Finally, AOD is calculated from the equation

$$\tau(\lambda) = \sum_{i=1}^{n} \sum_{k=1}^{k_{\text{top}}} E_{\text{ext}}(\lambda, n_{r_i}, r_{\text{eff}_i}) \times c_{ik} \times \rho_{d_k} \times d_k, \quad (4)$$

where $E_{\text{ext}}$ is the extinction coefficient (a function of wavelength $\lambda$, refractive index $n_{r_i}$, and effective radius $r_{\text{eff}_i}$), $c$ is aerosol mixing ratio, and $d$ is layer depth. Indices $i$ and $k$ denote aerosol species and model layers respectively; $n = 15$ denotes the number of GOCART aerosol species.

For each of the summations (1), (3), and (4), mixing ratios of aerosol species are horizontally linearly interpolated to the observation location. No extrapolations are performed in the vertical for surface observations, as their locations are assumed to coincide with the first model level.

Representativeness error for a surface observation is assigned based on the character of the site after Elbern et al. (2007), using a formula given by $\varepsilon_{\text{repr}} = \alpha \varepsilon_{m} (\Delta x / L_{\text{repr}})^{1/2}$, where $\varepsilon_m$ is measurement error, $\alpha$ is a tunable parameter, $\Delta x$ is model grid size, and $L_{\text{repr}}$ represents the observation’s radius of influence. The parameter $\alpha$ can be specified in the namelist. Its default value is set to 0.5. Radii of influence for observations are prescribed equal to 10 km, 4 km, and 2 km for rural, suburban, and urban sites, respectively. The total observation error is calculated as $\varepsilon_{\text{obs}} = (\varepsilon_{m}^2 + \varepsilon_{\text{repr}}^2)^{1/2}$.

Only surface measurements that fall below specified thresholds are accepted (default values are set to 100 µg m$^{-3}$ for PM$_{2.5}$ and to 150 µg m$^{-3}$ for PM$_{10}$). Also, an observation is rejected if its distance from the background is greater than these maximum allowable values. Depending on the user’s preference, an observation can also be rejected if a difference between its actual elevation and model terrain height interpolated
to its geographic location exceeds a threshold specified in the namelist. Characteristics of the error for different instruments and the default values can be easily modified (in the GSI distribution files convinfo, chemmod.f90 and read_anowbufr.f90).

To reduce volume and diminish the correlation of satellite observation errors, thinning of AOD observations is recommended to a resolution that is comparable to the model grid size. Thinning options can be specified in the namelist.

For aerosol options other than GOCART, PM$_{2.5}$ or PM$_{10}$ are read as PM2_5_DRY or PM10 from WRF-Chem output so that summations (1) and (3) are not required. The rationale for such an approach is discussed in the section on background error. Calculation of surface PM observation errors and data selection for the assimilation follows the implementation for GOCART.

### 3.2 Specification of background error

In GSI, error correlation length scales and variances can vary zonally and vertically. They can be calculated as forecast statistics using the NMC method (Parrish and Derber, 1992) or the ensemble method (Fisher, 2003). Computer code to produce a file containing these statistics for meteorological state variables and desired aerosols formatted for the GSI is available for download with WRF Data Assimilation system at http://www.mmm.ucar.edu/wrf/users/wrfda/downloads.html.

For GOCART parameterization, state variables include 15 aerosol species. As an illustration, vertical profiles of standard deviations and horizontal correlation length scales for OC$_1$, OC$_2$, and sulphate are shown in Fig. 1. These statistics were derived for a month-long period in the 2012 summer, over a domain spanning eastern North America, with 24 km grid resolution, using NMC method applied to 24 and 48 h forecasts. In the GOCART case, increments to each aerosol species are obtained using background error statistics for individual aerosol species. We will not reflect on the realism of the statistics derived using the NMC method in this manuscript, but only point out that accounting for uncertainty in emission sources and aerosol parameterization...
deficiencies should be considered when estimating model errors. Pagowski and Grell (2012) discuss this topic in detail.

An alternative approach is also available where increments to individual species are calculated based on their a priori contribution to the total aerosol mass. This is expressed as the sum of 15 aerosols species accounting for multiplication factors of sulphate and organic carbon (hereafter, “ratio approach”). With this approach, statistics for the total aerosol are used to minimize the 3-D-Var cost function and need to be provided in the background error input file. The choice of any of the two approaches is determined in the namelist. Also, error correlation length scales and standard deviations can be tuned for optimal performance and modified by factors specified in the namelist.

For parameterizations other than GOCART, specifying background error statistics for a large number of aerosol species is in our opinion overly burdensome, especially because such statistics may not be reliable given the large uncertainties in emissions and in the state of science in aerosol modelling. Therefore, for these parameterizations, we require that background error statistics are provided for a WRF-Chem output variable PM2_5_DRY/PM10 when PM$_{2.5}$/PM$_{10}$ observations are assimilated. This variable is also a state variable for which an increment will be calculated.

3.3 Running GSI and aerosol assimilation cycle

A comprehensive user’s guide for GSI is available at http://www.dtcenter.org/com-GSI/users/docs/users_guide/GSIUserGuide_v3.2.pdf. Also, an on-line tutorial is available and group tutorials are given at least once a year (http://www.dtcenter.org/com-GSI/users/tutorial/index.php). Thus, only a cursory description of the assimilation is given here. Our package provides a default configuration and shell scripts for assimilating PM$_{2.5}$, PM$_{10}$, and MODIS AOD with WRF-Chem GOCART parameterization.

Specifically, for aerosol assimilation, in addition to an input file with aerosol background statistics, a user needs to provide WRF-Chem output in netcdf format, observations files in BUFR format (normally a single file for PM$_{2.5}$ and PM$_{10}$, and/or file with
MODIS AOD), a namelist specifying options for the assimilation, plus a configuration file anavinfo. The latter file contains the names of aerosol species as state variables for which minimization of the 3-D-Var cost function is performed. Normally, entries in anavinfo would include either GOCART species or PM$_{2.5}$/PM$_{10}$. We note that a simultaneous assimilation of meteorological variables is also possible, although regressions between meteorology and chemistry have not been derived nor coded.

On the output, GSI overwrites the input WRF-Chem file. For quality control and to visualize increments, we suggest using ncdiff, a component of netcdf manipulation software NCO available at http://nco.sourceforge.net. For GOCART, the output WRF-Chem file contains an analysis of aerosol species. No further processing is required to issue the next forecast. For other aerosol options, increments to individual aerosol species need to be calculated using the ratio approach and added to the background. They will constitute initial conditions for the following forecast. We again recommend using NCO software for this procedure. Sample increments to OC$_1$, OC$_2$, and sulphate on the first model level (i.e. assumed to be at the surface) are shown in Fig. 2. Their magnitudes and spatial patterns are related to the specification of background error statistics for individual aerosol species. Surface and satellite observations were assimilated to produce this figure.

We routinely employ a six-hour assimilation cycle that includes both assimilation of standard meteorological observations and aerosol observations.

The impact of aerosol assimilation has been well documented in the publications cited in Sect. 1. For illustration, Fig. 3 shows bias and spatial correlation with respect to AIRNow measurements calculated for forecasts issued over a month-long period during summer 2012 with and without assimilation of surface observations of PM$_{2.5}$. GOCART parameterization was used with the ratio approach. The improvement in the forecasts’ quality is significant, though not long-lasting. Reasons for a relatively quick deterioration of the aerosol forecasts were briefly noted in Sect. 1 and are elaborated in detail in Pagowski and Grell (2012) and Jiang et al. (2013).
4 Conclusions

We described our implementation of the assimilation of PM$_{2.5}$ and PM$_{10}$, surface observations and satellite MODIS AOD level 2 retrieval using the GSI and WRF-Chem. Along with aerosol assimilation, computer codes for formatting the observations are included in the package. Also, an example configuration and sample input files for an assimilation exercise are supplied.

We recommend that prospective users become familiar with a general application of the GSI as described in the User’s Guide and in the on-line tutorial.

We hope that the availability of this implementation will lead to further development of the aerosol and chemical data assimilation system that may include wider range of observations. GSI is a community-based system and user contributions are encouraged.

Acknowledgements. This work has been supported by the US Weather Research Program within the NOAA/OAR Office of Weather and Air Quality, NOAA Health of the Atmosphere Program, and US Air Force Weather Agency.

References


2496

Fig. 1. Vertical profiles of standard deviations (top) and horizontal correlation length scales (bottom) for OC$_1$, OC$_2$, and sulphate derived for a North American domain (see text for details).
Fig. 2. Sample analysis increments of $\text{OC}_1$, $\text{OC}_2$, and sulphate (clockwise from the top left) on the first model level.
Fig. 2. Sample analysis increments of OC1, OC2, and sulphate (clockwise from the top left) on the first model level.

Fig. 3. Bias (left) and spatial correlation (right) calculated for forecasts issued over a month-long period in summer 2012 for the North America domain with and without assimilation of surface observations of PM$_{2.5}$. 