



# A case study of aerosol data assimilation with the Community Multi-scale Air Quality Model over the contiguous United States using 3D-Var and optimal interpolation methods

Youhua Tang<sup>1,2</sup>, Mariusz Pagowski<sup>4,5</sup>, Tianfeng Chai<sup>1,2</sup>, Li Pan<sup>1,2</sup>, Pius Lee<sup>1</sup>, Barry Baker<sup>1,2</sup>, Rajesh Kumar<sup>6</sup>, Luca Delle Monache<sup>6</sup>, Daniel Tong<sup>1,2,3</sup>, and Hyun-Cheol Kim<sup>1,2</sup>

<sup>1</sup>NOAA Air Resources Laboratory, College Park, MD, USA

<sup>2</sup>Cooperative Institute for Climate and Satellites, University of Maryland, College Park, MD, USA

<sup>3</sup>Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA

<sup>4</sup>NOAA Earth System Research Laboratory, Boulder, CO, USA

<sup>5</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

<sup>6</sup>National Center for Atmospheric Research, Boulder, CO, USA

**Correspondence:** Youhua Tang (youhua.tang@noaa.gov)

Received: 21 June 2017 – Discussion started: 11 July 2017

Revised: 3 October 2017 – Accepted: 17 November 2017 – Published: 22 December 2017

**Abstract.** This study applies the Gridpoint Statistical Interpolation (GSI) 3D-Var assimilation tool originally developed by the National Centers for Environmental Prediction (NCEP), to improve surface PM<sub>2.5</sub> predictions over the contiguous United States (CONUS) by assimilating aerosol optical depth (AOD) and surface PM<sub>2.5</sub> in version 5.1 of the Community Multi-scale Air Quality (CMAQ) modeling system. An optimal interpolation (OI) method implemented earlier (Tang et al., 2015) for the CMAQ modeling system is also tested for the same period (July 2011) over the same CONUS. Both GSI and OI methods assimilate surface PM<sub>2.5</sub> observations at 00:00, 06:00, 12:00 and 18:00 UTC, and MODIS AOD at 18:00 UTC. The assimilations of observations using both GSI and OI generally help reduce the prediction biases and improve correlation between model predictions and observations. In the GSI experiments, assimilation of surface PM<sub>2.5</sub> (particle matter with diameter < 2.5 µm) leads to stronger increments in surface PM<sub>2.5</sub> compared to its MODIS AOD assimilation at the 550 nm wavelength. In contrast, we find a stronger OI impact of the MODIS AOD on surface aerosols at 18:00 UTC compared to the surface PM<sub>2.5</sub> OI method. GSI produces smoother result and yields overall better correlation coefficient and root mean squared error (RMSE). It should be noted that the 3D-Var and OI methods used here have several big differences besides the data assimilation schemes. For instance, the OI uses relatively big

model uncertainties, which helps yield smaller mean biases, but sometimes causes the RMSE to increase. We also examine and discuss the sensitivity of the assimilation experiments' results to the AOD forward operators.

## 1 Introduction

The existing US National Air Quality Forecasting Capability (NAQFC) run by the National Oceanic and Atmospheric Administration (NOAA)/National Centers for Environmental Prediction (NCEP) provides daily 48 h ozone and PM<sub>2.5</sub> (particle matter with diameter < 2.5 µm) forecasts using the Community Multi-scale Air Quality (CMAQ) modeling system with 12 km horizontal grid resolution. Many contributions toward improving the NAQFC can be found in the literature, including updated emission, meteorology, chemical mechanism and lateral boundary conditions (Pan et al., 2014; Tang et al., 2008; Lee et al., 2017). However, biases still contaminate current predictions.

Chemical data assimilation techniques have been developed to improve initial conditions of chemical transport models (CTMs) and yield better prediction (Elbern et al., 1997, 2000, 2007; Elbern and Schmidt, 1999, 2001; Bocquet et al., 2015) by blending the information from a model estimate

(referred to as prior or background) and from observations in certain methods. One method is direct blending, e.g., using observed chemical mass concentrations to correct the modeled mass concentrations, which is relatively straightforward as they are directly comparable. As most monitoring data are located near surface, this method was usually applied to the near-surface field. Another method is indirect guessing, e.g., comparing satellite-retrieved aerosol optical depth (AOD) with modeled AOD to estimate the biases of modeled column mass concentrations and make the corresponding adjustment. Tang et al. (2015) used both surface  $\text{PM}_{2.5}$  and MODIS AOD to adjust the initial condition of a CTM with the optimal interpolation (OI) method, and the model biases were successfully reduced. The OI correction was applied to  $7 \times 7$  or  $11 \times 11$  grid cells horizontally, and vertically from the ground to the boundary layer top for surface  $\text{PM}_{2.5}$  assimilation. Adhikary et al. (2008) and Chai et al. (2017) also used similar local OI correction to reduce model biases. A more complex method is using Gridpoint Statistical Interpolation (GSI) (Wu et al., 2002; Purser et al., 2003a, b), a 3D-Var method, developed by NOAA/NCEP. It has been applied to assimilate the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosols in the Weather Research and Forecasting model coupled with Chemistry (WRF–Chem) with its 3D-Var method (Pagowski et al., 2014; Liu et al., 2011). In the next sections, we describe the method that extends the GSI to assimilate CMAQ aerosols, comparing it to OI correction. A comparison for their 1-month performances will be discussed.

## 2 Methodology and settings

The baseline model setting used in this study is similar to Tang et al. (2015), except that the CMAQ model was changed from version 5.0.2 to version 5.1 for more refined chemical mechanisms and physical schemes. The meteorology is provided by the WRF-ARW (version 3.4.1) driven by the NCEP FNL (Final) Global Analysis  $1^\circ \times 1^\circ$  analysis field and was reinitialized every 24 h. Both the meteorological and air quality models have 12 km horizontal resolution over the contiguous United States (CONUS), with 42 vertical sigma layers and the domain tops at 50 hPa. Detailed settings of the CMAQ model can be found in Tang et al. (2015). Chai et al. (2017) described the OI algorithm for aerosol assimilation in detail, and that study used a relatively old CMAQ model (version 4.7.1) with 22 vertical layers. In order to compare the two methods of adjusting the initial condition, we assimilate same surface  $\text{PM}_{2.5}$  data, the USEPA Air Quality System (AQS) data, four times every day (00:00, 06:00, 12:00 and 18:00 Z). The 18:00 Z assimilation uses additional MODIS AOD data from Terra and Aqua satellites, whose overpass times over the CONUS domain range from 14:00 to 22:00 UTC, depending on the date and location. In this study, we only used the MODIS AOD at 550 nm wavelength

(Level-2 MODIS AOD from the Collection 051 with the best quality, a quality flag of 3) (Levy et al., 2009).

### 2.1 Settings for data assimilations

In this study, we compare two assimilation methods: optimal interpolation and GSI's 3D-Var. Optimal interpolation is carried out in a similar way to that described in the OI4 case of Tang et al. (2015) for assimilating surface  $\text{PM}_{2.5}$  and aerosol optical depth retrieved from Aqua–Terra MODIS sensors at 550 nm.

$$X^a = X^b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{O})^{-1}(\mathbf{Y} - \mathbf{H}X^b), \quad (1)$$

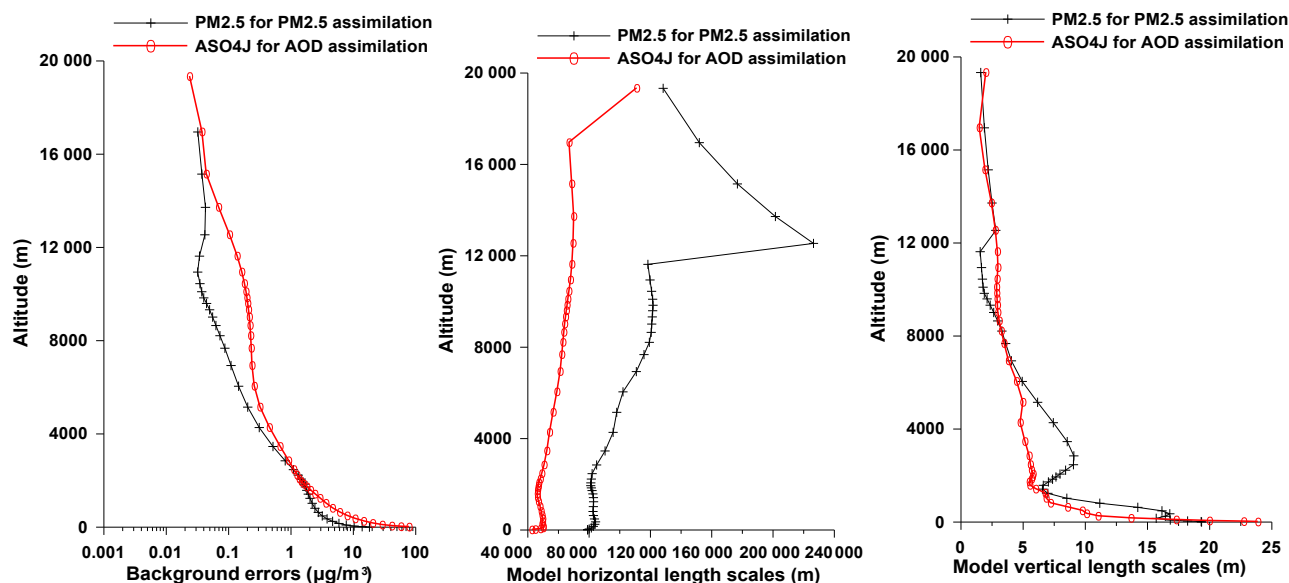
where  $X^a$  and  $X^b$  are the analyzed and background (prior modeled) concentrations or aerosol optical depth (AOD) data,  $\mathbf{B}$  and  $\mathbf{O}$  are the background and observation error covariance matrices,  $\mathbf{H}$  is the observational operator and  $\mathbf{H}^T$  is its matrix transpose and  $\mathbf{Y}$  is the observation vector. The relative uncertainty setting is also the same as in Tang et al. (2015), when the background relative uncertainties have horizontal and diurnal variations. Before the OI assimilation, the AQS surface  $\text{PM}_{2.5}$  monitoring data and MODIS AOD data are put into their corresponding model grids. If there was more than one active surface  $\text{PM}_{2.5}$  station in the same model grid, the observation was taken from their average. The OI adjustment is made on each  $11 \times 11$  grid cell horizontally throughout the whole domain, and the analyzed field  $X^a$  is obtained. Then the ratio of  $X^a/X^b$  at each grid point is used to scale all aerosol components throughout the vertical column. The surface  $\text{PM}_{2.5}$  OI is applied from the surface to the height of the planetary boundary layer (PBL), and the MODIS AOD is used to adjust above-PBL aerosol after deducting the adjusted below-PBL AOD.

The GSI's 3D-Var uses a similar approach (Pagowski et al., 2014; Liu et al., 2011) for its cost function

$$J = \frac{1}{2}(x_a - x_b)^T \mathbf{B}^{-1}(x_a - x_b) + \frac{1}{2}(\mathbf{H}x_a - \mathbf{O}_o)^T \mathbf{O}^{-1}(\mathbf{H}x_a - \mathbf{O}_o) + J_c, \quad (2)$$

where  $x_a$  and  $x_b$  are the analyzed and background (a prior modeled) concentrations or AOD data,  $\mathbf{B}$  and  $\mathbf{O}$  are the background and observation error covariance matrices,  $\mathbf{H}$  is the observational operator,  $\mathbf{O}_o$  is the observations and  $J_c$  refers to the constraint terms. Both GSI's 3D-Var and OI use spatially varied background bias and observation to make the adjustment. However, each OI adjustment used in this study is made in each  $11 \times 11$  grid horizontally, and its effect expands up to the PBL height for its surface  $\text{PM}_{2.5}$  assimilation. The GSI's cost function reduction is performed for the whole domain, and its effect of the adjustment can be expanded on much greater horizontal and vertical scales, defined by its horizontal and vertical length scales, respectively.

In this study, we applied similar uncertainty settings for the OI as in Tang et al. (2015), in which the background's relative



**Figure 1.** GSI's background errors and length scales used in this study (ASO<sub>4J</sub> is the CMAQ's accumulation-mode sulfate aerosol).

uncertainties have horizontal and diurnal variation (Fig. 2 of Tang et al., 2015) for PM<sub>2.5</sub> assimilation over the PM<sub>2.5</sub> monitoring locations, and a fixed relative uncertainty of 80 % is applied to the other areas. In the OI horizontal adjustment, its B matrix has cut-off Gaussian-like horizontal correlations for each 11 × 11 grid adjustment. For OI's AOD assimilation, we use the fixed relative uncertainty of 80 % for modeled AOD. MODIS AOD retrieval quality depends on the land surface properties, and its uncertainty can range from 5 % over ocean to 20 % over land (Remer et al., 2005). Although the surface in situ-measured AQS PM<sub>2.5</sub> has much less measurement error than the satellite-retrieved AOD in most scenarios, the single surface station's representative error over the 12 km grid cell could be higher than the MODIS AOD, which is in 10 × 10 km resolution. In this study, we simply choose the relatively uncertainty of 10 % for both MODIS AOD and surface PM<sub>2.5</sub> observations. In this study, the PM<sub>2.5</sub> observations are located at 674 surface sites over CONUS, and all of them are used in data assimilations.

The settings of GSI are similar to Pagowski et al. (2014), in which the background error and length scales have vertical variance. Figure 1 shows that the vertical profiles of background errors and length scales for PM<sub>2.5</sub> (for PM<sub>2.5</sub> assimilation) and accumulation-mode sulfate (ASO<sub>4J</sub>, for AOD assimilation), which can be calculated using the GSI NMC (National Meteorological Center) method (Parrish and Derber, 1992) through the use of a tool called GEN\_BE (Descombes et al., 2015). ASO<sub>4J</sub> is one of the CMAQ aerosol species used in AOD data assimilation. Other aerosol species have similar model or background errors. The uncertainties of modeled aerosols are high in the lower layers and decrease with altitude. The model's horizontal length scale in-

creases the extent of the assimilation's horizontal expansion. In Fig. 1, the horizontal length scale for modeled PM<sub>2.5</sub> increases significantly around the altitude of 12 km, where the tropopause is usually located. The model's vertical length scale indicates the extent of the assimilation's vertical expansion, which is related to the strength of vertical advection, diffusion and convection. Below PBL, the vertical length scale is usually stronger than that in the upper layers. GSI was mainly used for assimilating meteorological variables, such as temperature and humidity, which usually have strong latitude variation. Therefore, the vertical profiles of background errors and length scales used in GSI for meteorological variables were usually varied depending on latitudes. In this study, our assimilation target is aerosols, whose concentrations and the corresponding variance could be caused by many factors, including emission, transport and chemical transformation, and whose latitude dependences are not significant. Therefore, we simply use one set of the vertical profiles for each aerosol species throughout the whole domain, which has no temporal variation, either. We use the same constant observation error of 0.1 in GSI for surface PM<sub>2.5</sub> and MODIS AOD. In both settings of GSI and OI, their background errors are far greater than the observation errors in lower altitudes, which push the adjusted values toward the observed surface PM<sub>2.5</sub>. For AOD assimilations, GSI's main increment is also in low altitudes due to its background error profile (Fig. 1), while OI's AOD assimilation applies one adjusting ratio to the whole column for each grid cell.

## 2.2 Calculations of aerosol optical depths

To assimilate AOD in the CMAQ model, we need to convert CMAQ aerosol chemical compositions. CMAQ includes

two methods for calculating AOD: the Mie method and the reconstruction method (Binkowski and Roselle, 2003). The Mie method calculates the aerosol optical extinction coefficient (AOE) from modeled aerosol physical characteristics, including the index of refraction, volume concentration and aerosol size distribution. It does not require aerosol chemical composition information and can handle the aerosols' internal mixture, in which each aerosol particle is composed of a solid core and coating layers of various compositions due to the aerosol uptake. In theory, the Mie method should yield an accurate AOE if all the model's aerosol physical properties are correct. However, that condition was hard to reach in many circumstances. Also, aerosol mass concentrations are often available in both models and observations, and a convenient method is needed to directly convert aerosol mass concentrations to AOE. Therefore, CMAQ provides another empirical approach, the reconstruction method (RM), which uses the mass concentrations of aerosol chemical compositions to calculate the total AOE with a lookup table. The RM assumes that all aerosols are externally mixed. It calculates each composition's AOE, and it sums them up to get the total AOE (Binkowski and Roselle, 2003) for the wavelength at 550 nm.

$$\begin{aligned} \text{AOE} \left[ \frac{1}{\text{km}} \right] = & 0.003 \cdot f(\text{RH}) \cdot \{(\text{NH}_4)_2\text{SO}_4 + \text{NH}_4\text{NO}_3\} \\ & + 0.004 \cdot \{\text{organic mass}\} + 0.01 \cdot \{\text{light absorbing carbon}\} \\ & + 0.001 \cdot \{\text{fine soil}\} + 0.0006 \cdot \{\text{coarse mass}\} \\ & + 0.00137 \cdot f_s(\text{RH}) \cdot \text{sea salt} \end{aligned} \quad (3)$$

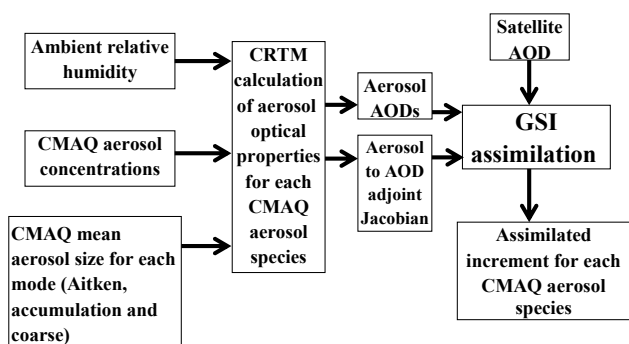
In CMAQ's RM, only the AOE from ammonium, sulfate, nitrate and sea salt are dependent on relative humidity (RH).  $f(\text{RH})$  and  $f_s(\text{RH})$  are two RH-dependent lookup tables of aerosol hygroscopic growth for sulfate/nitrate/ammonium and sea salt, respectively. (The original RM (Binkowski and Roselle, 2003) in earlier versions (< 5.1) of CMAQ does not include sea salt's RH dependence in AOE calculation.) All other aerosols are converted from their mass concentrations to the corresponding AOE with simple fixed constants. The conversion method does not need aerosol size distribution, but just the mass concentrations of aerosol compositions and ambient RH. Due to its simplicity and convenience, the RM is widely used not only for CMAQ aerosols, but also for converting observed aerosol mass concentrations to AOE (Malm et al., 1994; Roy et al., 2007). Tang et al. (2015) also use the RM-calculated AOD for OI assimilation, which is carried out here for this OI assimilation. In contrast, the Mie method based on aerosol physical characteristics is relatively hard to use in data assimilations as the data assimilations target the mass concentrations of aerosol compositions, not the aerosol physical characteristics directly. Although we can use adjusted aerosol mass concentrations to calculate the corresponding change in aerosol physical characteristics with CMAQ routines and then link it to AOD adjustment, the additional conversion makes these calculations difficult.

In theory, we should use the same forward method, such as RM, to calculate AOD used in GSI. However, the CMAQ's current RM is relatively simple, only for a single wavelength, 500 nm, without considering aerosol distribution. The current RM can not be directly used to calculate a multi-wavelength AOD, though this study only uses the 550 nm AOD. The existing GSI has its own tool, the Community Radiative Transfer Model (CRTM, Han et al., 2006; Liu and Weng, 2006), to handle its AOD calculation. CRTM provides GSI not only with the forward AOD, but also with the calculations of tangent-linear, adjoint and K-matrix for multiple wavelengths. As the GSI is tightly coupled with the CRTM for its AOD-related operations, we need to go through the CRTM to utilize GSI's existing AOD assimilation capability. However, the current version of CRTM does not support CMAQ's aerosol species and only handles GOCART (Goddard Chemistry Aerosol Radiation and Transport; Chin et al., 2000, 2002) aerosol species. GOCART's aerosol species are similar to the classification of aerosol used in the CMAQ RM, including sulfate, dust, sea salt, black carbon (BC) and organic carbon (OC), which also assumes that the aerosols are externally mixed. It is possible for us to represent CMAQ aerosols in GOCART aerosol categories to calculate their optical properties. Table 1 shows how the CMAQ 5.1 Aero6 species (Sonntag et al., 2014) are mapped to CRTM's GOCART aerosol. CMAQ aerosols have three size modes: Aitken (*i* mode), accumulation (*j* mode) and coarse mode (*k* mode), representing super-fine nucleation particles, aged coagulated particles and coarse particles, respectively (Binkowski and Roselle, 2003). Each size mode has its own lognormal size distribution (Whitby and McMurry, 1997). We applied the CMAQ-averaged aerosol size for each mode (Aitken, accumulation or coarse) to the CRTM AOD calculation. Unlike the CMAQ RM in which only AOE from sulfate, nitrate, ammonium and sea salt are dependent on RH, all the calculations for GOCART AOE except hydrophobic BC/OC depend on ambient RH, wavelength and aerosol sizes, which differs from RM's one-size-fits-all method. It should be noted that the GSI AOD assimilation uses the 54 CMAQ species, not the GOCART species. The concentrations of the 54 species, aerosol sizes and ambient RH are fed into CRTM, which generates the AOD of each CMAQ species and the corresponding adjoint Jacobian that is needed for the AOD assimilation (Fig. 2). In GSI's AOD assimilation, the control variables are the 54 CMAQ aerosol species, unlike the GSI's PM<sub>2.5</sub> assimilation in which only PM<sub>2.5</sub> is the control variable and all CMAQ aerosol species are adjusted based on the PM<sub>2.5</sub> incremental ratio. In this study, GSI does not assimilate MODIS AOD and surface PM<sub>2.5</sub> simultaneously due to the complexity of handling two observations types. For the 18:00 Z GSI assimilation, the assimilation was actually conducted in two steps: the first step is the GSI AOD assimilation, and the second step is the surface PM<sub>2.5</sub> assimilation, which was carried out upon the AOD assimilation adjustment. After the assimilations, we



**Table 1.** CMAQ 5.1 aerosol species and their mapping to GOCART aerosol optical properties in CRTM. (CMAQ species in *italic* are Aitken-mode aerosols, those in **bold** are coarse-mode aerosols and the rest are the accumulation-mode aerosols.)

Aerosol species in CMAQ 5.1	GOCART aerosols in CRTM
<i>ASO4I ANO3I ANH4I ACLI</i> <b>ASO4J ANO3J ANH4J ASO4K ANO3K ANH4K</b>	Sulfate aerosol
<i>AECI</i> <b>AECJ</b>	Black carbon aerosol
<i>APOCI APNCOMI</i> <b>APOCJ AOTHRJ</b> <i>AXYL1J AXYL2J AXYL3J</i> <b>ATOL1J ATOL2J ATOL3J</b> <i>ABNZ1J ABNZ2J ABNZ3J</i> <b>AISO1J AISO2J AISO3J</b> <i>ATRP1J ATRP2J</i> <b>ASQTJ AALK1J AALK2J</b> <i>AORG CJ AOLGBJ AOLGAJ</i> <b>APAH1J APAH2J APAH3J</b> <i>APN-COMJ</i>	Organic carbon aerosol
<i>AFEJ AALJ ASIJ ACAJ</i> <b>AMGJ AKJ AMNJ ACORS ASOIL</b>	Dust aerosol
<i>ANAJ ACLJ</i> <b>ACLK ASEACAT</b>	Sea salt aerosol

**Figure 2.** The schematic diagram of the GSI AOD assimilation.

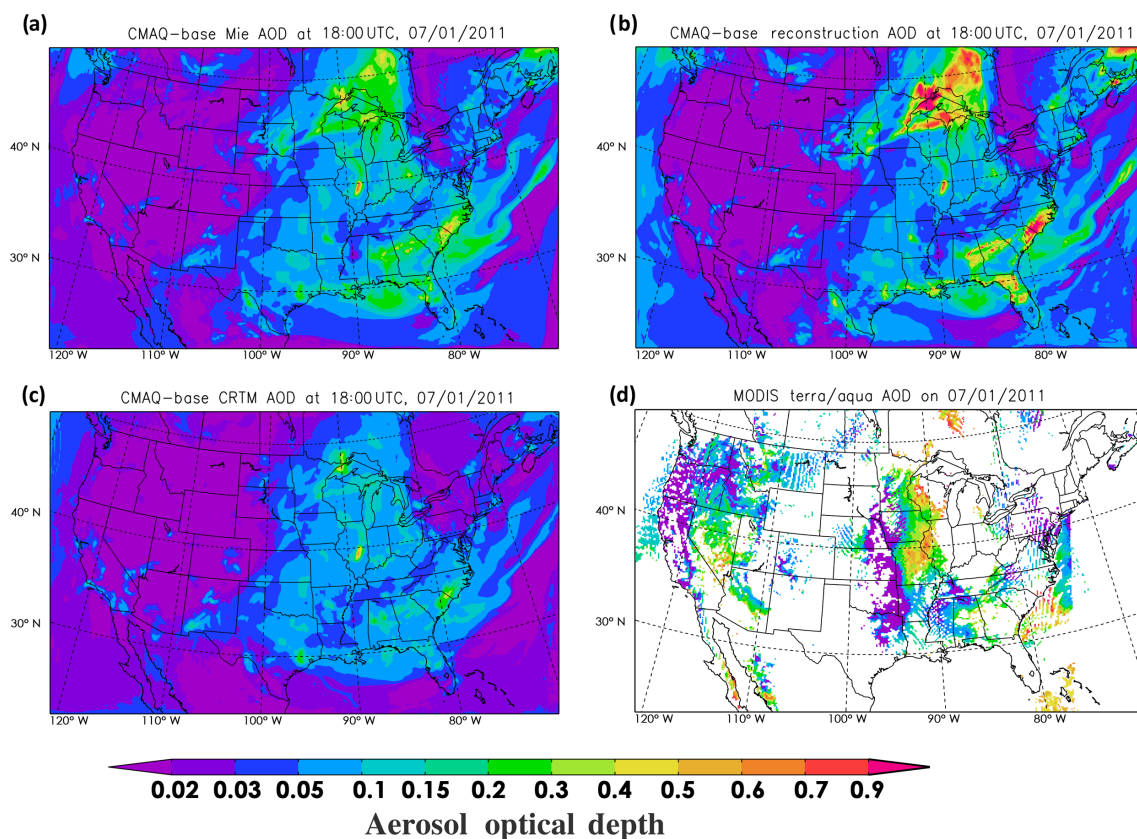
obtain the adjusted CMAQ aerosol concentrations, which are used as the initial condition of the next cycle of the CMAQ forward run.

Figure 3 shows one example of the AOD calculations from the three methods mentioned above compared to Terra/Aqua MODIS AOD data. The MODIS AODs in Fig. 3 are the daily data, and the overpass times of Terra and Aqua satellites on 1 July 2011 range from 15:00 to 22:00 UTC over the CONUS. Both OI and GSI use the AOD assimilation window of  $\pm 2$  h. During this event, some wildfires occurred in southern Canada and Wisconsin, and North Carolina and South Carolina, which caused the relatively high AOD values. These high AODs are also confirmed by the MODIS retrievals (Fig. 3d). Figure 3 shows that the different AOD calculations from the same CMAQ aerosol mass loadings yield a similar spatial distribution pattern with noticeable quantitative differences, and in general we see RM AOD > Mie AOD > CRTM AOD. In most regions, especially over the western United States, e.g., Nevada, MODIS AOD is higher than the three CMAQ AODs. The Mie method AOD and CRTM AOD are generally lower than MODIS AOD. Only the RM shows some sporadic overestimated AOD over

southern Canada and Wisconsin, and North and South Carolina. These differences will lead to corresponding differences in data assimilation adjustments. Roy et al. (2007) compared the CMAQ AODs with surface AOD observations from the nephelometer instrument over 25 Interagency Monitoring of Protected Visual Environment (IMPROVE) sites (most of them were located in national parks), and they found that the CMAQ RM yielded a surface AOD that is too high, but that agreed well with AERONET (Aerosol Robotic Network) and MODIS AOD in summer 2001. In Roy et al. (2007), CMAQ used the static lateral boundary condition, which may miss some elevated aerosol loadings (Lu et al., 2016). It is very likely that their CMAQ over-predicted near-surface AOD loading and underpredicted elevated AOD loading, but calculated total column AOD at the right magnitude. In another word, the converting factors of the RM used in Eq. (3) may be too high if their CMAQ predicted correct aerosol mass concentration over those national park sites.

### 2.3 PM<sub>2.5</sub> calculations

Besides the MODIS AOD, the data assimilations of OI and GSI also use surface PM<sub>2.5</sub> observations to make adjustments. The OI method for surface PM<sub>2.5</sub> assimilation was described in Tang et al. (2015). Pagowski et al. (2014) described the GSI method for surface PM<sub>2.5</sub> assimilation used in WRF–Chem in which aerosol size bins are fixed. CMAQ’s PM<sub>2.5</sub> calculation is slightly different as it is not defined by fixed size bins, but rather three size modes: Aitken, accumulation and coarse modes (*i*, *j*, *k* modes) (Appel et al., 2010).



**Figure 3.** AOD calculations using the Mie method (a), reconstruction method (b) and CRTM from the CMAQ's base case (before assimilations) (c) compared to MODIS AOD on 1 July 2011 (d). All the AODs are at 550 nm wavelength.

$$\begin{aligned}
 \text{PM}_{2.5} = & \text{PM}_{25\text{AT}} \cdot (\text{SO}_{4i} + \text{NO}_{3i} + \text{NH}_{4i} + \text{Na}_i + \text{Cl}_i \\
 & + \text{EC}_i + \text{POC}_i + \text{OTHER}_i) \\
 & + \text{PM}_{25\text{AC}} \cdot (\text{SO}_{4j} + \text{NO}_{3j} + \text{NH}_{4j} + \text{Na}_j + \text{Cl}_j \\
 & + \text{EC}_j + \text{EL}_j + \text{POC}_j + \text{SOA}_j + \text{OTHER}_j) \\
 & + \text{PM}_{25\text{CO}} \cdot (\text{SO}_{4k} + \text{NO}_{3k} + \text{NH}_{4k} + \text{Na}_k + \text{Cl}_k \\
 & + \text{SOIL}_k + \text{OTHER}_k), \quad (4)
 \end{aligned}$$

where  $\text{PM}_{25\text{AT}}$ ,  $\text{PM}_{25\text{AC}}$  and  $\text{PM}_{25\text{CO}}$  are the mass scaling factors for the three modes, indicating the mass ratios of each mode falling in the  $\text{PM}_{2.5}$  size range (Jiang et al., 2006), and their value ranges should be between 0 and 1, varied with time and location. POC is the primary organic carbon, and SOA represents the 23 secondary organic aerosols (Table 1). CMAQ 5.1 also includes eight additional mineral elements: Fe, Al, Si, Ti, Ca, Mg, K and Mg (Table 1) in its accumulation mode, represented by  $\text{EL}_j$ . The species OTHER represents all other aerosols (Table 1) in each mode. Equation (4) describes the forward operator for calculating  $\text{PM}_{2.5}$ , which is fed into GSI as background  $\text{PM}_{2.5}$  concentration. After the GSI estimates the total  $\text{PM}_{2.5}$  increment, the increment is proportionally allocated to each aerosol size mode with the corresponding mass scaling factor and then to each aerosol species. Thus, the final adjustment that OI and GSI

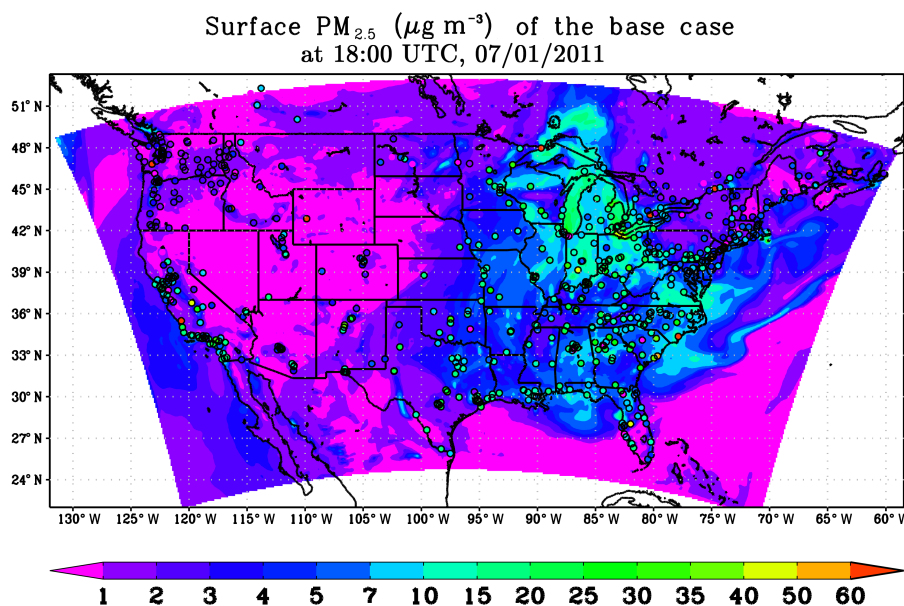
give back to CMAQ is the mass concentration increment for each CMAQ aerosol species.

### 3 Results and discussions

As the GSI and OI methods assimilate both surface  $\text{PM}_{2.5}$  and AOD, the impacts of these adjustments can be compared. Figure 4 shows the CMAQ raw predictions (referred to as base-case run) of surface  $\text{PM}_{2.5}$  compared to the measurements. In most regions west of  $100^\circ \text{W}$ , this a priori case underestimated the surface  $\text{PM}_{2.5}$ . Surface measurement also shows some sporadic high  $\text{PM}_{2.5}$  values in the eastern United States, missed by the base-case run. The goal of the data assimilations is to reduce these errors.

#### 3.1 The impact of data assimilation on aerosol mass concentration

Since both  $\text{PM}_{2.5}$  and aerosol AOD represent the concentrations of 54 aerosol species (Table 1), to understand the data assimilation method performances, we choose the accumulation-mode or  $j$  mode sulfate ( $\text{ASO4J}$ ) to illustrate the impact of data assimilation. Other aerosol species have similar adjustments in the GSI and OI assimilations. We set



**Figure 4.** Predicted surface  $\text{PM}_{2.5}$  from CMAQ base run compared to surface measurement (color-coded circle) at 18:00 UTC, 1 July 2011.

up three experiments to show the assimilation effect, one with just  $\text{PM}_{2.5}$ , one with AOD and one with both for the 18Z cycle. Figure 5a and b show that the GSI assimilation with surface  $\text{PM}_{2.5}$  yields more significant changes than the corresponding OI assimilation. The OI assimilation of  $\text{PM}_{2.5}$  leads to more localized increments compared to GSI since the OI increment is spread over  $11 \times 11$  grid cells (i.e., an area of  $17\,424\text{ km}^2$ ), while the GSI increment spread depends on the horizontal length scales. Their adjustments also differ over certain areas. For instance, the base case underestimates  $\text{PM}_{2.5}$  in Chicago, but it overestimates  $\text{PM}_{2.5}$  in southeastern Wisconsin (Fig. 4). The OI adjustment also has local variation: increase ASO4J in Chicago and decrease it over southeastern Wisconsin (Fig. 5b). The GSI's result differs significantly due to its length scales and cost function reduction over the whole domain. It yields overall  $\text{PM}_{2.5}$  reduction in the surrounding areas of Chicago, with smaller correction. Similar differences are also noted in other areas and other periods.

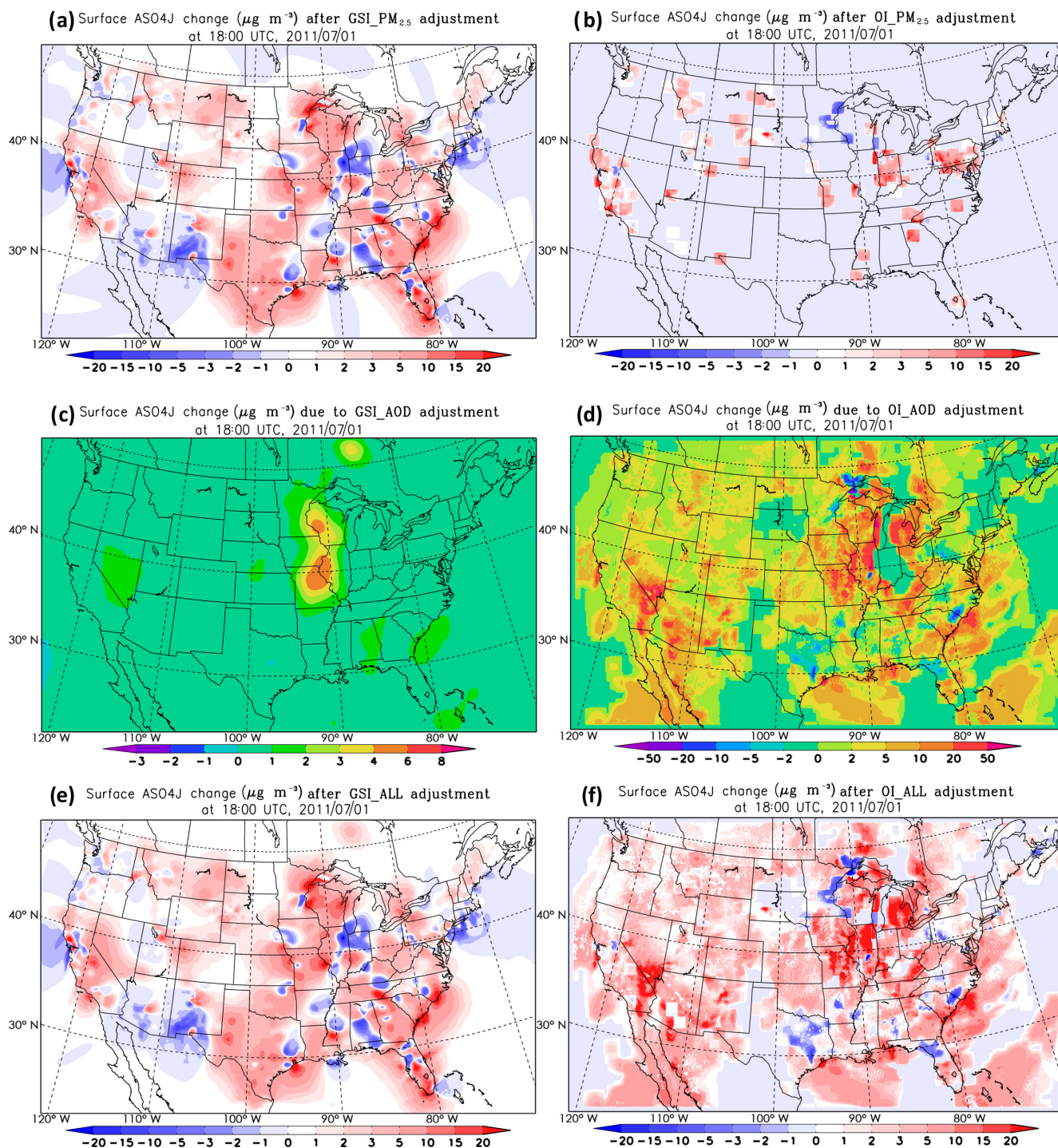
The impacts of AOD assimilations via GSI and OI also differ significantly (Fig. 5c and d). The model relative uncertainty used in the OI assimilation is relatively high, which caused the strong adjustment in ASO4J (Fig. 5d) due to OI's AOD assimilation over Nevada and southern California, Illinois and Michigan. The OI assimilation of AOD increases ASO4J over most of the model domain except in certain areas, such as northeastern Minnesota and southern Canada and the eastern part of the border of North Carolina and South Carolina, where the OI assimilation leads to a decrease in ASO4J (Fig. 5d). The GSI assimilation of AOD tends to be more moderate and smoother. Its increment is almost 1 order of magnitude smaller than the variation of the OI AOD as-

simulation (Fig. 5c). Figure 5c shows that the majority of increments of the GSI AOD assimilation occur in eastern California and Nevada, Missouri and Illinois (surrounding areas of St. Louis) and Wisconsin.

Figure 5e and f show the combined effect of assimilation of both surface  $\text{PM}_{2.5}$  and AOD on surface ASO4J from GSI and OI methods, respectively. For GSI, the impact of assimilating surface  $\text{PM}_{2.5}$  is greater than that of assimilating MODIS AOD. The OI assimilations show the opposite effect with AOD assimilation having a bigger impact than its surface  $\text{PM}_{2.5}$  assimilation.

It should be noted that the difference between GSI and OI increments is not only in their horizontal distribution, but also in their vertical distribution. GSI has a height-dependent background error profile (Fig. 1) while OI applies the uniform adjustment ratio to either the whole column for AOD assimilation or below PBL for surface  $\text{PM}_{2.5}$  assimilation. Figure 6 shows the model-predicted PBL height at this time. In most portions of the CONUS domain except the Rocky Mountain region, the PBL height is less than 3 km. This means that the impact of OI  $\text{PM}_{2.5}$  assimilation should be below 3 km for most regions. Considering the locations of available surface  $\text{PM}_{2.5}$  monitoring stations and the regional distribution of background  $\text{PM}_{2.5}$ , the vertical extension of the increments of OI  $\text{PM}_{2.5}$  should be more limited. Figure 7 shows the ASO4J increments similar to Fig. 5 but for the model's eighteenth layer, or roughly 2 km above ground. Figure 7b shows that the OI  $\text{PM}_{2.5}$  adjustment at the 2 km layer only appears in sporadic locations, such as Raleigh/Durham (central North Carolina), Atlanta and Denver downwind areas. Compared to Figs. 5b and 6, we find that in most CONUS regions, the OI  $\text{PM}_{2.5}$ 's ASO4J incre-

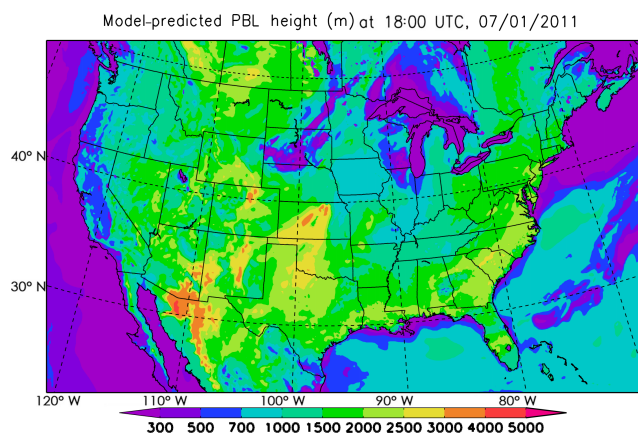




**Figure 5.** Surface accumulation-mode sulfate (ASO4J) changes after GSI (left) and OI (right) assimilations with surface  $\text{PM}_{2.5}$  and MODIS AOD at 550 nm.

ments do not show up in the 2 km layer because either their PBLs are lower than 2 km, or there is no strong surface adjustment due to the locations of monitoring stations. Conversely, GSI  $\text{PM}_{2.5}$  assimilation shows the horizontal distribution of the ASO4J increment at 2 km, similar to its surface

increment (Figs. 7a and 5a), but with a much smaller magnitude than the increment of OI  $\text{PM}_{2.5}$  assimilation at the 2 km layer. The GSI AOD assimilation also yields a similar ASO4J adjustment pattern with smaller increments than OI AOD (Figs. 7c and 5c). The strongest adjustment at the



**Figure 6.** Model-predicted PBL height (m) at 18:00 UTC, 1 July 2011.

2 km layer comes from the OI AOD assimilation, which uses the same adjustment ratio for the whole vertical column. Its impact on 2 km ASO4J is more than 1 order of magnitude stronger than the corresponding GSI assimilation due to its relatively high uncertainty setting. In terms of combined effect at the 2 km layer, for GSI, the AOD assimilation is almost as equally important as the surface  $\text{PM}_{2.5}$  assimilation, depending on the region (Fig. 7e). For the OI assimilation, the AOD's impact dominates at the 2 km layer as the surface  $\text{PM}_{2.5}$ 's effect only appear sporadically (Fig. 7f).

### 3.2 The impact of data assimilation on AODs

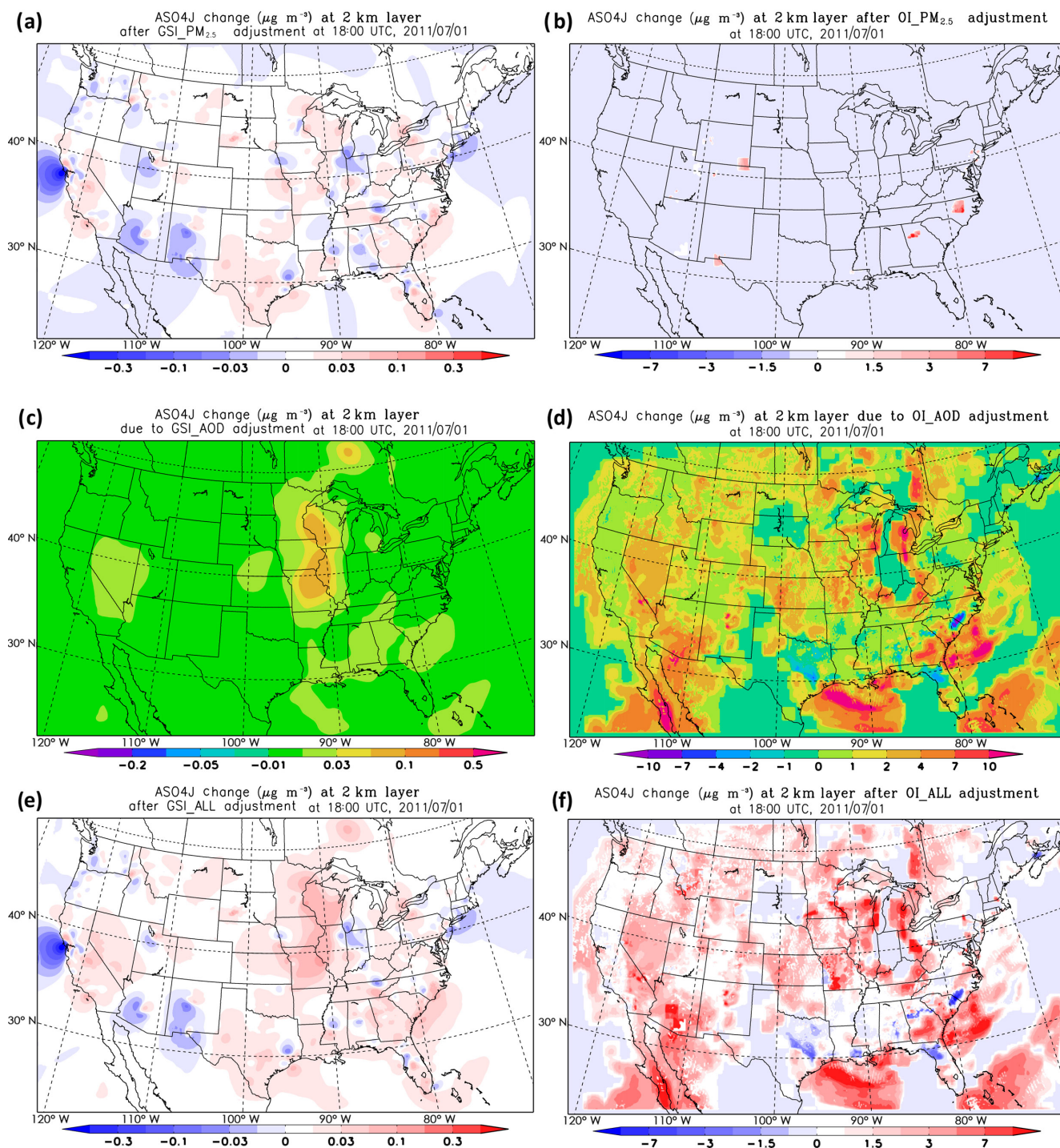
The above discussion is about the data assimilations' impacts on one aerosol mass concentration. It is also still necessary to assess the impacts on column AODs as the AOD is used here to carry out assimilations. Fortunately, both CRTM and RM are composition-based methods for externally mixed aerosols, and we can easily calculate the AOD changes due to the changes of aerosol mass concentrations. Figure 8 shows the CRTM AOD changes due to GSI assimilation (left panels), and RM AOD changes due to OI assimilation (right panels). Their spatial distribution patterns are very similar to the corresponding surface ASO4J increments in Fig. 5 as most high aerosol loadings are near surface. However, the AOD increment's value range of GSI  $\text{PM}_{2.5}$  assimilation is more than 1 order of magnitude lower than that of OI  $\text{PM}_{2.5}$ . One reason is that the CRTM method yields 2 or 3 times lower AOD than the RM with the same aerosol loading (Fig. 2b and c). Another reason, or the major reason, is that the GSI  $\text{PM}_{2.5}$  assimilation has a much lower increment on total aerosol mass loading, reflected by the magnitude difference between Fig. 6a and b. As the GSI assimilation uses a strong altitude-dependent background error profile (Fig. 1), it makes its major adjustment near surface and yields smaller overall adjustment for total column aerosol mass loading. Conversely, the OI  $\text{PM}_{2.5}$  assimilation applies the same adjustment ratio to

the aerosol mass below PBL, and the adjustment could be much stronger than that of GSI  $\text{PM}_{2.5}$  assimilation for the elevated layers below PBL. We can see similar patterns and differences due to their AOD assimilations (Fig. 8c and d). The AOD increments due to GSI AOD and OI AOD have a similar horizontal distribution to their corresponding ASO4J mass increments (Figs. 5c, 7c and 8c; Figs. 5d, 7d, and 8d), but the OI AOD has much stronger adjustments on total AOD and elevated aerosol mass. Generally, the GSI AOD tends to increase AOD or aerosol column loading in the northern central United States and over Nevada (Fig. 8c). The OI AOD assimilation has a similar AOD increment over these two regions, but a decreased AOD or aerosol column loading over Texas, northern Florida, southern central Canada and the border of North Carolina and South Carolina (Fig. 8d). It should be noted that the OI  $\text{PM}_{2.5}$  assimilation yields AOD enhancements in the southern and eastern United States, though just sporadically (Fig. 8b). Therefore, there is a conflict as the OI  $\text{PM}_{2.5}$  and OI AOD assimilations pointed to opposite adjusting directions over some areas, implying that the RM AOD could yield some overpredictions. In this situation, the combined OI assimilations will increase below-PBL aerosol mass according to the adjustment of OI  $\text{PM}_{2.5}$ , but reduce the above-PBL aerosol mass to fit the overall AOD reduction and obtain a compromise of results over these two regions (Fig. 8f). This conflict-resolving process will change the vertical distribution of CMAQ aerosols. Figure 8 shows that the overall AOD increments are mainly due to  $\text{PM}_{2.5}$  assimilation in GSI and AOD assimilation in OI. The OI adjustments of AOD are much stronger than those of GSI, which is mainly due to adjustments at elevated layers.

### 3.3 The overall assimilation impacts over longer periods

We continue the CMAQ runs after the 18:00 UTC assimilations. After 1 h, or at 19:00 UTC, we compared their surface  $\text{PM}_{2.5}$  prediction with the corresponding measurement again to see their impacts (Fig. 9). At this time, the effects of data assimilations had been transported to downstream areas. Without the data assimilation, the base case (Fig. 9a) systematically underestimated the  $\text{PM}_{2.5}$  in the western United States, which is consistent with that shown in Fig. 4. Both assimilations correct some of the biases, which is more obvious in areas west of 90° W. For instance, both assimilations correct the  $\text{PM}_{2.5}$  underprediction in Iowa. In some locations, OI obtains a better result, such as in southern Nevada and southern Utah, and Kansas, where only one surface observation is available. In other locations, the GSI achieves better results, such as in southeastern Wisconsin. Sometimes the data assimilations overcorrected and yielded worse results than the base run, such as the OI's overestimation over Lake Michigan and GSI's overprediction over central Illinois (Fig. 9b and c). The overcorrection issue is more evident in the OI run as the adjustment of the OI assimilation is stronger





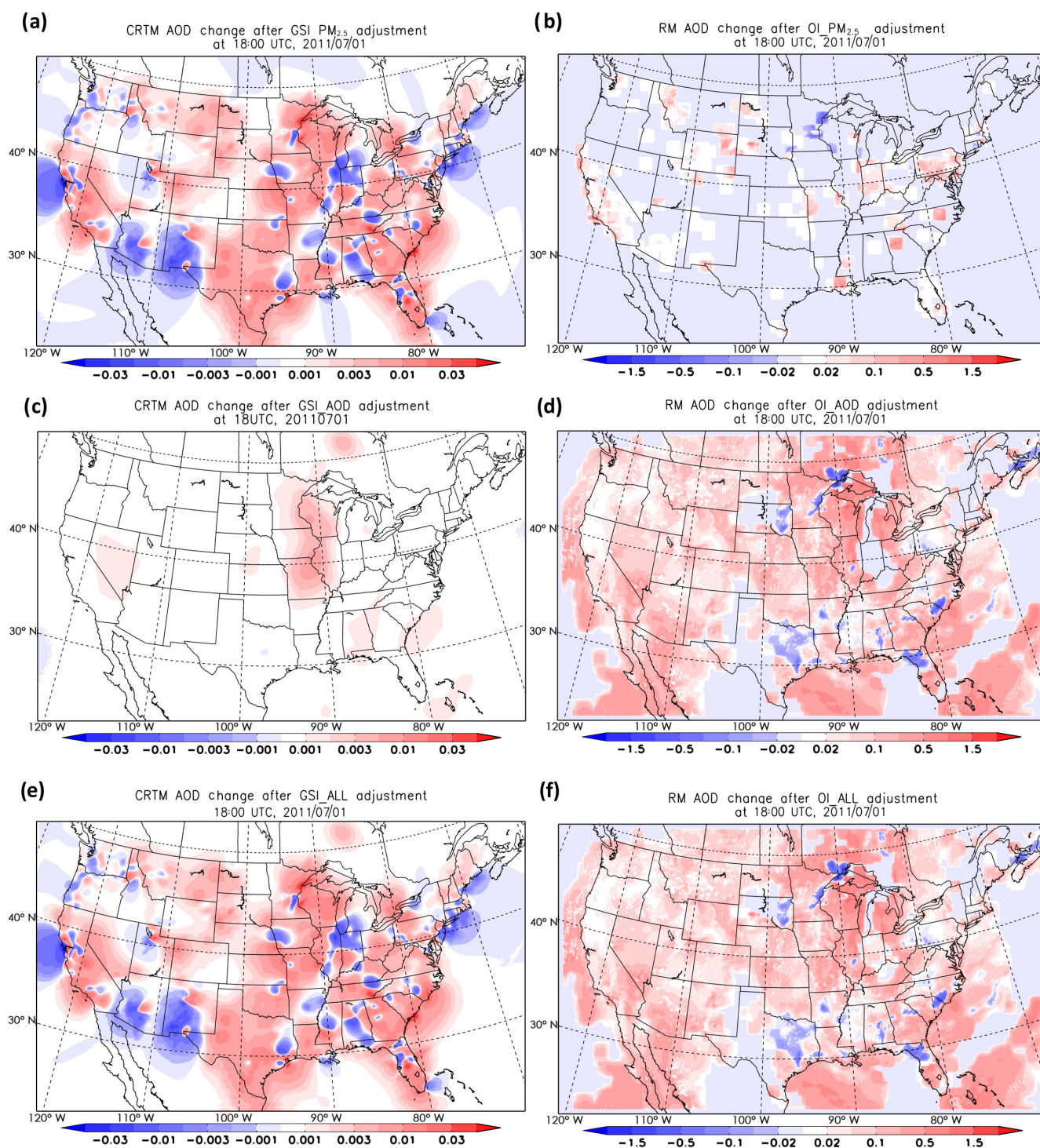
**Figure 7.** Same as Fig. 5 but for the model's 2 km layer.

than that of GSI, due to OI's stronger setting for model uncertainties. This strong setting is actually helpful sometimes; for instance, the OI assimilation is strong enough to correct the underpredicting bias over southeastern North Carolina, where GSI's moderate correction only helps reduce that bias.

The most evident side effect of the OI's overcorrection is the increase of root mean squared error (RMSE).

In this study, we employ four-cycles-per-day data assimilations; the MODIS AOD assimilation is only applied in the 18:00 UTC cycle for both the GSI and OI. We continue these runs for the whole of July 2011 and conduct 1-month con-

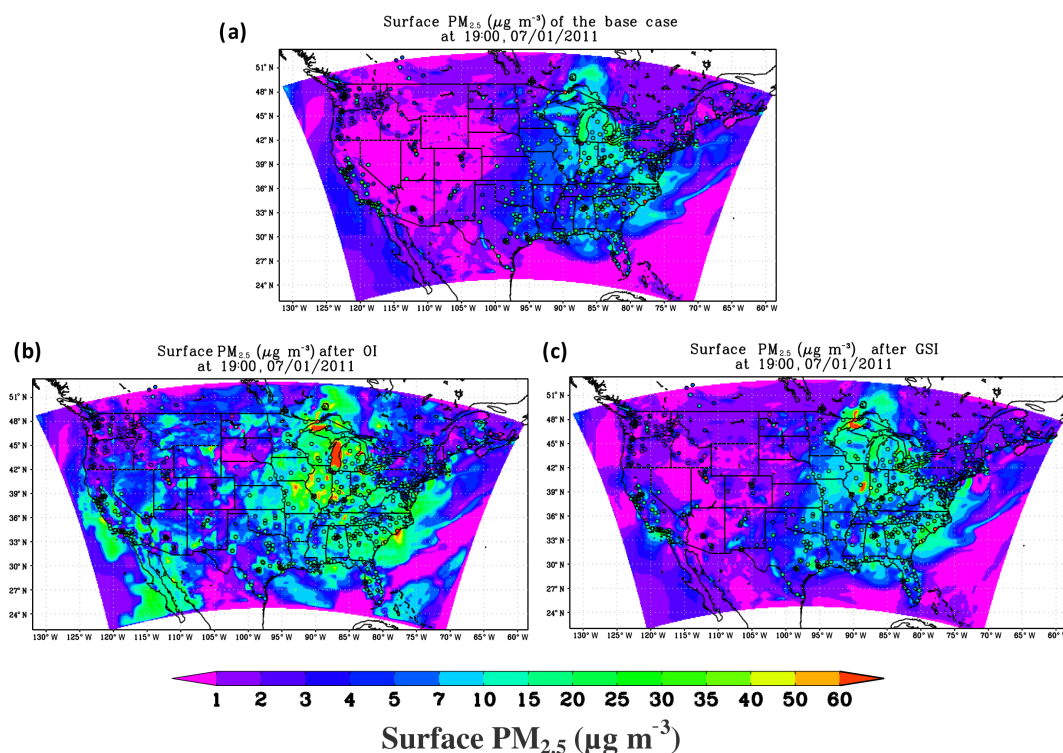




**Figure 8.** Same as Fig. 5 but for AOD (at 550 nm) change. The left panels show the CRTM AOD change due to GSI adjustments, and the right panels show the RM AOD change for OI.

tinuous verification based on the 674 CONUS surface sites. In order to avoid using the same data in data assimilation and verification, we skip verifying the initial condition of each 6 h run. For 18:00 UTC data assimilation and forecast,

the verification is conducted from 19:00 to 00:00 UTC, and the 00:00 UTC cycle verification is conducted from 01:00 to 06:00 UTC. The verification for the rest of the cycles follows a similar time shift. Figure 10 shows the time-series plots

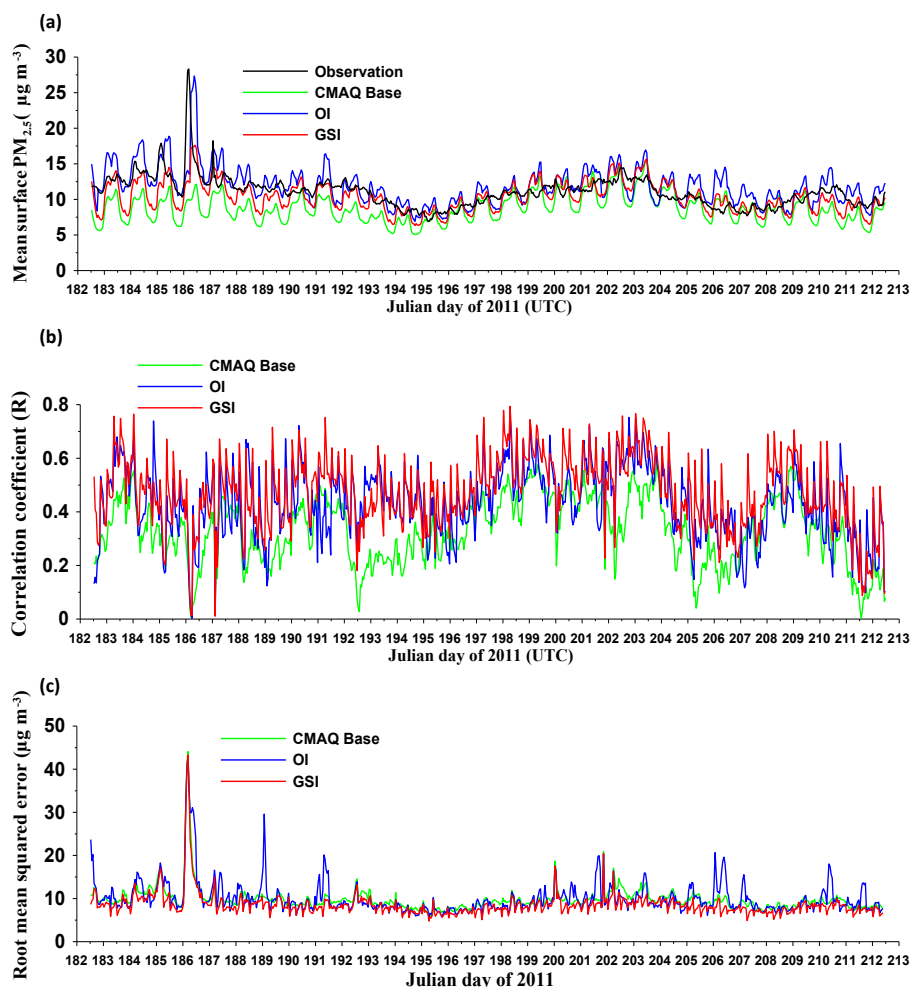


**Figure 9.** Predicted surface  $\text{PM}_{2.5}$  from the base case (a), OI (b) and GSI (c) runs compared to surface observation at 19:00 UTC, 1 July 2011.

of these CMAQ predictions for surface  $\text{PM}_{2.5}$ , their correlation coefficient ( $R$ ) and the RMSE. Before 14 July or Julian day 195, the CMAQ base prediction systemically underpredicted the  $\text{PM}_{2.5}$ . After that date, the base model tends to underpredict during daytime, but slightly overpredict during nighttime, except for the last three days of July 2011 when the overall underprediction appeared again. Both GSI and OI help reduce these underprediction biases. The OI's correction is stronger due to its stronger setting. It became more evident during the  $\text{PM}_{2.5}$  peak period caused by firework emissions on the night of 4 July (US Independence Day), which is about early morning of 5 July (Julian day 186) in UTC time (Fig. 10a). The OI assimilation caught the  $\text{PM}_{2.5}$  spike caused by fireworks, though its peak time was later than that of the observation. The GSI assimilation showed moderate correction that was not strong enough to fully match observations. The OI run shows some overprediction during the nighttime, especially later in July. Figure 10b clearly shows the effect of four assimilations per day, represented by the four-times-per-day enhancement of GSI and OI's correlation coefficients. For most of this month, we can generally see  $\text{GSI} > \text{OI} > \text{CMAQ base}$  for  $R$ . The difference of  $R$  between GSI and OI is much smaller than that between the assimilation runs and the CMAQ base case. GSI's  $R$  is consistently better than the CMAQ base, while the OI's  $R$  is not always better, e.g., during the Julian days of 191 and 207, though we can still find several periods when OI yielded the highest

$R$ . In terms of RMSE, the GSI's RMSE is always lower than that of the base run, and the OI run shows some RMSE spikes (Fig. 10c) due to its localized correction and strong setting.

Table 2 shows the corresponding statistics for the whole domain and the specific regions. The CMAQ base case systematically underpredicted surface  $\text{PM}_{2.5}$  with a mean bias of  $-2.25 \mu\text{g m}^{-3}$  over CONUS during this period. With data assimilations, the OI and GSI runs obtain mean biases of 0.77 and  $-0.73 \mu\text{g m}^{-3}$ , respectively. Besides this effect, their correlation coefficient,  $R$ , is also improved: the base case's  $R$  is 0.3, and  $R$ s of OI and GSI runs are 0.38 and 0.44 over CONUS, respectively. Their indexes of agreement (IOA) have corresponding improvements, too. Among the regions, the data assimilation yields most significant improvements over the Pacific coast and the southeastern United States, where the CMAQ base case has relatively poor correlation coefficients. In all of these regions, the GSI yields overall best correlation coefficient and RMSE, while the OI run has the smallest mean biases except for the southern central region, where the GSI's mean bias is the lowest. OI's localized correction and strong setting help to obtain better mean biases, but they cause the overcorrection issue as well as the increased RMSE. Over the northeastern United States where relatively dense surface observations are available, the OI's RMSE increase is relatively small. Although both assimilations generally improve the  $\text{PM}_{2.5}$  prediction, their performance can be highly varied depending on the region. For



**Figure 10.** Time-series comparisons of CMAQ surface  $PM_{2.5}$  (base, OI and GSI) vs. the observation over the CONUS domain for their mean value (a), correlation coefficient (b) and root mean squared error (c).

instance, over Rocky Mountain states, where both surface observations and MODIS AOD are limitedly available with complex terrain, the GSI and OI assimilations slightly improve the  $R$  and IOA, though they evidently improve the mean bias of surface  $PM_{2.5}$ .

All these runs have daytime underprediction bias issue from 12:00 to 00:00 UTC. The four-cycles-per-day data assimilations help reduce that bias, but the bias trend is still there. It implies that the data assimilation for CMAQ's initial condition has certain limitations, and it may not be able to solve the prediction bias alone. In addition to uncertainties in initial conditions, forecast errors also depend on errors in the models' meteorology, chemistry and emissions, and we need other adjustments to correct these biases.

#### 4 Conclusions

In this study, we expanded the GSI assimilation to CMAQ 5.1's Aero6 aerosol species. The four-cycles-per-day aerosol data assimilations for surface  $PM_{2.5}$  and AOD were carried out with GSI and OI (Tang et al., 2015) methods over the CONUS. The results were compared to surface  $PM_{2.5}$  observations, and this showed that both assimilations generally improved the aerosol predictions. The increments resulting from our OI assimilation are spread in  $11 \times 11$  horizontal grid cells, while the increment spread in GSI (a 3D-Var assimilation technique) is controlled by its background error variances and horizontal and vertical length scales. GSI's cost function reduction is performed for the whole domain. The differences in formulation of these two methods led to their different patterns of adjusting the initial conditions, and GSI yielded smoother and horizontally broader adjustment, but much weaker vertical increment. The local OI method uses a simple one-ratio-fits-all method for its vertical adjustment



**Table 2.** Regional statistics of the three simulations (CMAQ\_BASE, OI and GSI) for surface PM<sub>2.5</sub> in Jul 2011.

Regions	Simulations	Mean bias ( $\mu\text{g m}^{-3}$ )	Root mean squared error ( $\mu\text{g m}^{-3}$ )	Correlation coefficient, <i>R</i>	Index of agreement
CONUS	BASE	−2.25	10.57	0.3	0.54
	OI	0.77	11.07	0.38	0.59
	GSI	−0.73	9.42	0.44	0.64
Northeastern United States	BASE	−3.36	11.02	0.27	0.54
	OI	0.78	11.17	0.42	0.64
	GSI	−1.18	9.70	0.44	0.66
Pacific coast	BASE	−2.24	8.74	0.23	0.40
	OI	0.71	9.13	0.40	0.60
	GSI	−0.80	7.86	0.44	0.61
Southeastern United States	BASE	−1.86	11.05	0.23	0.50
	OI	0.12	9.88	0.41	0.63
	GSI	−0.58	9.30	0.41	0.63
Rocky Mountain states	BASE	−3.22	12.14	0.09	0.24
	OI	1.59	13.28	0.13	0.30
	GSI	−1.63	11.75	0.16	0.28
Northern central	BASE	−1.55	12.29	0.22	0.49
	OI	0.96	13.83	0.28	0.51
	GSI	−0.26	11.12	0.35	0.59
Southern central	BASE	−1.1	9.49	0.09	0.4
	OI	0.38	8.07	0.27	0.52
	GSI	−0.1	7.58	0.27	0.54

up to PBL height for PM<sub>2.5</sub> assimilation, or the whole column for AOD assimilation. This OI method can use strong setting to achieve a better mean bias, but has the side effect of an RMSE increase due to its localized correction. Overall, GSI's adjustment yield better results, even with its moderate setting of data assimilation parameters, shown by its better statistics. One important reason is GSI's whole-domain cost function reduction, which helps constrain its RMSE, and a longer horizontal length scale, especially for PM<sub>2.5</sub> assimilation (Fig. 1), which helps expand its adjusting effect to relatively broad areas. Both assimilations highly depend on the available observations. Compared to GSI's massive code, the OI code is much smaller and portable, which may be a good choice for some light-duty usages.

AOD assimilations have more issues than the surface PM<sub>2.5</sub> assimilation, which is not only about the methods for converting aerosol mass concentrations to AOD that relies on the model for ambient RH, aerosol sizes and speciation, but also depends on a prior model for the aerosol vertical distribution. The CRTM AOD is about 2–3 times lower than RM AOD with the same aerosol mass loadings (Fig. 3). From the existing evidence, it is seen that the converting factors used in RM AOD may be too high (Roy et al., 2007), or the RM's one-size-fits-all method may not resolve highly varied aerosol size impact on AOD calculation. There is another is-

sue of aerosol specification that is particularly important for organic aerosols as CMAQ 5.1 has 23 SOAs plus primarily emitted OC (POC). All of the organic aerosols were assumed to have the same optical properties in CRTM and RM, which is obviously an approximate assumption. Liu et al. (2016) show that SOA optical properties could be highly varied depending on the chemical species, aging time and ambient NO<sub>x</sub> concentrations. Both CRTM and RM assume that all aerosols are externally mixed, which may not best fit the situation in the real world. These uncertainties need to be addressed in future through verification with observations. In this study, we only adjusted the aerosol mass concentrations, and we did not adjust the aerosol composition, size and vertical distributions, which could have a big impact on the AOD calculations. The data assimilation for the initial condition also has its limitations, and its adjustment effect could fade away with time if there is a persistent model bias. Therefore, more frequent adjustment would be helpful. Unfortunately the MODIS AOD assimilation used in this study is applied only once per day in the data assimilations due to data availability. More satellite AOD data, such as VIIRS (Visible Infrared Imaging Radiometer Suite) and GOES-R (Geostationary Operational Environmental Satellite-R Series) ABT (Advanced Baseline Imager) AODs, should make this assimilation more useful. Another approach is to use a

more complex and costly four-dimensional (4-D) variational data assimilation (4D-Var) to correct the model's persistent biases, which integrated the data assimilation method within the CTM (Chai et al., 2006). All of these issues should be addressed in future studies.

**Code availability.** This study includes forward simulations and data assimilation tools. The meteorological code of WRF 3.4.1 can be downloaded from [http://www2.mmm.ucar.edu/wrf/users/download/get\\_source.html](http://www2.mmm.ucar.edu/wrf/users/download/get_source.html). CMAQ 5.1 can be downloaded from <https://www.cmascenter.org/download.cfm>. The GSI data assimilation tool can be downloaded from <http://www.dtcenter.org/com-GSI/users.v3.5/downloads/index.php>. All other codes and the modified codes can be provided upon request.

**Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** We gratefully acknowledge the funding support from NASA HAQAST (grant no. NNX16AQ19G) and NOAA USWRP (grant no. NA16OAR4590120).

Edited by: Samuel Remy

Reviewed by: three anonymous referees

## References

- Adhikary, B., Kulkarni, S., Dallura, A., Tang, Y., Chai, T., Leung, L. R., Qian, Y., Chung, C. E., Ramanathan, V., and Carmichael, G. R.: A regional scale chemical transport modeling of Asian aerosols with data assimilation of AOD observations using optimal interpolation technique, *Atmos. Environ.*, 42, 8600–8615, 2008.
- Appel, K. W., Roselle, S. J., Gilliam, R. C., and Pleim, J. E.: Sensitivity of the Community Multiscale Air Quality (CMAQ) model v4.7 results for the eastern United States to MM5 and WRF meteorological drivers, *Geosci. Model Dev.*, 3, 169–188, <https://doi.org/10.5194/gmd-3-169-2010>, 2010.
- Binkowski, F. S. and Roselle, S. J.: Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component: 1. Model description, *J. Geophys. Res.*, 108, 4183, <https://doi.org/10.1029/2001JD001409>, 2003.
- Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camañó, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, *Atmos. Chem. Phys.*, 15, 5325–5358, <https://doi.org/10.5194/acp-15-5325-2015>, 2015.
- Chai, T., Carmichael, G. R., Sandu, A., Tang, Y., and Daescu, D. N.: Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements, *J. Geophys. Res.-Atmos.*, 111, D02301, <https://doi.org/10.1029/2005JD005883>, 2006.
- Chai, T., Kim, H. C., Pan, L., Lee, P., and Tong, D.: Impact of Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol optical depth (AOD) and AirNow PM<sub>2.5</sub> assimilation on Community Multi-scale Air Quality (CMAQ) aerosol predictions over the contiguous United States, *J. Geophys. Res.-Atmos.*, 122, 5399–5415, <https://doi.org/10.1002/2016JD026295>, 2017.
- Chin, M., Rood, R. B., Lin, S.-J., Muller, J.-F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global model GOCART: model description and global properties, *J. Geophys. Res.-Atmos.*, 105, 24671–24687, <https://doi.org/10.1029/2000JD900384>, 2000.
- Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., and Higurashi, A.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements, *J. Atmos. Sci.*, 59, 461–483, 2002.
- Descombes, G., Auligné, T., Vandenberghe, F., Barker, D. M., and Barré, J.: Generalized background error covariance matrix model (GEN\_BE v2.0), *Geosci. Model Dev.*, 8, 669–696, <https://doi.org/10.5194/gmd-8-669-2015>, 2015.
- Elbern, H. and Schmidt, H.: A 4D-Var chemistry data assimilation scheme for Eulerian chemistry transport modelling, *J. Geophys. Res.-Atmos.*, 104, 18583–18598, 1999.
- Elbern, H. and Schmidt, H.: Ozone episode analysis by four dimensional variational chemistry data assimilation, *J. Geophys. Res.-Atmos.*, 106, 3569–3590, <https://doi.org/10.1029/2000JD900448>, 2001.
- Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modelling, *J. Geophys. Res.-Atmos.*, 102, 15967–15985, <https://doi.org/10.1029/97JD01213>, 1997.
- Elbern, H., Schmidt, H., Talagrand, O., and Ebel, A.: 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environ. Modell. Softw.*, 15, 539–548, 2000.
- Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmos. Chem. Phys.*, 7, 3749–3769, <https://doi.org/10.5194/acp-7-3749-2007>, 2007.
- Han, Y., van Delst, P., Liu, Q., Weng, F., Yan, B., Treadon, R., and Derber, J.: JCSDA Community Radiative Transfer Model (CRTM) – Version 1, NOAA Tech. Rep. NESDIS 122, 33 pp., NOAA, Silver Spring, MD, 2006.
- Jiang, W., Smyth, S., Giroux, E., Roth, H., and Yin, D.: Differences between CMAQ fine mode particle and PM<sub>2.5</sub> concentrations and their impact on model performance evaluation in the lower Fraser valley, *Atmos. Environ.*, 40, 4973–4985, 2006.
- Lee, P., McQueen, J., Stajner, I., Huang, J., Pan, L., Tong, D., Kim, H., Tang, Y., Kondragunta, S., Ruminski, M., and Lu, S.: NAQFC developmental forecast guidance for fine particulate matter (PM<sub>2.5</sub>), *Weather Forecast.*, 32, 343–360, 2017.
- Levy, R. C., Remer, L. A., Tanré, D., Mattoo, S., and Kaufman, Y. J.: Algorithm for remote sensing of tropospheric aerosol over dark targets from MODIS: Collections 005 and 051, NASA Goddard Space Flight Center, Greenbelt, MD, available at: [https://modis-images.gsfc.nasa.gov/\\_docs/ATBD\\_MOD04\\_C005\\_rev2.pdf](https://modis-images.gsfc.nasa.gov/_docs/ATBD_MOD04_C005_rev2.pdf) (last access: 20 December 2017), revision2, 2009.

- Liu, J., Lin, P., Laskin, A., Laskin, J., Kathmann, S. M., Wise, M., Caylor, R., Imholt, F., Selimovic, V., and Shilling, J. E.: Optical properties and aging of light-absorbing secondary organic aerosol, *Atmos. Chem. Phys.*, 16, 12815–12827, <https://doi.org/10.5194/acp-16-12815-2016>, 2016.
- Liu, Q. and Weng, F.: Advanced doubling-adding method for radiative transfer in planetary atmospheres, *J. Atmos. Sci.*, 63, 3459–3465, 2006.
- Liu, Z., Liu, Q., Lin, H.-C., Schwartz, C. S., Lee, Y.-H., and Wang, T.: Three-dimensional variational assimilation of MODIS aerosol optical depth: implementation and application to a dust storm over East Asia, *J. Geophys. Res.-Atmos.*, 116, D22306, <https://doi.org/10.1029/2011JD016159>, 2011.
- Lu, C.-H., da Silva, A., Wang, J., Moorthi, S., Chin, M., Colarco, P., Tang, Y., Bhattacharjee, P. S., Chen, S.-P., Chuang, H.-Y., Juang, H.-M. H., McQueen, J., and Iredell, M.: The implementation of NEMS GFS Aerosol Component (NGAC) Version 1.0 for global dust forecasting at NOAA/NCEP, *Geosci. Model Dev.*, 9, 1905–1919, <https://doi.org/10.5194/gmd-9-1905-2016>, 2016.
- Malm, W. C., J. F. Sisler, D. Huffman, R. A. Eldred, and Cahill, T. A.: Spatial and seasonal trends in particle concentration and optical extinction in the United States, *J. Geophys. Res.*, 99, 1347–1370, 1994.
- Pagowski, M., Liu, Z., Grell, G. A., Hu, M., Lin, H.-C., and Schwartz, C. S.: Implementation of aerosol assimilation in Gridpoint Statistical Interpolation (v. 3.2) and WRF-Chem (v. 3.4.1), *Geosci. Model Dev.*, 7, 1621–1627, <https://doi.org/10.5194/gmd-7-1621-2014>, 2014.
- Pan, L., Tong, D. Q., Lee, P., Kim, H., and Chai, T.: Assessment of NO<sub>x</sub> and O<sub>3</sub> forecasting performances in the US National Air Quality Forecasting Capability before and after the 2012 major emissions updates, *Atmos. Environ.*, 95, 610–619, <https://doi.org/10.1016/j.atmosenv.2014.06.020>, 2014.
- Parrish, D. F. and Derber, J. C.: The National Meteorological Center spectral statistical interpolation analysis system, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Purser, R. J., Wu, W.-S., Parrish, A. F., and Roberts, N. M.: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous Gaussian covariances, *Mon. Weather Rev.*, 131, 1524–1535, 2003a.
- Purser, R. J., Wu, W.-S., Parrish, A. F., and Roberts, N. M.: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances, *Mon. Weather Rev.*, 131, 1536–1549, 2003b.
- Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R. R., Ichoku, C., Levy, R. C., Kleidman, R. G., and Eck, T. F.: The MODIS aerosol algorithm, products, and validation, *J. Atmos. Sci.*, 62, 947–973, 2005.
- Roy, B., Mathur, R., Gilliland, A. B., and Howard, S. C.: A comparison of CMAQ-based aerosol properties with IMPROVE, MODIS, and AERONET data, *J. Geophys. Res.-Atmos.*, 112, D14301, <https://doi.org/10.1029/2006JD008085>, 2007.
- Sonntag, D. B., Baldauf, R. W., Yanca, C. A., and Fulper, C. R.: Particulate matter speciation profiles for light-duty gasoline vehicles in the United States, *J. Air Waste Manage.*, 64, 529–545, <https://doi.org/10.1080/10962247.2013.870096>, 2014.
- Tang, Y., Lee, P., Tsidulko, M., Huang, H.-C., McQueen, J. T., DiMego, G. J., Emmons, L. K., Pierce, R. B., Thompson, A. M., Lin, H.-M., Kang, D., Tong, D., Yu, S.-C., Mathur, R., Pleim, J. E., Otte, T. L., Pouliot, G., Young, J. O., Schere, K. L., Davidson, P. M., and Stajner, I.: The Impact of chemical lateral boundary conditions on CMAQ predictions of tropospheric ozone over the Continental United States, *Environ. Fluid Mech.*, 9, 43–58, <https://doi.org/10.1007/s10652-008-9092-5>, 2008.
- Tang, Y., Chai, T., Pan, L., Lee, P., Tong, D., Kim, H., and Chen, W.: Using optimal interpolation to assimilate surface measurements and satellite AOD for ozone and PM<sub>2.5</sub>: a case study for July 2011, *J. Air Waste Manage. Assoc.*, <https://doi.org/10.1080/10962247.2015.1062439>, 2015.
- Whitby, E. R. and McMurry, P. H.: Modal aerosol dynamics modeling, *Aerosol Sci. Tech.*, 27, 673–688, 1997.
- Wu, W.-S., Purser, R. J., and Parrish, D. F.: Three-dimensional variational analysis with spatially inhomogeneous covariances, *Mon. Weather Rev.*, 130, 2905–2916, 2002.