Air Quality Forecast Verification Using Satellite Data

S. Kondragunta,* P. Lee,† J. McQueen,# C. Kittaka,@,++ A. I. Prados,& P. Ciren,** I. Laszlo,* R. B. Pierce,* R. Hoff, & AND J. J. Szykman##

*NOAA/NESDIS/Center for Satellite Applications and Research, Camp Springs, Maryland
† Science Applications International Corporation, Camp Springs, Maryland
#NOAA/NWS/National Centers for Environmental Prediction, Camp Springs, Maryland
@ Science Applications International Corporation, Hampton, Virginia
& Joint Center for Earth Systems Technology, University of Maryland, Baltimore County, Baltimore, Maryland
**QSS, Inc., Camp Springs, Maryland
++NASA Langley Research Center, Norfolk, Virginia
##Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina

(Manuscript received 13 January 2006, in final form 30 March 2007)

ABSTRACT

NOAA’s operational geostationary satellite retrievals of aerosol optical depths (AODs) were used to verify National Weather Service developmental (research mode) particulate matter (PM2.5) predictions tested during the summer 2004 International Consortium for Atmospheric Research on Transport and Transformation/New England Air Quality Study (ICARTT/NEAQS) field campaign. The forecast period included long-range transport of smoke from fires burning in Canada and Alaska and a regional-scale sulfate event over the Gulf of Mexico and the eastern United States. Over the 30-day time period for which daytime hourly forecasts were compared with observations, the categorical (exceedance defined as AOD > 0.55) forecast accuracy was between 0% and 20%. Hourly normalized mean bias (forecasts – observations) ranged between −50% and +50% with forecasts being positively biased when observed AODs were small and negatively biased when observed AODs were high. Normalized mean errors are between 50% and 100% with the errors on the lower end during the 18–22 July 2004 time period when a regional-scale sulfate event occurred. Spatially, the errors are small over the regions where sulfate plumes were present. The correlation coefficient also showed similar features (spatially and temporally) with a peak value of −0.6 during the 18–22 July 2004 time period. The dominance of long-range transport of smoke into the United States during the summer of 2004, neglected in the model predictions, skewed the model forecast performance. Enhanced accuracy and reduced normalized mean errors during the time period when a sulfate event prevailed show that the forecast system has skill in predicting PM2.5 associated with urban/industrial pollution events.

1. Introduction

Congressional mandate (H. R. 4. Energy Act of 2002) required the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) to develop and deploy an air quality modeling system capable of issuing nationwide hourly ozone and PM2.5 (particles smaller than 2.5 μm in median diameter) forecast guidance. In response to the mandate, NOAA has developed an air quality forecast capability in partnership with the U.S. Environmental Protection Agency (EPA). The initial capability is based on linking the NWS operational forecast model, Eta, with the Community Multiscale Air Quality (CMAQ) model (Davidson et al. 2004). In September 2004, the NWS began issuing hourly ozone forecast guidance operationally and PM2.5 guidance experimentally in the northeastern United States using the CMAQ modeling system. The CMAQ model is run offline in a forecast mode using North American Mesoscale (NAM)–Eta Model winds, surface–air interface dynamics, precipitation, and cloud-related variables to drive the transport and chemistry. Using inputs of precursor emissions (provided by the EPA), a set of initial conditions, and parameterized gas phase and heterogeneous photo-
chemical processes, the CMAQ model computes concentrations of ozone (Otte et al. 2005). In a separate developmental test configuration, the forecast fields of PM2.5 concentrations are produced with an expanded set of chemical reactions, but based on chemical input only from emissions inventory data. Predicted concentrations data are then used in the model to calculate aerosol optical depth. Uncertainties in ozone and PM2.5 forecasts arise from uncertainties in precursor emissions, meteorology, physics, chemistry, and initial/boundary conditions in the model. A good forecast verification system is essential to diagnose forecast errors in space and time.

The CMAQ forecast verification system for ozone and PM2.5 consists of EPA’s AIRNOW (http://www.airnow.gov) measurements. While EPA’s ozone-monitoring network is dense, with coverage extending from rural to urban areas, the PM2.5 monitoring network is much sparser, and mostly located in the urban areas. This poses a problem for PM2.5 forecast verification over large spatial domains. Satellite observations have full spatial coverage and are best suited to verify air quality forecasts on different spatial and temporal scales. However, satellites do not directly measure surface PM2.5. Satellites measure column extinction due to particles (aerosol optical depth) and that depends on atmospheric aerosol loading, particle type, particle size, and relative humidity. In addition, retrieval of optical depths depends on surface characteristics (a dark surface gives better results) and cloud-free conditions.

Individual-layer aerosol extinction coefficients in the 22 layers of the CMAQ model are computed using PM2.5 concentrations. These layer extinctions are then integrated to calculate column optical depth. This allows satellite-measured aerosol optical depths (AODs) to be directly compared with CMAQ AODs. Satellite data allow the evaluation of the model with respect to the whole column, complementing typical evaluations where model predictions at the surface are compared with ground observations. Although the evaluation of surface predictions is critical and essential to assess forecast accuracy, evaluating processes aloft that impact the surface concentrations downwind is equally important in diagnosing what causes forecast error. Geostationary satellites are viable tools for such evaluations as they measure column amounts (near surface, low-level, and free troposphere) over wide spatial domain at high temporal frequency.

In this study, we explored the possibility of using AOD measurements from NOAA’s Geostationary Operational Environmental Satellite (GOES) to verify CMAQ AOD forecasts. During July and August of 2004, NOAA conducted a field campaign called the New England Air Quality Study (NEAQS). This field campaign used aircraft-, ship-, ground-observations, and models to understand the state of the air quality of the New England area. For this campaign, the NWS ran the CMAQ model in a research mode for PM2.5 forecasts. Thirty days (15 July–15 August 2004) of these forecasts are used in this study to compare with satellite observations, to help to evaluate test predictions, and to demonstrate the applicability of using satellite data in air quality forecast verification.

For quantitative applications such as verification of air quality forecasts, there is a need to use accurate retrievals. Therefore, prior to using the GOES AODs in the model verification study, we conducted extensive evaluation of GOES AOD data. This is necessary because the GOES AOD algorithm is not only a single-channel retrieval, but it also uses the same visible channel to retrieve both aerosol optical depth and surface reflectance. Current state-of-the-art aerosol retrievals over land are expected to have uncertainties within 0.15 \( \tau \pm 0.05 \), where \( \tau \) is the aerosol optical depth. Studies have shown that for multichannel AOD retrievals from nonoperational climate sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Multiangle Imaging Spectroradiometer (MISR), 71% and 66% of global aerosol retrievals in the visible channel fall within this uncertainty, respectively (Levy et al. 2005; Kahn et al. 2005). Although an operational sensor, NOAA’s Advanced Very High Resolution Radiometer (AVHRR) sensor has been providing aerosol retrievals over ocean for the last two decades with accuracies within expected uncertainty of 37% for various operational applications (Stowe et al. 1997; Ignatov et al. 2004). AVHRR assumes one aerosol model over the whole globe and uses a single-channel retrieval approach (Ignatov et al. 2004). Similarly, expected uncertainty for GOES AOD retrieval is between 18% and 34% (Knapp et al. 2002). Prados et al. (2007) analyzed two years of GOES-12 AOD data and showed that, for AODs greater than 0.1, retrieval uncertainties are less than 20%. In this study we analyzed GOES AOD retrievals for the July–August 2004 time period to show that they fall within this expected uncertainty and are therefore suitable for operational air quality forecast verification.

2. Summer 2004

The summer of 2004 was affected by the long-range transport of smoke from biomass burning in Canada and Alaska, in addition to fires burning inside the United States. Although there were several small fires burning in the United States between July and August
2004, the presence of a large-amplitude trough over the eastern United States for much of late July resulted in the transport of smoke from the fires burning in Alaska and Canada during this period. This was the worst fire season in 50+ years in Alaska. According to the National Interagency Fire Center (NIFC), Alaskan wildland fires burned over 3.4 million acres during late June and early July. Based on the GOES Imager observations of fire counts in Canada in the last five years, there were fewer fires in 2004 than in the fire seasons of 2002 and 2003. However, despite the small number of fires in 2004, the acreage burned was very high. According to Canadian National Forest Service report, 3.2 million Canadian acres burned in 2004, as compared with 2.8 and 1.6 million acres in 2002 and 2003, respectively.

Figure 1 shows the sequential GOES AOD images at a resolution of $4 \times 4 \text{ km}^2$ over the United States from 17 to 22 July 2004. The smoke-rich air began to move southward on 17 July with the outer edges of the smoke plume covering the northern part of the United States even while the center of the plume hovered over southern Canada. By 18 July, the smoke plume moved farther down to Kansas and Oklahoma. The smoke plume stretched from coastal Louisiana and Texas to Quebec, Canada, on 19 July. This smoke transport occurred in a region of cloud-free air just ahead of an extensive cloud shield to the north of the Great Lakes, and just behind the well-defined comma cloud associated with the low pressure system over the northeastern United States. On 20 July, the smoke plume continued to be elongated as it advected to the southeast, with remnants of the smoke plume evident over the Appalachian Mountains. Lidar observations from the Regional East Atmospheric Lidar Mesonet (REALM) network show that the smoke subsided as it moved from north (Madison, Wisconsin) to southeast (Baltimore, Maryland, and New York City, New York), from 6 km down into the boundary layer (http://alg.umbc.edu/REALM/index.html). During 20–22 July, the smoke-rich air was observed over Maryland, with elevated smoke layers at around 2–6 km. On 21 July, a fraction of the smoke mixed into the boundary layer (Engel-Cox et al. 2006).

Figure 2 shows the sequential 1300 UTC CMAQ AOD images at $12 \times 12 \text{ km}^2$ resolution over the United States for 17–22 July. Since the CMAQ forecast did not include influences from the Alaskan and Canadian wildfires, the CMAQ AOD during this period is dominated by a more local aerosol event and does not predict the southward transport and elongation of the smoke plume over the northern and south-central United States. Instead, on 17 July localized regions of elevated aerosol loading were predicted over Ohio with an extensive region of high aerosol loading extending from southern Mississippi northeastward toward South Carolina and off the northeastern seaboard. During the next five days of predictions, the Ohio aerosols increased and were advected eastward over Pennsylvania and Maryland within the northern portion of the developing low pressure system. The aerosol loading along the Gulf Coast also built during this period, but remained over the Gulf Coast states and southeastern United States as a weak cold front moved out over the Atlantic. From 17 to 19 July, much of the aerosol transport predicted by CMAQ occurs in association with the developing low pressure system and is therefore obscured by clouds in the GOES AOD imagery. However, from 20 to 22 July, after the Alaskan smoke plume has broken up, much of the southeastern United States is cloud free. During this time period the CMAQ AOD prediction does a reasonable job of capturing the spatial distribution, but underestimates the magnitude of the aerosol loading in the southeastern United States.

Figure 3 shows distributions of total gravitational mass, the ratio of total carbonaceous mass (TCM) to sulfate aerosol mass, TCM, and sulfate masses from the EPA surface speciation network on 20 July 2004 (EPA 1997). The distribution of EPA total gravitational mass shows elevated (>30 μg m$^{-3}$) surface aerosols in the southeastern United States and Michigan. The ratio of TCM to sulfate mass is high and consistent with smoke from the Alaskan and Canadian wildfires to the west of the Mississippi and Ohio River valleys while sulfates dominate in Virginia, the Carolinas, and northern Georgia. In the rest of the eastern United States, the TCM to sulfate ratio ranges from 0.75 to 1.25, indicating that smoke accounts for a significant fraction of the surface fine aerosol mass.

3. Data

a. GOES AOD measurements

The AOD measurements from NOAA’s GOES Imager are available in near–real time once every 30 min during the sunlit portion of the day. GOES AODs are derived using measurements made in the visible band (0.52–0.72 μm) using a lookup table (LUT) approach (Knapp 2002; Knapp et al. 2005). Note that the optical depths reported are for 550 nm. LUTs are created using the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) radiative transfer model (Vermote et al. 1997). Measurements in the visible band are at 1 km × 1 km resolution but the infrared channels used in the cloud-screening algorithm are measured at 4 km × 4 km resolution. Therefore, the GOES AODs
FIG. 1. Sequential images of GOES AOD 1300 UTC retrievals for (a) 17, (b) 18, (c) 19, (d) 20, (e) 21, and (f) 22 Jul 2004. Areas in white are clouds, and areas with black/gray color are the regions where retrievals are flagged by the data quality screening procedures.
FIG. 2. Sequential images of CMAQ AOD 1300 UTC forecasts for (a) 17, (b) 18, (c) 19, (d) 20, (e) 21, and (f) 22 Jul 2004.
are retrieved at 4 km × 4 km spatial resolution. For this study, these data were further regressed onto a 12 km × 12 km CMAQ Lambert conformal (LCC) projection to project the satellite data onto CMAQ’s spatial scale and projection. The Eta-CMAQ modeling system uses a Lambert conformal map projection of an Arakawa C staggered horizontal grid with a 12-km horizontal grid spacing (Otte et al. 2005). In the regridding process, the GOES AOD pixel is assigned to an Eta-CMAQ model cell if the center of the AOD pixel falls within the four corners of the model cell. After the AOD data are rebinned into the model cells, an average AOD value is calculated based on the number of rebinned satellite data within each model cell.

Satellite-measured visible radiances, from which AODs are retrieved, have contributions from ground (surface reflectance), aerosols, and background atmosphere (ozone absorption, water vapor absorption, Rayleigh scattering, aerosol scattering). The surface contribution is estimated by using a clear-sky composite of the second darkest pixel from the previous 21 days, and the state of the atmosphere is defined by climatological ozone and water vapor, and the *U.S. Standard Atmosphere, 1976* for Rayleigh scattering. An AOD of 0.02 is assumed for clear-sky background aerosol based on Knapp et al. (2005). A continental aerosol model with a single scattering albedo of 0.89 at 550 nm is assumed in creating the algorithm lookup tables (Hess et al. 1998). The Stowe et al. (1999) cloud-screening algorithm, developed for AVHRR, has been adapted and modified to screen for cloudy pixels. The threshold-based cloud-screening tests use measured reflectivities in the visible channel and brightness temperatures in the infrared channels to determine if pixels are contaminated by thick clouds, semitransparent clouds, and high-altitude thin clouds (Knapp et al. 2005). Despite

![Image](http://journals.ametsoc.org/doi/pdf/10.1175/2007JAMC1392.1)
these cloud-screening tests, certain pixels may remain unscreened because of the 4-km spatial resolution (e.g., popcorn cumulus clouds). To account for subpixel cloud contamination, the algorithm also contains various quality control steps that use spatial variability tests to screen for outliers. As a part of the validation efforts, we modified some of the thresholds in these screening tests. These and other quality control procedures are described in section 3a(2).

In addition to cloud contamination, GOES AOD data have other sources of uncertainties. These include aerosol model assumptions, errors in estimation of surface reflectivities, variations in atmospheric ozone and water vapor concentrations, calibration errors, and errors associated with radiative transfer modeling. GOES AOD retrieval errors associated with calibration errors were minimized by using updated calibration. As of 1 October 2006, the GOES-12 visible channel degraded by 22.5% since it became operational on 1 April 2003. The GOES-12 visible channel is vicariously calibrated using MODIS to account for calibration drift (Wu and Sun 2005). While careful screening and reprocessing can minimize retrieval errors due to cloud contamination and calibration errors, errors due to aerosol model assumptions and errors in surface reflectance retrieval are hard to reduce. Since the same visible channel is used for retrieving surface reflectance and aerosol optical depth, errors in surface reflectance will translate into retrieved optical depth. These errors can occur over bright surfaces and when cloud shadows result in underestimation of surface reflectance. For aerosol models, it is not possible to dynamically choose an aerosol model that varies in space and time. Current operational sensors such as the GOES Imager with only five channels (one visible, one near-IR, and three IR channels) do not provide enough information to pick an aerosol model based on observations. The operational GOES AOD algorithm uses a continental aerosol model as it represents the most commonly observed aerosol types over the United States (water soluble nitrate and sulfate, soot, and dust aerosols). This has been shown to work well for GOES AOD retrievals over the United States (Knapp 2002; Knapp et al. 2005; Prados et al. 2007). Additionally, based on Stowe et al. (1997), it must be noted that assumptions of aerosol phase function is more important than aerosol model. Changes in aerosol single scattering albedo associated with aerosol model affects the amount of backscattered radiation (phase function), and therefore one can tune aerosol phase function by keeping the same aerosol model to get an agreement between satellite retrievals and ground observations (Stowe et al. 1997). However, in absence of the knowledge of the actual phase function of aerosol mixture in the atmosphere, it is important to make a reasonable assumption of aerosol model.

Although the continental aerosol model assumption works well, it is generic and can be very different from actual aerosol types present in the atmosphere during events such as biomass burning (Fig. 3). Speciated observations made on 20 July 2004, shown in Fig. 3, clearly indicate that sulfate dominated in the southeast and eastern United States in general but carbonaceous aerosols dominated in the central United States. Therefore, we tested the sensitivity of retrievals to aerosol model assumptions by reprocessing GOES AOD data with two different types of biomass-burning aerosol models, one used by MODIS and one used by MISR (Remer et al. 2006; Kahn et al. 2001). The Kahn et al. (2001) model, which is a mixture of carbonaceous (54%), sulfate (25%), dust (12%), and black carbon (9%) with a single scattering albedo of 0.78 at 550 nm has recently been updated, and their new model is less absorbing with a single-scattering albedo between 0.89 and 0.91 (Reid et al. 2005). We used the Remer et al. (2006) biomass-burning aerosol model, which is similar to Reid et al. (2005). Results from these sensitivity runs are described in section 3a(2).

1) GOES AOD DATA QUALITY

The GOES AOD data are screened for outliers using various threshold-based tests. These thresholds were determined based on an extensive analysis done by Knapp (2002). GOES AOD data for retrieved surface reflectivities less than 0.5% (to avoid cloud shadows) and greater than 15% (to avoid bright surfaces) are screened. To eliminate residual cloud contamination, a screen using a threshold value of 0.15 for AOD standard deviation corresponding to average AOD in a 3 × 3 box (8 pixels) surrounding the target pixel is used. If AOD standard deviation is greater than 0.15, the data are flagged. These spatial variability tests are very common to satellite retrievals (Zhang and Reid 2006; Martins et al. 2002). In addition to this, the screening procedures require that, for each target pixel, surrounding pixels in a 5 × 5 box are cloud-free from cloud-screening tests. These data-screening thresholds were carefully chosen by investigating the bias between GOES AODs and ground observations as a function of these parameters (surface reflectance and AOD standard deviation). Although these screening tests flag bad retrievals and eliminate obvious outliers, the tests are very conservative and flag some good data as well; for the July–August 2004 time period, 4% of good retrievals were flagged as bad retrievals. This analysis is
simple and can be repeated by GOES AOD data users to refine the thresholds based on their individual needs.

2) Evaluation of GOES AOD retrievals

GOES AOD retrievals were compared with Aerosol Robotic Network (AERONET) level-1.5 AOD measurements from 16 stations in the central and eastern United States for July and August 2004 (http://aeronet.gsfc.nasa.gov). This domain is consistent with the CMAQ modeling domain. GOES versus AERONET matchups were performed within ±15 min of coincident observations. For each GOES observation time period, GOES AOD overpass data corresponding to each AERONET station were created by averaging AOD from surrounding pixels in a 3 × 3 box. For AERONET data, observations coincident within ±15 min of GOES observation time were averaged. The three panels in Fig. 4 show scatterplots of GOES AODs versus AERONET AODs for the three different aerosol models used in processing GOES AOD data. GOES AODs agree well with AERONET observations for continental aerosol model with an RMS (bias and noise) value of 0.112 (22% for mean AOD of 0.5) and a slope of 0.84. These results are consistent with Prados et al. (2007) who validated two years of GOES AOD retrievals by comparing them with MODIS and AERONET AODs. Prados et al. (2007) show that for stations in the eastern United States, the correlation coefficient is 0.79 with a slope value of 0.78 and RMS error of 0.12. For stations in the central and western United States, the correlation coefficient is much lower (<0.4). Prados et al. (2007) also report that for AODs greater than 0.1, the retrievals are within 20% of AERONET AODs while errors are greater than 30% for AODs lower than 0.1. For GOES AOD retrievals using MODIS biomass-burning aerosol model, the RMS and slope values were 0.150 and 0.733, respectively, when compared with AERONET. In MODIS biomass-burning aerosol model, aerosol optical and physical properties are a function of absolute AOD (Remer et al. 2006). The larger RMS value when the MODIS biomass-burning aerosol model is used indicates that the model is not compatible with observations. Similarly for GOES AOD retrievals using the MISR biomass-burning aerosol model of Kahn et al. (2001), the RMS and slope values were 0.0907 and 0.645, respectively, when compared with AERONET. Although the RMS values are lower, the slope values deviate more from unity. The use of the MISR smoke aerosol model resulted in lowering the magnitude of GOES AOD values for situations in which aerosol loading was high and perhaps consisting of mostly

![Fig. 4. Scatterplot of GOES AODs vs AERONET AODs for (a) continental aerosol model, (b) MODIS biomass-burning aerosol model, and (c) MISR biomass-burning aerosol model.](http://journals.ametsoc.org/doi/pdf/10.1175/2007JAMC1392.1)
smoke aerosol. As a result, the scatter in the data (outliers) decreased and improved RMS values. However, using biomass aerosol model distorts the agreement when aerosols are a mixture of various aerosol types. Although smoke from forest fires in Canada and Alaska dominated the Midwest region, smoke plume dissipated into several small plumes over the Appalachian Mountains. The use of a biomass-burning aerosol model is perhaps best suited for the central United States but not for the eastern United States where, in addition to remnants of dissipated smoke plumes, there was a regional-scale sulfate pollution in the boundary layer below 3 km (Engel-Cox et al. 2006). Figure 5 shows the 2-month mean bias between GOES AODs and AERONET AODs for SERC station, located in Maryland. The bias is slightly variable with time of the day but overall flat (near zero) for retrievals using continental aerosol model (bottom panel in Fig. 5). For GOES AOD retrievals using biomass-burning aerosol models, the biases are larger than those obtained for the continental aerosol model (top panel for MODIS aerosol model and middle panel for MISR aerosol model in Fig. 5). Results are similar for other 15 stations in the central and eastern United States. It is difficult to assess the true nature of the aerosol mixture in the atmosphere. Most satellite aerosol retrieval algorithms choose an aerosol model a priori using different approaches. For example, the MODIS algorithm chooses its models depending on the geographical region. An ideal situation would be for satellite radiances to provide the information on aerosol type so appropriate model can be dynamically chosen. Although it is desirable to have that capability, having an aerosol model that accounts for a wide variety of aerosols can prevent the algorithm from being tuned for episodic cases such as biomass burning or dust storms. We recognize that the choice of aerosol model will continue to be a challenge for retrieving aerosol optical depths. Using agreement of GOES AODs with AERONET AODs as guidance, we find that the continental aerosol model is preferred over the biomass-burning aerosol model and thus used GOES AODs retrieved using LUTs created from the continental aerosol model in the CMAQ evaluation work.

b. CMAQ runs

The CMAQ modeling system uses National Centers for Environmental Prediction (NCEP) NAM–Eta Model (12 × 12 km² spatial resolution and 60 vertical layers up to 0.25 hPa) gridded meteorological model predictions at hourly intervals for chemical transport, air–surface interactions, and emissions processing (Lee et al. 2005). Interface processors that handle the vertical and horizontal interpolation of the meteorological fields collapse the 60 vertical layers into 22 terrain-following sigma levels in the CMAQ model (Lee et al. 2005). The CMAQ chemistry configuration is optimized (e.g., with reduced chemical species) to provide predictions in real time (Byun and Ching 1999). The boundary conditions for ozone come from NCEP’s Global Forecast System and climatological profiles are used for PM2.5. For initial conditions, a 6-h cycling system was developed and run 4 times per day to initialize CMAQ gas-phase-only chemistry and soil fields to reduce spinup of soil and chemical constituents. However, for the developmental aerosol tests, a 24-h cycling system was utilized. CMAQ hourly outputs of AODs from 15 July to 15 August 2004 were obtained from NWS/NCEP for this study.

The CMAQ aerosol module, developed by Binkow-
ski and Shankar (1995), has three aerosol modes (Aitken mode, accumulation mode, and coarse mode). Species included in these modes are sulfate, ammonium, nitrate, organic (primary and secondary), elemental carbon, and water. Interaction between these modes is one-way toward coarse mode when fine-mode particles grow beyond 2.5 μm in diameter. There is no coagulation between fine and coarse mode. Aerosol extinction coefficient at 520 nm in the CMAQ model β is calculated according to Eq. (1):

$$\beta = \frac{3\pi}{\lambda} \int_{-\infty}^{\infty} \frac{Q}{\alpha} \frac{dV}{\ln\alpha} d\ln\alpha,$$

where \(\alpha\) is particle size parameter, \(V\) is volume of the particle, \(\lambda\) is wavelength, and \(Q\) is extinction efficiency calculated using approximations to Mie theory (Binkowski 1999). AODs are calculated by integrating \(\beta\) values from each layer in the CMAQ model.

The dominant aerosol types in the CMAQ model are ammonium sulfate, ammonium nitrate, organic, and elemental carbon. Extinction at 1800 UTC 21 July 2004 for example was primarily due to sulfate. Peak surface sulfate concentration in the model domain was about 37 µg m\(^{-3}\) whereas nitrate, organic carbon, and elemental carbon were 22, 1.8, and 0.123 µg m\(^{-3}\), respectively. In contrast, on 23 July 2004 after the episode ended, peak concentrations of sulfate, nitrate, organic carbon, and elemental carbon were 5, 4, 4.5, and 0.01 µg m\(^{-3}\), respectively. Evaluation of CMAQ model aerosol components has proven to be particularly difficult because of lack of observations. Not only spatially (horizontally and vertically) and temporally sparse, most observations are limited to total particulate mass, without chemical information. Mebust et al. (2003) used visibility data and speciated data from the Interagency Monitoring of Protected Visual Environments (IMPROVE; http://vista.cira.colostate.edu/improve) network to evaluate the aerosol component for simulations of the 11–13 July 1995 episode and found that CMAQ predictions underestimated ground-level PM2.5 concentrations by 38.5% and PM10 concentrations by 51%. They found that the CMAQ model simulates the sulfate concentrations better than other species. CMAQ PM2.5 evaluation for the study period in this paper shows that model predictions agree with surface PM2.5 observations to within 25% (Mathur et al. 2005). Mathur et al. (2005) also found that sulfate predictions agree well with surface observations, but organic carbon was underestimated by 70% because of contributions of long-range transport of smoke that the model did not include. However, when satellite-derived MODIS AODs were assimilated into the CMAQ model to account for missing biomass-burning aerosols, the organic aerosol concentrations improved by 23% to 42% and decreased the discrepancy between model predictions and observations (Rao et al. 2006).

Sources of uncertainties in CMAQ predictions include uncertainties in anthropogenic emissions, transport, lack of emissions from biomass burning (prescribed and wild), static boundary conditions, etc. While retrospective model runs can be tuned and improved to perform better, running the model in forecast mode and evaluating the forecasts in near–real time requires tools such as satellites to provide the data in near–real time. Sources of emissions from random events such as forest fires, dust storms, and NO emission pulses from soil after a precipitation event or a lightning event cannot be incorporated into real-time forecasts unless that information is made available in near–real time. Until the forecast system is ready for ingesting information on such episodes in real time, forecast accuracy will be limited during extreme events such as forest fires and dust storms.

4. Statistical analysis

CMAQ forecasts of hourly AOD were evaluated against GOES observations with standard statistical techniques to determine correlation coefficient \(r\), mean bias (MB), normalized mean bias (NMB), root-mean-square error (RMSE), and normalized mean error (NME):

$$MB = \frac{1}{N} \sum_{i=1}^{N} (\text{CMAQ} - \text{GOES}),$$

$$NMB = \frac{\sum_{i=1}^{N} \frac{|\text{CMAQ} - \text{GOES}|}{\text{GOES}}}{\sum_{i=1}^{N} \text{GOES}} \times 100\%,$$

$$NME = \frac{\sum_{i=1}^{N} |\text{CMAQ} - \text{GOES}|}{\sum_{i=1}^{N} \text{GOES}} \times 100\%,$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\text{CMAQ} - \text{GOES})^2}{N}}^{1/2}.$$
main were used in the statistical analysis; despite missing data fields due to the presence of clouds, hundreds of grid cells with observed AOD values were analyzed. Forecast accuracy for predicting an exceedance (AOD > 0.55) or a nonexceedance (AOD < 0.55) was computed according to Kang et al. (2007):

\[
\text{Accuracy (Exceedance)} = \frac{b}{(a + b + d)} \times 100\% \quad \text{and} \quad \text{Accuracy (Nonexceedance)} = \frac{c}{(a + c + d)} \times 100\%.
\]

For the correlation between observed (GOES) and predicted (CMAQ), variable \(a\) defines the number of points where both GOES observations and CMAQ predictions are greater than 0.55 but GOES observations are lower, variable \(b\) defines the number of points where both GOES observations and CMAQ predictions are greater than 0.55 and CMAQ predictions are greater than 0.55 but GOES observations are lower, variable \(c\) defines the number of points where both GOES observations and CMAQ predictions are lower than 0.55, and variable \(d\) defines the number of points where GOES observations are higher than 0.55 but CMAQ predictions are lower. The AOD value of 0.55 was chosen as it generally corresponds to surface PM2.5 concentrations of 35 \(\mu g\) m\(^{-3}\), a newly revised EPA standard for 24-h average (http://www.epa.gov/air/criteria.html). This proportionality between GOES AOD and surface PM2.5 was determined by averaging the slope values for correlation between GOES AOD and surface PM2.5 measurements for those stations that showed correlation coefficient greater than 0.6. Of the 256 stations analyzed, 95 stations showed correlation greater than 0.6. Good correlations are obtained when the dynamic range extends from low (near zero) to high (\(\sim 70 \mu g\) m\(^{-3}\)). When the data are segregated into exceedances (>35 \(\mu g\) m\(^{-3}\) and nonexceedances (<35 \(\mu g\) m\(^{-3}\)), the correlation between satellite-measured AOD and surface PM2.5 concentration drops. For example, for the station that has highest correlation coefficient (0.92) between GOES AOD and surface PM2.5 when all data are included, the correlation coefficient drops to 0.87 when only data with PM2.5 concentrations greater than 35 \(\mu g\) m\(^{-3}\) are included. The correlation and slope also varied from station to station. This variability is attributed to variations in boundary layer depth, relative humidity, and aerosol vertical profile (Engel-Cox et al. 2006; Gupta et al. 2006). Similar slope values were reported for correlation between MODIS AOD and surface PM2.5 by Al-Saadi et al. (2005) and Wang and Christopher (2003).

5. Results and discussion

a. Analysis of CMAQ performance prior to the impact of long-range transport of smoke

The air quality forecast verification period that we chose was dominated by the long-range transport of smoke into the United States (CMAQ model domain) after 15 July 2004. Accordingly, the forecasts for 1700 UTC 21 July and 15 July 2004 are used to understand the model performance with and without the influence of long-range smoke transport of smoke. The four panels in Figs. 6a–d show comparisons of GOES and CMAQ AODs at 1700 UTC 15 July 2004. Observed AODs over most of the domain were below 0.5 with a few scattered regions of higher AOD values. Most of these high AODs that appear are pockets of clear scenes in the midst of cloudy scenes (e.g., Nebraska). These high AOD retrievals are near the cloud edges. Satellite retrievals are known to show high AOD values around the rim of the clouds (Ignatov et al. 2004). Aerosols are in a high relative humidity environment near the clouds and they can swell and scatter more light leading to high optical depths. Although not independently validated for these specific scenarios (high AODs near clouds), these data are retained in our analysis as the data were processed through our cloud-screening algorithm and spatial variability tests to look for subpixel cloud contamination. Although CMAQ forecasts for 1700 UTC 15 July show elevated AODs associated with sulfate plumes over the Atlantic, high AODs observed over Texas in GOES data are not present in CMAQ simulations. The GOES Imager data do not show the elevated AODs over the Atlantic Ocean due to the presence of clouds except for a few small regions. Differences between GOES and CMAQ plotted in Fig. 6c are within ±0.1 except for certain small areas in western Texas and central Iowa, where differences are as high as 0.3. It should be noted though that except over western Texas, these large differences are associated with isolated high AODs observed in GOES near-cloudy regions; overall CMAQ is biased slightly low. The histogram plot (Fig. 6d) shows GOES and CMAQ data distribution for values greater than 0.55 (>35 \(\mu g\) m\(^{-3}\)). GOES AODs are spread between 0.55 and 1.4 whereas CMAQ has only few data points with AODs greater than 0.55. In contrast to this, on 21 July 2004 the smoke plume has already impacted the United States (Fig. 6e). The elevated AODs due to long-range transport of smoke and the sulfate event, well simulated by CMAQ, are observed over the mid-Atlantic, eastern seaboard, Gulf of Mexico, and Texas (Figs. 6e,f). Except for the high values over Texas as-
associated with transported smoke aerosol, CMAQ predictions of high AODs over western New York, the Atlantic Ocean, and Gulf of Mexico agree with GOES AOD observations. However, most of the central and southeastern United States was under cloud cover and did not provide observations to compare with CMAQ predictions. Differences between GOES and CMAQ AOD values are large over Texas because smoke aerosols were not included in CMAQ (Fig. 6g). Although CMAQ predicted high AODs associated with sulfate episodes in the eastern United States, its values were slightly biased low in those areas because a fraction of observed high AODs in the eastern United States is due to transported smoke aerosol.

GOES values are higher than the CMAQ predictions in regions where observed AODs are high, and GOES AODs are lower than CMAQ values in regions where observed AODs were small. On 21 July, in addition to the long-range transport of the smoke aloft, there was a regional-scale sulfate pollution event in the lower atmosphere. The sulfate event was well captured by the CMAQ model and is consistent with the appearance of a peak between 0.55 and 0.7 in the histogram plot (Fig. 6h). This peak was not present on 15 July 2004 (Fig. 6d).
Despite the presence of high AOD pixels in CMAQ predictions on 21 July 2004, the number of high AOD pixels in GOES is substantially larger than in CMAQ. It is likely that there are uncertainties in the CMAQ emissions database for local non-biomass-burning sources as well that need further investigation.

b. Spatial and temporal patterns

Figures 7a–e show the correlation coefficient, mean bias, normalized mean bias, root-mean-square error, and normalized mean error derived using the data at each grid cell over the 30-day time period. The correlation coefficient ranged between -0.2 and 0.8 over land with some regions over the ocean showing strong negative correlation (Fig. 7a). There is no specific geographical pattern in the correlation coefficient except that the correlation is high over Gulf of Mexico and the eastern seaboard, which corresponds to the well-forecast sulfate event by the CMAQ model. Features seen in the spatial pattern of correlation coefficient values agree well with those seen in normalized mean error (Fig. 7e). Normalized mean errors are very high in
Fig. 7: Spatial plots of (a) normalized mean bias, (b) correlation coefficient, (c) root-mean-square error, (d) mean bias, and (e) normalized mean error between GOES and CMAQ AODs for data between July and August 2004.
most of the model domain (60%–80%) except in the region where correlation between GOES and CMAQ is high ($r = -0.8$). In those regions (Gulf of Mexico and the eastern seaboard), normalized mean errors are between 40% and 60%. These regions correspond to the regional-scale sulfate event that resulted in a large-scale haze up and down the East Coast in addition to remnants of the transported smoke. Relatively smaller errors associated with the sulfate event indicate that the model forecasts have some skill in predicting urban/industrial pollution events. The CMAQ model has a negative bias over land ($\sim -0.3$) and positive bias ($\sim 0.2$) over the ocean; spatially, however, these regions of high/low bias do not correlate with errors. The region of high negative bias in the western part of the domain is also the region where ammonia emissions are generally underestimated in the CMAQ model (Binkowski and Roselle 2003). RMS differences varied from 0 to 0.8 with a mean of $\sim 0.4$ with smaller values over ocean and larger values over the land. This corresponds to smaller AOD values over ocean relative to land.

Time series of correlation coefficient values show that correlation peaked between 20 and 25 July (Fig. 8a). During this period, high sulfate concentrations were predicted and corroborated with satellite observations. When smoke dominated the atmosphere or when background aerosol conditions prevailed, the CMAQ predictions did not correlate well with observations. This is evident in the time evolution of mean bias, which changes sign from negative to positive between July and August (Fig. 8b). For 15–17 July, when the smoke did not reach the United States, the mean bias is near zero or slightly positive. The bias decreases during the rest of July because of high observed AODs associated with smoke and sulfate. Once the smoke cleared, marked by lower aerosol concentrations in the atmosphere during August 2004, CMAQ overpredicted the aerosol concentrations. This can also be seen in RMS errors, which decrease from higher values in July to smaller values in August (Fig. 8e).

Normalized mean errors over the 30-day time period ranged from 40% and 100% with relatively lower values during the 17 to 21 July 2004 period when the sulfate event coincided with the transported smoke. Relative errors are larger in August than in July, reflecting the smaller magnitude of observed AODs.

The forecast accuracy for an exceedance (AOD > 0.55) calculated [Eq. (6)] shows that the range of accuracy is between 0% and 20% (Fig. 9), whereas for a nonexceedance (AOD < 0.55) the accuracy is between 50% and 99% (Fig. 10). Values of forecast accuracy for an exceedance are lower than those typically observed for forecast variables. For example, evaluation of CMAQ ozone forecasts for this study period showed the accuracy values to be $\sim 24\%$ (Lee et al. 2008). While these individual grid-based statistics are expected to indicate the skill of the forecast model, they are computed by direct pairing of observed and model-predicted values without accounting for differences in sampling techniques. Kang et al. (2007) noted this and recommend various weighting techniques that include pairing of aggregated grid cells (e.g., size of a city) and observations. Our future studies will include these weighting techniques to minimize errors associated with sampling. A lack of biomass-burning aerosol sources in the model and exclusion of PM10 from AOD calculations are the reasons for the reduced accuracy in CMAQ aerosol forecasting when compared with ozone forecasting. The forecast accuracy for a nonexceedance is relatively higher because the number of grids where both CMAQ and GOES values are lower than 0.55 are disproportionately higher relative to number of grids for exceedances. There is, however, a wide spread (diurnal variation) to the forecast accuracy values in the nonexceedances case (40%–90%) between 15 July (day 197) and 28 July (day 210) (Fig. 10). Note that for both exceedances and nonexceedances, statistical analysis was carried out with data for cloud-free conditions because of a lack of GOES observations in the presence of clouds. The regional-scale sulfate event to the east of the Ohio region, well simulated by the CMAQ model, coincided with a low-pressure cloud system that obscured GOES from detecting the pollution below the clouds. Therefore, the forecast accuracy analysis is biased toward the satellite data and events corresponding to long-range transport of smoke, obscuring the strengths of the CMAQ model. Our future studies will focus on a regional-scale sulfate event associated with a multiday high pressure system, providing several continuous clear-sky days for assessing CMAQ PM2.5 forecast skills.

6. Conclusions

The forecast period from 15 July to 15 August 2004, for which comparisons of hourly forecasts and observations were made in this study, was dominated by long-range transport of smoke from fires burning in Canada and Alaska and a regional-scale sulfate event over the eastern United States. Over this 30-day period, the categorical forecast accuracy was between 0% and 20% for exceedances, and for nonexceedances, between 40% and 99%. For discrete forecasts, the mean bias (forecasts – observations) ranged between $-0.2$ and $+0.1$. Normalized mean bias ranged between $-50\%$
and +50% with forecasts being positively biased when observed AODs were small and negatively biased when observed AODs were high. Normalized mean errors are between 50% and 100% with the errors on the lower end during 18–22 July, in areas of the domain where a regional-scale sulfate event occurred. Spatially, the errors were low over the regions where higher sulfate concentrations were observed. Correlation coefficient values also showed similar features (spatially and temporally) with a peak value of ~0.6 during 18–22 July 2004. The inability of the GOES Imager to detect aerosols over cloudy regions skewed the statistics toward the period when long-range transport of the smoke event dominated. However, enhanced accuracy and reduced normalized mean errors during the time period when a sulfate event prevailed shows that the forecast system is very capable of issuing PM2.5 forecasts associated with urban/industrial pollution events.

**FIG. 8.** Temporal plots for comparison of coincident observations (GOES) and forecasts (CMAQ) for the 30-day time period beginning 15 Jul 2004.
To improve forecast accuracy, predictions must include the effects of large fires and dust events.

It should be noted that the use of satellite data to evaluate the CMAQ model complements efforts to evaluate the model-predicted PM2.5 concentrations using ground observations. The unprecedented advantage of satellite observations is the availability of data in near–real time across wide spatial scales and at high temporal frequency. Satellite data are column amounts and thus can provide evaluation of the model from a column perspective.

This study has used only the AOD data from GOES. However, other satellite datasets exist and can be used to verify or improve forecasts of ozone, PM2.5, and their precursors. The Ozone Monitoring Instrument (OMI) on the National Aeronautics and Space Administration Aura platform is already providing tropospheric column amounts of nitrogen dioxide (NO2) and other ozone precursors such as formaldehyde (HCHO) that can be used to evaluate model predictions. These species can also be assimilated into the model to constrain precursor emissions. Additional satellite data will come from the recently launched Global Monitoring Ozone Experiment (GOME-2) on the Meteorological Operation (MetOp) platform by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). NOAA is planning to provide trace gas and aerosol products from GOME-2 to users in near–real time for various air quality applications. These observational tools are expected to enhance our understanding of the atmospheric physics and chemistry and help to incorporate the new science into forecast models.

Acknowledgments. This work was funded in part by NOAA GIMPAP and G-PSDI programs. The authors acknowledge Rohit Mathur of NOAA/OAR and Paula Davidson of NOAA/NWS for releasing the experimental CMAQ forecasts used in this study. Although internally reviewed, the views expressed in this paper are those of the authors and must not be interpreted as those of the National Oceanic and Atmospheric Administration or U.S. government.

REFERENCES


Engel-Cox, J. A., and Coauthors, 2006: Integrating lidar and sat-
ellite optical depth with ambient monitoring for 3-dimensional particulate characterization. *Atmos. Environ.*, 40, 8056–8067.


