Description and Verification of the NOAA Smoke Forecasting System: The 2007 Fire Season

GLENN D. ROLPH,* ROLAND R. DRAXLER,* ARIEL F. STEIN,+,++ ALBION TAYLOR,+,++ MARK G. RUMINSKI,# SHOBHA KONDRAKUNTA,# JIAN ZENG,+,## HO-CHUN HUANG,@ GEOFFREY MANIKIN,& JEFFERY T. MCQUEEN, & AND PAULA M. DAVIDSON**

*NOAA/Air Resources Laboratory, Silver Spring, Maryland
†Earth Resources and Technology, Annapolis, Maryland
#NOAA/NESDIS/STAR, Camp Springs, Maryland
@Scientific Applications International Corporation, Camp Springs, Maryland
&NOAA/NWS/National Centers for Environmental Prediction, Camp Springs, Maryland
**NOAA/NWS/Office of Science and Technology, Silver Spring, Maryland

(Manuscript received 23 May 2008, in final form 20 October 2008)

ABSTRACT

An overview of the National Oceanic and Atmospheric Administration’s (NOAA) current operational Smoke Forecasting System (SFS) is presented. This system is intended as guidance to air quality forecasters and the public for fine particulate matter (≤2.5 μm) emitted from large wildfires and agricultural burning, which can elevate particulate concentrations to unhealthful levels. The SFS uses National Environmental Satellite, Data, and Information Service (NESDIS) Hazard Mapping System (HMS), which is based on satellite imagery, to establish the locations and extents of the fires. The particulate matter emission rate is computed using the emission processing portion of the U.S. Forest Service’s BlueSky Framework, which includes a fuel-type database, as well as consumption and emissions models. The Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model is used to calculate the transport, dispersion, and deposition of the emitted particulate matter. The model evaluation is carried out by comparing predicted smoke levels with actual smoke detected from satellites by the HMS and the Geostationary Operational Environmental Satellite (GOES) Aerosol/Smoke Product. This overlap is expressed as the figure of merit in space (FMS), the intersection over the union of the observed and calculated smoke plumes. Results are presented for the 2007 fire season (September 2006–November 2007). While the highest FMS scores for individual events approach 60%, average values for the 1 and 5 μg m⁻³ contours for the analysis period were 8.3% and 11.6%, respectively. FMS scores for the forecast period were lower by about 25% due, in part, to the inability to forecast new fires. The HMS plumes tend to be smaller than the corresponding predictions during the winter months, suggesting that excessive emissions predicted for the smaller fires resulted in an overprediction in the smoke area.

1. Introduction

Beginning in 2003, in response to congressional direction to the National Oceanic and Atmospheric Admin-
expand the capability to include quantitative predictions of airborne fine-particle matter (diameter \( \leq 2.5 \mu m \) – PM\(_{2.5}\)), with initial operational deployment targeted in 2014. Exposure to high levels of particulate matter from large fires can impact sensitive populations including children, people with respiratory illnesses, and the elderly (e.g., Kaiser 2005). The PM prediction process is very complex, involving contributions from direct sources (primary PM) and from secondary production by reaction of precursor species in the atmosphere (secondary PM). Current state-of-the-science simulations are not yet able to predict PM with the accuracy and reliability needed for NOAA’s operational predictions. In the interim, while the efforts to build a quantitative PM prediction capability are progressing, NOAA has developed a smoke forecast capability that takes account of a significant source of PM\(_{2.5}\) in the United States each year. Recent episodes of large wildfires (e.g., October–November 2007 in California) have emphasized this need: although PM\(_{2.5}\) from wildfires is only one component of the total particulate matter measured at ground level, it contributes a significant fraction during severe smoke episodes.

NOAA’s Office of Oceanic and Atmospheric Research (OAR), the NWS National Centers for Environmental Prediction (NCEP), and the EPA working together research air quality forecasting issues for transition into NWS operations. The former OAR Air Resources Laboratory (ARL) Atmospheric Sciences and Modeling Division (ASMD), now the U.S. EPA Atmospheric Modeling Division, along with NCEP have developed air quality simulations that have been adapted, tested, and implemented into the NAQFC in NWS operations for ozone predictions (Otte et al. 2005; Eder et al. 2006; Lee et al. 2008). Forecasting the movement of smoke from large wildfires has also been an ongoing research activity within ARL Headquarters since May 1998 when tremendous smoke from Central America (Fig. 1) was transported over Texas, the Southeast, and the mid-Atlantic states, and into Canada (Tanner et al. 2001; Peppler et al. 2000). This occurred as the burn season was at its height in Central America and Mexico and an amplified weather pattern enabled the transport of the smoke to extend far to the north and east. The smoke created a health hazard across large areas of Texas and reduced visibility at the surface and aloft throughout the Gulf coast region.

Consequently, because of additional fires in Florida in the summer of 1998, the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess 1997, 1998) was configured to produce a 48-h forecast of 3-h average concentrations of PM\(_{2.5}\) from wildfires over Florida and Central America. Calculations were made on a 10-km horizontal resolution concentration grid over Florida using the NCEP regional (40 km) Eta Model and on a 25-km concentration grid covering Mexico using NCEP’s Global Forecast System [GFS, formerly called the Aviation (AVN) run] with a horizontal resolution of 111 km.

**FIG. 1.** GOES-East visible image from 14 May 1998 showing the extensive smoke area extending from the Gulf of Mexico into the western Gulf states and points north.
Source locations were provided by NOAA’s National Environmental Satellite and Data Information Service (NESDIS), which analyzed polar-orbiting (Advanced Very High Resolution Radiometer, AVHRR) and geostationary [Geostationary Operational Environmental Satellite (GOES)] satellite infrared imagery channels to determine “hotspot” locations. A hotspot is an anomalously warm pixel or pixels relative to the surrounding environment that is observed in thermal infrared imagery. The U.S. Forest Service (USFS) provided daily information on the number of acres burned for a particular location. From this information, a gross estimate of particle emissions of total suspended particulate (TSP) were calculated based upon an average rate of 68 kg h\(^{-1}\) ha\(^{-1}\) from estimated values reported by Levine (1994) and Crutzen and Andrea (1990) for the tropics. Obviously, the uncertainty in the emission rates were quite large and were calibrated by comparing HYSPLIT predictions to observed satellite quantities. The results were posted daily to the ARL Web server for use by state and local environmental officials.

In 1998, fire and smoke analysis via satellite imagery was handled by NESDIS in a rudimentary fashion, with an analysis performed only over the region of interest and as conditions warranted. It quickly became apparent that a more routine and robust system was needed to identify and monitor fire locations and smoke movements across North America. In response, an operational fire and smoke program was initiated at NESDIS primarily to support the NWS’s needs in the regions affected by the smoke. From this program, the Hazard Mapping System (HMS) was created in 2001 (Ruminski et al. 2006). HMS is an interactive tool used to identify fires and smoke produced over North America in an operational environment using multiple environmental satellites from both NOAA and the National Aeronautics and Space Administration (NASA). ARL began automating the smoke forecasts using the HMS fire locations, which served as a basis for the current smoke forecasting system that became operational at the NWS in March 2007.

This paper presents an overview of the current operational Smoke Forecasting System (SFS), a description of the HMS system, the smoke emissions, the HYSPLIT model configuration, the meteorological data, the model verification system, and results for the 2007 fire season.

2. Operational system

a. Overview

Building upon a relatively simple system already producing smoke forecasts from fires, NOAA began producing a daily operational smoke particulate (PM\(_{2.5}\)) forecast product in March 2007. This system, built and tested by a team of NOAA scientists (OAR, NESDIS, NWS) together with help from USFS researchers’ in incorporating wildfire smoke emissions, was designed to provide local air quality forecasters and the public with predictions of the transport and dispersion of particulate matter from large wildfires. The following sections describe the operational system.

b. Smoke and fire detection using HMS

Fire locations for the dispersion simulation are obtained daily for the prior day’s fires from the NOAA/ NESDIS HMS (Ruminski et al. 2007). The HMS is an interactive processing system that allows the trained satellite analysts in the Satellite Analysis Branch (SAB), within the Satellite Services Division (SSD), to manually integrate data from various automated fire detection algorithms with geostationary and polar-orbiting images. The result is a quality controlled display of the locations of fires and significant smoke plumes. There is no attempt to distinguish between wildfires and agricultural/prescribed burns.

Currently, the HMS analysis domain is adjusted seasonally, covering the coterminous United States, Hawaii, Mexico, and Central America daily from October through March and expanding northward to include Alaska and Canada from spring into autumn during this region’s prime wildfire and burning season. The analysis for Mexico and Central America is performed by the Servicio Meteorologico Nacional in Mexico by arrangement and incorporated within the analysis generated by the SAB. Any of the regions can be analyzed during off-peak periods as conditions warrant.

The HMS incorporates imagery from seven NOAA and NASA satellites, which allows for continuous monitoring. Geostationary data are obtained from GOES-11 and GOES-12 (also known as GOES-West and GOES-East, respectively) and offer high temporal resolution (data refresh of 15 min) and a nominal spatial resolution of 1 km for visible imagery (which is used for smoke detection) and 4 km for the 3.9-\(\mu\)m band (which is employed for hotspot detection). Polar-orbiting data are currently provided by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on both the NASA Terra and Aqua spacecraft as well as the AVHRR on NOAA-15/17/18. The polar data provide a higher nominal resolution of 1 km for the 3.9-\(\mu\)m band but at lower temporal refresh rates. Low- and midlatitude locations are scanned twice per day by each of the polar-orbiting satellites while higher latitudes receive more frequent coverage (up to six orbits per day in Alaska and northern Canada). The MODIS Terra and NOAA-17 spacecraft have similar equatorial crossing times near 1030 and 2230 LST (however, the NOAA-17

Unauthenticated | Downloaded 08/17/22 04:19 PM UTC
Smoke detection is achieved exclusively with visible band imagery. Owing to the nature of smoke plumes for active fires (the source remains at a fixed location) and the need to often discriminate between smoke and clouds, the primary platform used for smoke detection is GOES. The satellite viewing angle in conjunction with the solar zenith angle are such that optimal smoke detection is achieved for the contiguous United States with \textit{GOES-I11} (centered over the equator at 135°W) in the morning and with \textit{GOES-I12} (centered over the equator at 75°W) in the evening. Occasionally, polar imagery from the evening \textit{NOAA-I5} pass is employed due to its crossing near sunset, which allows for enhanced smoke discrimination. Polar imagery is also used more frequently in the northern regions of Alaska and Canada due to the more frequent coverage, which allows for animation of the imagery with a shorter time interval than at lower latitudes.

To produce an accurate representation of the smoke from fires, the source locations of the fires producing smoke must be adequately defined. For this reason, considerable time and effort are expended to generate the most comprehensive fire analysis possible. This process involves manually inspecting each of the automated detections and evaluating it for accuracy and also inspecting the raw satellite imagery for additional fires that the automated algorithms may not have detected due to strict screening thresholds. Additional details on this phase of the analysis can be found in Ruminski et al. (2006).

When smoke is identified, the analyst draws contours that depict an approximate concentration (thickness) and areal extent at a given time or time interval. As noted earlier, one of the methods used to identify smoke plumes is to locate a feature that has a fixed source. This can then be verified by noting whether a hotspot is also present at the source of the plume; although occasionally a smoke plume is the only indication of a fire and there is no corresponding hotspot detected because the fire may have had a very short burn duration or is not seen through clouds, forest canopy, or due to its location on a sloped surface. An additional means for differentiating between smoke and clouds is to compare the visible to infrared imagery. Clouds can be seen in infrared imagery while smoke normally is not discernable owing to their different radiative properties.

While the SFS features the incorporation of real-time satellite fire detection data, these data are a source of uncertainty. A major issue is predicting which fires will be continuous throughout the forecast period. Some fires, such as large wildfires, are both easily detected and likely to continue to burn. However, this is not the case for most of the agricultural/prescribed burns and many of the small wildfires. The fire area is currently determined by the number of detections in a grid cell, which can also be problematic due to satellite sensor characteristics. Smoke detection using visible band imagery is not possible at night and the presence of clouds may hinder or completely eliminate the capability to detect smoke. During large fire outbreaks, smoke can become lofted and remain suspended for many days where it can mix with anthropogenic sources such that only a subjective estimate of the smoke demarcation is possible. There are several current and planned sensors, mainly on polar-orbiting satellites, that can provide enhanced smoke detection, but these were designed for research missions and have not been easily folded into an operational system such as the HMS due to limitations in coverage and timeliness. In future versions of HMS, the incorporation of additional satellite data and more advanced processing algorithms (Zhang and Kondragunta 2008) may alleviate some of these limitations.

c. Fire duration

Analysts denote which fires are producing smoke for inclusion in the forecast calculation. Initially, all fires were modeled to emit smoke continuously, which is generally representative for large wildfires. However, this is not appropriate for many of the agricultural/prescribed burns. Therefore, since May 2006, analysts have specified the time of initiation and the duration of each fire as determined by observing hot spots in satellite imagery for which smoke emissions are observed. In addition to more accurately representing the amount of emissions, the dispersion accuracy can also be improved through a more accurate representation of the time of emission due to wind shifts and vertical temperature–wind profile changes that occur during the course of the day. Typically, GOES is used to determine fire initiation and duration based on interrogation of visible imagery as well as information from the 3.9-μm band from the other available satellites. If polar imagery is available and representative for a particular emitting fire, it is used instead of GOES to define the areal extent of the fire due to its higher nominal spatial resolution. However, since the objective is to provide the most up-to-date information, a polar satellite depiction of a fire would not be considered representative if, for example, a MODIS image from 1830 UTC captured the fire in question but subsequent GOES imagery indicated a marked increase in the size and intensity of the fire.
and a corresponding increase in the amount of smoke emissions.

d. Emissions

The PM$_{2.5}$ emissions and plume rise are estimated from the emissions processing portion of the USFS’s BlueSky smoke modeling framework (Larkin et al. 2009; O’Neill et al. 2009) based on fire size and location. BlueSky (information online at http://www.fs.fed.us/bluesky) is a fire and smoke prediction tool that was originally developed for land and air quality managers to assist with wildfire containment and prescribed burning decisions while at the same time attempting to minimize the impacts of the smoke on the local population. The USFS system links together models of fire characteristics, meteorology, emissions estimation, smoke dispersion, and the graphical display of output products. The NOAA adaptation of BlueSky uses its emissions estimation component, which consists of the Emissions Production Module (EPM) integrated with the CONSUME model (Sandberg and Peterson 1984), and the National Fire Danger Rating System (NFDRS; Cohen and Deeming 1985) fuel loadings database. The BlueSky emissions module currently has the capabilities to create emissions for CO, CO$_2$, CH$_4$, nonmethane hydrocarbons (NMHC), PM, PM$_{2.5}$, and PM$_{10}$. NOAA’s operational implementation incorporates only the PM$_{2.5}$ emissions for the predictions; however, the incorporation of multiple species is being tested for implementation in future versions.

Plume rise is computed assuming an air parcel’s rise is based only on the buoyancy terms (Briggs 1969; Arya 1999) using the fire heat release (from BlueSky), the wind velocity, and the friction velocity during the day and the static stability at night. During the day the plume rise is limited to the top 75% of the mixed layer while at night the plume rise can be up to 2 times the mixed layer depth. In the smoke prediction computation, smoke particles are released at the final plume rise height from the center of each emission grid cell that contains one or more fire location(s).

As indicated previously, the number of fires detected in a region for a specific fire is used to estimate that fire’s size for input into the BlueSky system based upon a nominal detection resolution of 1 km$^2$. Zhang and Kondragunta (2008) observed the average fire size from half-hourly GOES subpixel fire data varied diurnally from less than 0.1 km$^2$ at night to more than 0.2 km$^2$ during the early to midafternoon for various ecosystem types between 2002 and 2005. To simplify the conversion, a standard value of 10% of the pixel area was used to estimate the active (or instantaneous) burn area for each fire detection regardless of the fire detection method; however, future development will be focused on better representing the burn area with recent GOES fire product research. The fire detects were then aggregated onto a 20-km-resolution grid, with each detection adding an additional 10 ha to the burning area. The fire detection process tags all fire locations with a starting time and duration. Fires identified as continuous will emit from their starting time through the end of the 48-h forecast. Noncontinuous fires, mostly associated with agricultural land clearing, only emit for the duration specified.

The aggregated emissions file is saved each day and then loaded the next day by the preprocessor. Any grid cell that had an emission on the previous day’s model run but not on the current model run is assigned an emission based on the previous day’s emission. However, the previous day’s emissions are assumed to decay at a rate of 75% per day until the emission cell has less than one fire detection, at which point the cell emission is set to zero. This process ensures that predictions treat large continuous fires as continuing to burn for at least 1 day if no new fire detection data are available (days when cloud cover or other factors may restrict fire detections), but decaying quickly enough so that smaller fires will not extend beyond 1 day of missing fire detections.

e. Meteorology

The dispersion model calculation currently uses hourly, 12-km horizontal resolution NCEP North American Mesoscale-Weather Research and Forecasting (NAM-WRF; Janjić 2003; Janjić et al. 2004) meteorological data fields over the CONUS and 3-h, 1° horizontal resolution NCEP Global Forecast System (GFS) data fields for any fire locations outside of the NAM-WRF domain. The NAM-WRF data fields were interpolated from the native 60-level hybrid sigma-pressure coordinate to a 22-level sigma coordinate system having approximately 10 levels below 1500 m until 19 December 2007 when a sigma-level NAM grid was incorporated into the SFS. The GFS data fields are provided on 23 pressure levels between the surface and 20 hPa with a vertical resolution of 25 hPa up to 900 hPa and 50-hPa resolution up to 50 hPa. The smoke forecast is run at NCEP once each day using the 0600 UTC NAM-WRF forecast cycle in order to provide air quality forecasters with a 48-h smoke forecast by 1300 UTC. HYPLIT requires a 24-h meteorological “analysis” in addition to the 48-h forecast because fire locations are reported for the day prior to when the forecast starts. The 24-h NAM-WRF and GFS “analyses” consist of a combination of initial fields (±0 h to +5 h forecasts) from the previous four model cycles (0000, 0600, 1200, and 1800 UTC).
f. HYSPLIT model configuration

Smoke forecasts are produced daily by the NWS using the ARL HYSPLIT dispersion model, a hybrid between Lagrangian and Eulerian approaches. Advection and diffusion calculations are made within a Lagrangian framework following the transport, while concentrations are calculated on a fixed grid. The transport and dispersion of a pollutant can be calculated by assuming the release of puffs with either a predefined Gaussian or top-hat (zero outside, one inside) horizontal distribution that increases with time, or from the turbulent dispersal of an initial fixed number of particles, or by combining both puff and particle methods by assuming a puff distribution in the horizontal and particle dispersion in the vertical direction. In this way, the greater accuracy of the vertical dispersion parameterization of the particle model is combined with the advantage of having fewer pollutant puffs to represent the horizontal distribution. The model can be configured to support a wide range of simulations related to the transport, dispersion, and deposition of other pollutants as well, such as volcanic eruptions, emissions of anthropogenic pollutants, and dust storms. HYSPLIT requires, at a minimum, three-dimensional fields of the vector wind components and temperature. Further detailed descriptions of the HYSPLIT model can be found in Draxler and Hess (1997, 1998).

For the SFS, HYSPLIT is configured to run over North and Central America (Fig. 2) once a day using the actual fire locations observed by satellite from the previous day. HYSPLIT releases Lagrangian particles with a top-hat mass distribution in the horizontal and a three-dimensional particle distribution in the vertical at each time step over the reported fire duration at the center of each aggregated emission grid cell. As mentioned earlier, the smoke particle release height is assumed to equal the final buoyant plume rise height following the method of Briggs (1969). The final plume rise is a function of the estimated heat release rate from the BlueSky emission algorithm and the forecast stability and wind speed at each release hour. The maximum number of particles is limited to 250 000 over the entire model domain. Smoke particles are subject to dry deposition (including turbulent diffusion) in addition to gravitational settling and wet removal using meteorological model precipitation. Particles are assumed to have an average diameter of 0.8 μm and a density of 2 g cm⁻³ (Ferrare et al. 1990; Reid and Hobbs 1998; Colarco et al. 2004; Muller et al. 2005). Wet removal is much more effective than dry deposition. Smoke particles in grid cells that have reported precipitation deposit as much as 90% of their mass within a few hours depending on the rainfall rate.

Two 15-km resolution air concentration output grids are defined for each simulation. One grid contains hourly averaged air concentrations (μg m⁻³) of primary PM2.5 from the surface up to 5 km for verification with satellite smoke plume observations (section 3). A second grid defines the layer in the lowest 100 m as hourly averaged air concentrations of PM2.5 for comparison with routine surface air quality measurements. The dispersion simulation consists of two parts: 1) a 24-h analysis simulation run for the previous day and 2) a 48-h forecast simulation, with the assumption that yesterday’s fires will continue to burn for the next 2 days for those fires that have been tagged as having a 24-h duration. The smoke particle positions and masses at the end of each analysis period are used to initialize the next day’s analysis simulation (Fig. 3). Particle age is limited to 72 h after release to help maintain the timely completion of smoke predictions, restricted by a 2-h run-time window.
g. Output products

The operational NWS graphical output for each forecast hour over the CONUS is posted daily (Fig. 4), as part of the NAQFC guidance, in the NWS National Digital Guidance Database (NDGD; information online at http://www.weather.gov/aq/). The NDGD was initially designed as a graphical tool to display numerical weather prediction elements, and leverages the development and graphical techniques used by the National Digital Forecast Database (NDFD; Glahn and Ruth 2003). The NDGD and NDFD allow the public to query a forecast database that includes numerical guidance products (NDGD) along with interpreted weather forecast products (NDFD) created by NWS field forecasters. Twice-daily air quality forecasts of surface ozone concentrations from the Community Multiscale Air Quality (CMAQ) model (Eder et al. 2006) have been made available to the public via the NDGD since August of 2004, followed by the HYSPLIT smoke forecasts in March 2007. In addition to being able to zoom in on areas of the United States, users can click on any location and retrieve hour-by-hour ozone and smoke particulate forecasts specific to that location. In addition to the NDGD, Gridded Binary 2 (GRIB2; information online at http://www.nco.ncep.noaa.gov/pmb/docs/grib2/grib2_doc.shtml) formatted data files are posted to the NWS Telecommunications Operation Center’s ftp server (information online at ftp://tgftp.nws.noaa.gov/pub/SL.us008001/ST.opnl/DF_gr2/DC.ndgd/AR.conus/) for retrieval via ftp by users who may want to use the product in another display or analysis program.

For research purposes, smoke particulate forecast maps, as well as a 30-day rotating archive of the regional forecast graphics, are available from NOAA ARL’s smoke product Web page (information online at http://www.arl.noaa.gov/smoke.php). A “real time” model evaluation Web page is also posted to the ARL Web site each day that includes model graphics, HMS smoke images, narratives, and statistical graphs. One unique feature of this Web page is that the user has the ability to manually overlay both HYSPLIT and HMS smoke plumes that represent the same time interval on a map as well as fire locations used in the model run.

3. Model evaluation

As mentioned previously, a real-time model evaluation Web page is posted daily on the NOAA ARL Web site to provide an evaluation of the analysis portion of the current day’s model forecast and the previous day’s 24-h forecast valid at the same time as the current analysis. The intent of the real-time evaluation is to provide air quality forecasters with tools to judge the applicability of the current forecast based upon how well the predicted smoke compares with the actual smoke detected from HMS on the previous day. An initial evaluation using the total column (0–5 km) PM$_{2.5}$ concentrations was chosen instead of the surface PM$_{2.5}$ concentrations because the satellite data utilized in the HMS only represent total-column concentrations. Comparison with PM$_{2.5}$ observations across the United States is complicated by the sparse data network and the fact that the anthropogenic contribution is not modeled by this system. However, a further study by the authors (Stein et al. 2008) compares the surface concentrations to observations for several large events when the smoke happened to pass over surface observing stations and

---

**Fig. 3.** Computational cycle of the operational SFS. Particles at the end of the previous day’s analysis are initialized on the grid at the beginning of the forecast period prior to emitting new particles.
the results are encouraging, although additional changes may be needed to the model configuration to improve the surface concentration values.

a. Plume matching

To generate the products needed for the model evaluation, HYSPLIT was configured to output the latitude and longitude positions of four selected smoke concentration contours (1, 5, 20, and 100 \( \mu g \ m^{-3} \)) each hour. These contours were chosen instead of a single contour value because of uncertainties in the emissions and the threshold concentrations representing the visible edge of the HMS-analyzed smoke plume. Because it is not possible to assign a fixed numerical threshold concentration to the HMS-analyzed plume, the best evaluation contour may vary from day to day. HYSPLIT contour positions were converted into the Geographic Information System (GIS) shapefile format for ready comparison to the HMS smoke plumes, which are also provided in shapefile format (information online at http://satepsanone.nesdis.noaa.gov/FIRE/fire.html).

In the model evaluation system, hourly HYSPLIT (0–5 km) plume shapefiles for each contour level are compared to the HMS shapefiles valid at the same time by producing the figure of merit in space (FMS) statistic (Mosca et al. 1998; Boybeyi et al. 2001) for each matched shape at a fixed concentration level. The FMS is defined as the ratio of the intersection to the union of the plume areas:

\[
FMS(\%) = \frac{A_{\text{HMS}} \cap A_{\text{HYS}}}{A_{\text{HMS}} \cup A_{\text{HYS}}} \times 100,
\]

where \( A_{\text{HMS}} \) and \( A_{\text{HYS}} \) are the areas of the HMS and HYSPLIT plumes, respectively, as defined by a specified contour value. FMS/100 is also called the threat score or critical success index (CSI; information online at http://www.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores/docs/QandA.html) in the meteorological community. FMS scores can range between 0% and 100% and, in general, a high value indicates good model performance. However, because there can be significant errors in the model input parameters (emissions, meteorology, release height, etc.) and because HMS plumes do not always represent the exact outline of smoke observed over the time period due to their inherent subjective nature, a low value does not necessarily indicate poor performance of model transport and diffusion processes alone. Figure 5 shows some examples from the fall of 2007 of individual plume FMS scores and their relation to plume matching. Figures 5a and 5b show two cases where the model (horizontal stippling) appears to capture the observed (gray shading) plumes, with FMS scores of 65% and 43%, respectively. Figure 5c is a
case where the plume direction was modeled well, but the width of the observed plume was much smaller, resulting in an FMS of only 19%. Finally, Fig. 5d is a case where the modeled plume was shifted slightly from the observed plume, contributing to the small area of overlap and an FMS of only 12%. Even if there is complete overlap of the two plumes, but the measured or calculated plume is much larger than the other, the FMS will be reduced considerably in magnitude. Although the FMS is an imperfect metric, it is useful for obtaining a day-to-day snapshot of the model’s performance. It should also be noted that HYSPLIT plumes not intersecting HMS plumes are not included in the analysis because of uncertainties in fire detections and the possibility that some of the detected smoke is due to fires not represented in the model.

To compute the FMS statistics, multiple HYSPLIT shapefiles valid at the same hour and intersecting the same HMS shapefile are first merged into one HYSPLIT shapefile before producing the FMS statistic. This is to account for the fact that an HNS smoke plume may contain smoke from several individual fires modeled separately by HYSPLIT and also because as the HYSPLIT plumes age and lower in concentration, they tend to become spotty at a given threshold level. Finally, wet deposition can also cause the plume to become spotty after it passes through an area of precipitation. The spotty or “broken” HYSPLIT plumes (horizontal stippling) in Figs. 5b and 5d are good examples.

Typical individual FMS scores during the 2007 fire season ranged from 1% to 60% for the 1 and 5 μg m⁻³ contours, with the higher numbers generally achieved during active fire periods. The 20 and 100 μg m⁻³ contours produced similar results; however, the number of matched shapes was much smaller (especially for the 100 μg m⁻³ contour), because only the most intense fires produced HYSPLIT smoke at this level. Figure 6 shows a histogram of the 5 μg m⁻³ FMS scores for the 24-h analysis period of 30 October 2007, which represents a typical result during an active fire period.

Because there can be hundreds of matched plumes in a given active fire day, a daily FMS score is computed by dividing the sum of the individual intersection areas by the sum of the individual union areas, which can then be compared with previous daily scores to determine the overall current model skill.

Before showing the daily FMS results for the 2007 fire season, a summary of the fire season is presented.

b. 2007 fire season results

According to preliminary numbers from the National Interagency Fire Center, by the end of December 2007 more than 9.3 million acres had burned across the United States. This season ranks second to the fire season record set in 2006 when almost 9.9 million acres burned. From March through May of 2007 most of the fires occurred in the southeastern United States, an area that experienced unusually dry conditions and a severe-to-extreme drought (NCDC 2007). Several very large fires contributed to unhealthful air quality in parts of Georgia and Florida, including the Georgia Bay Complex fire in southeast Georgia and northeastern Florida.
that burned more than 560 000 acres (GFC 2007) between mid-April and mid-June. This fire was declared the largest wildfire in recorded history in Georgia. Also in June, fire activity increased in parts of the West, Southwest, Pacific Northwest, and the Canadian Northwest. By mid-July major fires in Idaho, Montana, and central Canada combined to send large areas of smoke across the northern United States and southern Canada. The fires in Idaho and Montana continued to burn through mid-September with smoke impacting areas as far east as the Tennessee Valley and the southeastern United States. Fire activity began to diminish in September with only a few large fires observed in North Carolina and California. However, a devastating fire event began on 21 October in southern California that eventually burned more than 500 000 acres, destroyed more than 2200 homes, killed 10 people, and prompted the largest evacuation in California history (information online at http://www.oes.ca.gov). These fires, spread rapidly by the Santa Ana winds and a long-term drought, produced intense smoke plumes that initially moved west into the Pacific Ocean (Fig. 7a), only to be returned inland when the Santa Ana winds subsided on 26 October (Fig. 7b).

c. Model results

The daily FMS scores for the 24-h analysis’s 1 and 5 μg m⁻³ contours are shown in Figs. 8a and 8b, respectively, and the 24-h forecast’s 1 and 5 μg m⁻³ contours for the period 1 September 2006–1 November 2007 are shown in Figs. 8c and 8d (the 20 μg m⁻³ contour produced similar results, although the values were slightly lower, while the 100 μg m⁻³ contours had very few matched shapes to compare). Overall, the 5 μg m⁻³ contour produced slightly better results (mean value of 11.6%) than the 1 μg m⁻³ contour (mean value of 8.3%) throughout the year for the 24-h analysis period. The 24-h forecast period produced lower FMS scores with mean values of 6.1% and 9.1% for the 1 and 5 μg m⁻³ contours, respectively. The model forecast was comparable to the analysis during the peak fire months of May–September but the forecast did not perform well during the winter. The reason for the remarkable difference between the analysis and forecast periods during the winter months is twofold. First, most of the fires during the winter months are short-duration fires that do not necessarily burn in the same location the next day. Because we cannot forecast new fires, nor do we include planned prescribed burns, we can only run the forecast with the fires that were observed the day before up until the start of the model run at 0600 UTC. Therefore, plumes observed from new fires when compared with HYSPLIT plumes from the previous day’s fires result in few matches. Second, after the emission is terminated, HYSPLIT continues to transport the smoke into the next day, but the smoke from smaller agricultural/prescribed burns is usually no longer visible by satellite; hence, there is no observed plume to match against. This result implies that the model may be emitting too much smoke, especially for small fires. During the summer months, the fires are larger and continue to burn for several days, allowing a comparison to be made between the observed and forecast plumes.

Figure 9 presents the same results in a series of monthly boxplots, which provide further information on the distribution of daily FMS scores. Note that the 5 μg m⁻³ analysis contour, in addition to having higher FMS scores, generally has less scatter than the 1 μg m⁻³ analysis contour, particularly during November and December. These results are consistent with observations that the HMS plumes tend to be smaller during the colder months and hence provide a better match to the smaller areas covered by the 5 μg m⁻³ contours. This result is consistent with the previous conclusion that emissions may be too high and/or fire duration is too long for the smaller fires.

Another statistical method used to measure the overlap area of two shapes was developed by Warner et al. (2004). The two-dimensional measure of effectiveness (MOE) differs from the FMS in that the MOE includes areas of under- and overprediction, also called areas of false negative (AP$_N$) and false positive (AP$_P$), respectively. The MOE has two dimensions, with the x axis corresponding to the ratio of the area of intersection of the HYSPLIT and HMS plumes to the HMS area, and the y axis corresponding to the ratio of the area of intersection of the HYSPLIT and HMS plumes...
to the HYSPLIT area. Simplification leads to the $x$ axis corresponding to 1 minus the false-negative fraction and the $y$ axis corresponding to 1 minus the false-positive fraction:

$$
MOE = (x, y) = \left( 1 - \frac{A_{FN}}{A_{HMS}}, 1 - \frac{A_{FP}}{A_{HYS}} \right) 100,
$$
where $A_{\text{HMS}}$ is the HMS plume area and $A_{\text{HYS}}$ is the HYSPLIT plume area, as in the FMS.

Figure 10a shows an example of the MOE plot on 23 October 2007 for the $5 \, \mu g \, m^{-3}$ HYSPLIT 24-h analysis contour. The point (100:100) represents a perfect match between the HYSPLIT and HMS plumes; that is, both shapes have the same size and location. Points along the 1:1 line represent the HYSPLIT and HMS shapes that are identical in plume area but are shifted in space, so that a point at (0:0) signifies no overlap. Points in the upper-left portion of the plot are cases where the HYSPLIT plume is nearly covered by the HMS plume; however, the HMS plume is larger than the HYSPLIT plume (underprediction). Conversely, points in the lower-right portion of the plot are cases where the HMS plume is nearly covered by the HYSPLIT plume; however, the HYSPLIT plume is larger than the HMS plume (overprediction). MOE results were also computed for each threshold concentration (analysis and forecast) for the entire 2007 fire season and are presented in Fig. 10b. The MOE results also indicate that the $1 \, \mu g \, m^{-3}$ threshold tends to be overpredicted, whereas the higher thresholds tend to be underpredicted. The size of the $5 \, \mu g \, m^{-3}$ threshold contour is forecast correctly; however, the area of overlap of the two plumes can still be improved. The daily MOE plots vary from day to day and can be compared with the overall MOE values to give the air quality forecaster a general idea of the over- and underprediction of the forecast for a given day that the FMS score alone cannot provide.

d. Objective verification

As mentioned earlier, verification using the shape-matching approach is very dependent upon the physical size, shape, and orientation of the predicted and observed smoke plumes. Differences in smoke plume areas can be due to errors in the emissions and the meteorology, as well as problems with the detection of the smoke plume. Given that the observed HMS smoke plumes are subjective and are dependent on an analyst’s interpretation of the satellite imagery, an objective verification method was sought using satellite tools being developed by NOAA/NESDIS. The GOES Aerosol/Smoke Product (GASP), described by Knapp et al. (2005), is currently used to assist the satellite analysts in assigning a smoke concentration to an area of smoke. (GASP is also available online at http://www.ssd.noaa.gov/PS/FIRE/GASP/gasp.html.) GASP is derived using the concept that higher reflectance at a given location in visible imagery that is not due to clouds may be caused by aerosols. Therefore, a clear-sky composite reflectance reference image is generated for each location for each observation time by using the second darkest pixel from the previous 28 days. The current image is
screened for clouds and the surface reflectance is obtained from the background image, using lookup tables calculated with a radiative transfer model. The output from GASP is the aerosol optical depth (AOD), which is estimated using the lookup tables from the radiative transfer model and the calculated surface reflectance. Normally, the larger the difference between the calculated reflectance and the reference image from the previous 28 days, the higher the AOD.

NESDIS recently implemented a further improvement to the GOES observation product for smoke verification: the automated smoke detection and tracking algorithm (ASDA; J. Zeng and S. Kondragunta 2008, personal communication). Since the HYSPLIT forecast is for smoke concentration, the GOES AOD, which is derived using a continental aerosol model, is scaled for a smoke aerosol model. ASDA can separate biomass burning smoke aerosol from other types of aerosols such as dust and urban haze using AOD imagery and fire hotspots (Fig. 11). A pattern recognition technique is then used to track the smoke plumes that drift away from the fire sources. Finally, the AODs are converted to smoke concentrations assuming a mass extinction coefficient of 7.9 ± 4.5 m² g⁻¹ (Dobbins et al. 1994) with the smoke confined to the lowest 5 km. The conversion from an AOD to a smoke concentration allows for validation with the output from HYSPLIT. This algorithm has been applied for some recent major fires in the United States and southern Canada and validated using the Ozone Mapping Instrument (OMI) aerosol products. Results are encouraging with the agreement between the GOES-I2 smoke optical depth and the OMI total optical depth for

![Boxplots of daily FMS scores](image-url)
absorbing aerosol (aerosol index \(>2.0\)) being in the neighborhood of \(\pm 0.2\) AOD.

GASP and ASDA have two desirable properties for use with the SFS; they are derived routinely (every half hour over sunlit portions of the GOES domain) and provide an objective, quantitative estimate of AOD and smoke concentration. However, there are also limitations. As with other satellite-based estimates of aerosol content using visible channels, GASP performs better over darker surface backgrounds (ocean and moist continental areas) as compared to semiarid regions (such as portions of the western and central United States). While GASP employs GOES visible band imagery, which has 1-km nominal resolution, ASDA is limited to 4-km resolution because the GOES infrared band imagery used in the cloud-clearing portion of the algorithm is provided at 4-km resolution. Thus, some of the smaller smoke plumes are either not depicted or may have smoke concentration values that are diminished due to averaging with adjoining nonsmoke pixels. Retrievals are also not performed for high solar zenith angles due to a known high bias.

The NCEP Forecast Verification System (FVS; Shafren 2007) was modified to routinely compare the ASDA-detected fire-related plume concentrations with predicted HYSPLIT smoke concentrations. The FVS currently allows evaluation of 11 upper-air and 14 surface variables including both meteorological and trace variables. The modeled and observed smoke concentrations are first compared grid cell by grid cell according to a set of prescribed threshold values. Then, the numbers of forecasted \((F)\), observed \((O)\), and hit \((H,\text{forecasted and observed})\) values are tabulated and stored in a verification statistics database (VSDB) file. Figure 12 shows the relationship between the \(F, O,\) and \(H\) values and a traditional \(2 \times 2\) box definition. The FVS computes and displays various skill scores from the VSDB file and the entire process is fully automated for daily execution. A more detailed FVS description can be found online (http://www.emc.ncep.noaa.gov/mmb/papers/brill/FVShelpfile07.txt and http://www.emc.ncep.noaa.gov/mmb/papers/shafran/gridtob.pdf).

The NCEP FVS smoke PM\(_{2.5}\) verification for HYSPLIT performance began during the fall of 2006 and recent results can be found online (http://www.emc.ncep.noaa.gov/mmb/hchuang/web/html/score_mon.html, under the HYSPLIT Smoke Verification option). In addition to the full domain, this verification focuses on two subregions, the western and eastern United States. Seven threshold values are used to examine the model performance, including \(<1.0, >1.0, >2.0, >5.0, >10.0, >15.0,\) and \(>20.0\) \(\mu g\) \(m^{-3}\). Figure 13 shows the monthly average skill score of HYSPLIT smoke PM\(_{2.5}\) predictions for threshold values of \(>1.0\) \(\mu g\) \(m^{-3}\) and \(>5.0\) \(\mu g\) \(m^{-3}\) from May to October 2007. Scores for the entire domain ranged from 8.5% to 30% with a mean score of 15.1% for \(>1.0\) \(\mu g\) \(m^{-3}\) thresholds and 4.4% to 15.9% with a mean score of 8.2% \(>5.0\) \(\mu g\) \(m^{-3}\) thresholds. The mean skill scores are 11.2% (4.1%) and 13.7% (8.1%) for the eastern and western United States, respectively, for threshold values of \(>1.0\) \(\mu g\) \(m^{-3}\) (\(>5.0\) \(\mu g\) \(m^{-3}\)). The HYSPLIT fire-related plume forecasts are predicted reasonably well during the fire season for thresholds \(>1.0\) \(\mu g\) \(m^{-3}\). HYSPLIT generally performed better in
the summer months as well as in the western United States (skill score from 8.1% to 30%). Scores for the western United States in August 2007 can be as high as 11.5% and 8.0% for the threshold values of $>10.0 \, \mu g \, m^{-3}$ and $>15.0 \, \mu g \, m^{-3}$, respectively (results not shown here). The skill score for the higher threshold provides an estimate of model performance on the larger and more intense fire events.

Although the FVS cannot produce MOE values directly at this time, MOE values were computed from the output of other FVS parameters for the May–October 2007 forecast period and are presented in Fig. 14 for the 1 and 5 $\mu g \, m^{-3}$ thresholds. Again, note that the 1 $\mu g \, m^{-3}$ threshold produced the better forecast with the least over- or underprediction when compared with the 5 $\mu g \, m^{-3}$ threshold using the satellite-derived product.

Unlike the HMS plume-matching verification, the NCEP FVS verification indicates that the 1 $\mu g \, m^{-3}$ threshold provides a better fit between HYSPLIT- and ASDA-derived concentrations than the 5 $\mu g \, m^{-3}$ threshold. As these are two independent types of verification, one comparing grid cells of predicted and derived smoke concentrations (ASDA) and the other comparing predicted and observed smoke plume shapes (HMS), it is not unexpected that the results are different. The higher FMS scores for the 1 $\mu g \, m^{-3}$ threshold are likely due to the HYSPLIT plume being larger at that threshold and, therefore, covering more ASDA-identified smoke grid cells. At this point, although quantitative verification of predicted smoke concentrations is not possible, the objective verification is serving as a baseline for future enhancements to the smoke prediction system.

**FIG. 11.** GOES-12 smoke aerosol image depicting aerosol optical thickness at 2245 UTC 17 Aug 2007. The high values seen across the Ohio River valley correspond to smoke originating from fires in ID and MT.

**FIG. 12.** The definitions of forecast ($F$), observed ($O$), and hit ($H$) in the NCEP FVS.
4. Summary

The SFS has been implemented operationally as an interim smoke forecast tool in the NAQFC. The dispersion model, HYSPLIT, together with the BlueSky emission algorithm and the HMS satellite analysis, provide air quality forecasters and the public with a daily forecast of smoke concentration (PM$_{2.5}$) over North America from fires large enough to produce visible smoke from satellite.

The HMS, evolving since its inception in 2001, has proven to be a valuable tool for providing a source of model validation for the SFS and its most recent and planned enhancements are focused on improving the detection, depiction, and forecasting of the smoke generated from wildfires and agricultural/prescribed burns. One of the advantages of the system is that the detection and specification of smoke areas and smoke-producing fires is performed in near–real time using constantly updating environmental satellite data and covers all of North America. The ability to specify the initiation and duration of emissions by analysts using HMS has also allowed for a more realistic modeling capability.

Evaluation utilizing the HMS has shown daily values of FMS to be as high as 60% with monthly averages around 8% and 12% for the 1 and 5 µg m$^{-3}$ contours, respectively. The best results are typically obtained during active fire periods when large smoke plumes are generated.

Being objective, the use of NESDIS ASDA products within the NCEP verification system (FVS) has also provided an independent, near-real-time method of the verification of the SFS. Results for the summer of 2007 are consistent with the HMS verification for the 5 µg m$^{-3}$ threshold but perform even better for the 1 µg m$^{-3}$ threshold. This is most likely a result of comparing grid cell to grid cell in the ASDA verification, whereas the HMS method compares complete plume shapes.

Because the ASDA verification has only been running for one full season, both methods of verification...
will continue to be applied and the results will be posted to their respective Web sites.

Acknowledgments. This project was carried out under the auspices of NOAA’s National Air Quality Forecast Capability and the views expressed in this paper are those of the authors and do not necessarily represent those of NOAA. The authors thank Sim Larkin and Susan O’Neill of the Pacific Northwest Research Station, United States Department of Agriculture’s Forest Service, for their help with incorporating the BlueSky smoke modeling framework into the NOAA Smoke Forecasting System.

REFERENCES


