VERIFICATION AND VALIDATION OF DESERT DUST FORECASTS AND THEIR IMPACT ON RESPIRATORY HEALTH APPLICATIONS IN THE SOUTHWESTERN UNITED STATES

S.A. Morain and A.M. Budge

*Earth Data Analysis Center, University of New Mexico, Albuquerque, NM, USA-
(smorain,abudge)@edac.unm.edu

Commission VIII, WG VIII/2

KEY WORDS: Satellite remote sensing; Environmental pollution, Image understanding, Spatio-temporal modelling, Medicine, Multi-spectral remote sensing, Visualization

ABSTRACT:
The paper describes the performance of the Dust Regional Atmospheric Model (DREAM) after being nested within the National Centers for Environmental Prediction (NCEP/eta) operational weather model. The model system is designed to simulate dust entrainment, transport, and concentration under changing atmospheric and terrestrial conditions to better forecast dust episodes that impose health outcomes on populations at risk. The approach was to: (1) benchmark the model system’s performance after it was modified to fit a new domain in the southwest U.S.; (2) verify and validate model outputs statistically by comparing outputs from ground station observations; and (3), replace selected baseline terrain parameters in DREAM by assimilating comparable satellite data and comparing the results. Having base-lined, verified, and validated the model system’s performance, Earth observation data were assimilated sequentially into a series of new model runs. Terrain parameters were assimilated for barren ground dust sources, fraction of photosynthetically active radiation; digital topography, aerodynamic surface roughness, and soil moisture. Replacement of the baseline parameters with these assimilated parameters improved dust model performance without imposing negative impacts on observed meteorological fields. Major gains were made in modelling the onset of dust storms, the timing, of peak hour concentration, and duration of near-surface high dust concentrations. There was no verifiable improvement in measuring the magnitude of dust concentrations, even though the enhanced model predicted accurately the occurrence of dust storm events at most locations in the model domain.

1. INTRODUCTION
The incidence of respiratory and cardiovascular diseases is higher in populations having frequent exposure to high dust concentrations. The body of epidemiology linking dust and aerosols to health outcomes is growing rapidly (Griffin, 2007). For human health practice, the first challenge is to show that satellite Earth observation (EO) data can be integrated reliably into models that improve predictions of dust levels that trigger respiratory responses. Another challenge is for medical practice to extract from dust exposure data the consequent flow of pathogens and chemicals through airborne mechanisms, and to translate these findings into actionable human health interventions (Pope, 2004). While the medical community recognizes the adverse effects of PM$_{10}$ and PM$_{2.5}$ in patients with respiratory conditions, they lack proven technology for forecasting the onset and severity of dust episodes in sufficient time to issue alerts or to implement health interventions.

2. MODEL SYSTEM
The model used for operational weather forecasting is the National Centers for Environmental Prediction (NCEP), eta version (NCEP/eta). It simulates large-scale numerical solutions controlled by conservation of integral properties. It uses a non-linear horizontal advection numerical scheme that preserves energy and squared vorticity and controls non-linear energy cascade. With the eta vertical coordinate, which generates quasi-horizontal model levels, topography is represented by step-like elements. Physical parameterization includes land surface processes, turbulent mixing, convection, large-scale precipitation, lateral diffusion, and radiation. However, it does not simulate dust events. For simulating these events, the Dust Regional Atmospheric Model (DREAM) has been nested within the NCEP/eta simulator to form the model system used in this research (Janjic, 1984; Mesinger et al., 1988; Janjic, 1994; Nickovic et al., 2001). DREAM was originally developed for use in the Mediterranean region and was run as a European Center for Medium-Range Weather Forecast (ECMWF) product using initial and boundary conditions of one degree. Verification and validation of the model’s outputs are reported by Nickovic et al. (2004), and Perez et al. (2006). The DREAM/eta system is currently undergoing extensive V&V analyses. Preliminary results are given in (Morain and Sprigg, 2007).

3. THE V&V SYSTEM
Verification and validation of DREAM outputs has been done by making qualitative and quantitative (statistical) comparisons of model outputs with in-situ dust concentrations reported by ground based networks. Development efforts focused on three tasks: creating a model output archive; developing a data management system for web services; and defining statistical measures.

1355
3.1 Model output archive

The first task was to generate an archive of DREAM dust concentration data. This includes a daily DREAM model run for the 48-hour forecast beginning at 00:00:00 hours of the previous day. It also includes a twice-daily DREAM model run beginning in 2006. The archiving system is designed to execute three model runs per day and a rolling 72-hour forecast for the current day. The configuration of the model prevents concurrent execution of runs, so they are scheduled to minimize the potential for conflict. A single model run executes in approximately 5 hours, so a two hour buffer has been built into the execution schedule.

3.2 Data management and web services

The second task was to develop web services that permit system developers and health-care users to search for, access, and download dust concentration data generated by the DREAM model, as well as data collected by in-situ networks. Both the historical and daily forecasts are integrated into the DREAM data management system for delivery to public health decision support systems through simple object access protocols (SOAP) and web mapping service (WMS) interfaces published by the project (Budge et al., 2006).

The web service architecture allows users to search for and download both PM$_{2.5}$ and PM$_{10}$ particulate data from in-situ monitors, as well as DREAM model output values for specific locations. Users can download PM$_{2.5}$ or PM$_{10}$ data for a defined date range, or for a single day. Similarly, SOAP service functions allow one to download both in-situ and DREAM dust concentration values for a single station, or for all stations within the modelling domain; or to download data for a specific day, a 48-hour period corresponding to a DREAM model run, or a date range specified by the user. At present, data from in-situ monitors are not segregated into species. The downloadable in-situ values represent a composite measure of both geologically-derived and anthropogenically-produced particles (see Section 4.2).

3.3 Statistical measures

The third task was to create web services that allow developers to generate statistical measures and indices. One of these, the DREAM Data Access and Statistical Wizard, allows one to extract modeled dust values for specified X-Y coordinates at specified times, and combine them with in-situ values to generate statistics. In order to verify and validate the performance of consecutive versions of the model, web services have been designed to calculate measures of central tendency and measures of variability for both observed and modelled dust concentration values. These measures include the mean and standard deviation. Another set of statistics provides measures of association between these two variables. These include: mean observed value at each site; mean bias (0% if perfect); mean error (0% if perfect); normalized mean bias (0% if perfect); normalized mean error (0% if perfect); fractional bias (0% if perfect); fractional error (0% if perfect); and index of agreement (1 if perfect); the correlation coefficient (R); and the centered root mean square (RMS). These statistics can be obtained for a single station for a date range specified by the user.

4. IN SITU V&V DATA STREAMS

4.1 AIRNow reporting stations

Hourly PM$_{2.5}$ and PM$_{10}$ data are available from the U.S. Environmental Protection Agency (US/EPA) AIRNow network for the entire period 2006 to present through the DataFed’s AIRNow Web Coverage Service (WCS). These data are acquired daily as a comma separated value (CSV) file for all EPA stations within the DREAM domain for the previous 60 days. The daily reacquisition for the previous 60 days corrects data for stations that experienced delays in submitting values either to EPA’s network or to DataFed’s data ingest system.

During the development phase of the V&V system, a question arose regarding the timestamps encoded into the CSV files. Initially it was thought there was an undocumented offset to UTC, but subsequent discussions with DataFed revealed that timestamps encoded in the AIRNow data files varied by day and station, and that these timestamps were not consistently converted to UTC. This led DataFed to reconfigure its services to provide AIRNow data in UTC, regardless of the offset in the original data. This standard UTC format now provides unambiguous alignment of DREAM model outputs with well-defined ground observation times.

The current web interface has 94 PM$_{2.5}$ and 41 PM$_{10}$ sites for which modelled and observed data are co-located for side-by-side comparisons. Many sites have missing data for lengthy periods, especially for days of known dust events. It is suspected that in-situ sensors fail under extreme conditions and/or reporting of these events is delayed. It is unclear how many sites within the in-situ network have this problem, but it happens often that dust events of interest have missing data at many sites. It is sometimes possible to obtain data from the AIRNow website itself rather than through the DREAM web interface. Also, there is an obvious gap in station coverage for PM$_{10}$ in central Texas, a region known to experience widespread dust events. Most AIRNow sites are located in cities, making validation over rural areas difficult. It has been shown also that the MOD12Q1 data for northern Mexico (included in the modelling domain) improve validation statistics at US stations (Yin et al., 2007); yet, there are no in-situ measurements from Mexico for use in V&V.

4.2 Speciation in PM$_{10}$ and PM$_{2.5}$ dust

There are drawbacks to comparing model outputs with AIRNow data for PM$_{10}$ and PM$_{2.5}$, because each fraction contains materials that are not generated by natural atmospheric processes. A more robust approach for health applications is to V&V these fractions continuously on the basis of individual species’ concentrations.

PM$_{10}$ being larger in diameter and mass than PM$_{2.5}$, requires more momentum and higher wind speeds to be entrained. After lifting, this fraction also settles out of the atmosphere quicker. Because DREAM is strictly wind driven, and PM$_{10}$ is almost always mechanically entrained, the coarse fraction is a better indicator of atmospheric dust events than PM$_{2.5}$. However, in-situ PM$_{10}$ may be present in arid environments even in the absence of wind, and in such cases would not be predicted by DREAM. Anthropogenic concentrations often are present when DREAM predicts none. Fugitive dust from off-road vehicles, agricultural and construction dust clouds and emissions of larger pollutants from automobiles and factories add biases to...
PM$_{10}$. During non-windy conditions, it is still possible to observe other sources of PM$_{10}$ that DREAM has no way of simulating. Due to its relatively large size, PM$_{10}$ deposits in the upper thoracic region of the human respiratory system, and is often a concern for silicosis (Policard et al., 1952; Bar-Ziv and Goldberg, 1974; Norboo et al., 1991).

PM$_{2.5}$, on the other hand, may be present before, and linger after, weather-driven events. It penetrates deeper into the lungs and is a serious concern for chronic asthma, myocardial infarction, and other respiratory and cardiovascular conditions. Furthermore, its smaller size, makes validation more difficult. There are many more types of particles in the fine fraction. These finer particles include organic carbon as smoke from fires, soot from automobile emissions, and photochemical products. Other gases react photochemically forming ammonium sulfates and ammonium nitrates in this size range. Trace metals are produced via industrial emissions. Finally, natural aerosols are created mechanically as sea salt or windblown mineral dust. This research focuses on the mineral dust component, but these other components of PM$_{2.5}$ material complicate measurement of particulate concentrations and therefore model performance. Total PM$_{2.5}$ as referred to here, is the net concentration of all species in the air for that size range. DREAM has no anthropogenic emission module, so the other species and the anthropogenic signal in total PM$_{2.5}$ have been ignored.

The importance of speciation is evident in analyses of urban areas. For example El Paso, Texas experiences both desert dust storms and anthropogenic pollution episodes. DREAM can only model the former, so distinguishing the two using speciation is extremely beneficial for V&V. It is evident that during days of dust storms, the soil component comprises a much larger fraction of the total PM$_{2.5}$, while on non-windy days the other species dominate. While this is promising for V&V purposes, more frequent in-situ data are needed. Presently, only daily averages taken every third day are used for speciation, so DREAM can be validated discretely only at this frequency. Continuous hourly data are ideal, but are probably not feasible due to cost and time restraints.

5. MODEL RUNS AND STATISTICS

Model set-up and runs have been produced by project collaborators at the University of Arizona, Department of Atmospheric Sciences (Yin et al., 2005; Yin et al., 2007). The Preparation of EO data sets for assimilation into the system were provided by the University of New Mexico Earth Data Analysis Center (Morain and Budge, 2006a). For V&V purposes, the model has been tested for the southwest U.S. for dust storm events occurring on December 15-16, 2003, January 4-6, 2007, and February 23-25, 2007. The 2003 event was used to baseline the DREAM/eta system before and after EO data assimilation. In 2005-2006, a series of runs was executed using different combinations of assimilated numerical EO data to assess each data set’s relative importance for improving model performance. Subsequent model runs in 2007 were based on the expanding model archive (see Section 3.1) using the V&V statistical measures (see Section 3.2). Results were reported in Morain and Sprigg (2005). V&V consists of two parts: how well the model simulates atmospheric parameters important for dust entrainment; and, how well DREAM simulates dust clouds and dust movement.

5.1 Atmospheric dust entrainment

The baseline model run compared observed and modelled surface and upper air patterns, as well as vertical profiles for wind speed, wind direction, and temperature to verify that DREAM/eta did not adversely impacted the NCEP/eta simulator output. Statistics were calculated using modelled data and hourly measurements from 95 surface synoptic stations and 633 surface METAR stations in the modelling domain. In addition to other statistics, the index of agreement for wind speed, wind direction and temperature all exceeded 0.7, indicating that the model simulates these parameters fairly well. Table 1 shows the performance statistics before and after EO data assimilation (Morain and Budge, 2006b). The agreement indices in the bottom row indicate that only a slight improvement is achieved for wind speed and wind direction by assimilating EO data, but that a significant improvement is achieved in the surface temperature parameter. Overall, the slightly higher index values improve the ability of the model to simulate dust entrainment. It is expected that migrating from NCEP/eta to NCEP/NMM, a non-hydrostatic, high resolution version of the atmospheric model, will further improve atmospheric simulations.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Wind Speed (m/s)</th>
<th>Wind Direction (°)</th>
<th>Temp (K)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Obs.</td>
<td>5.53</td>
<td>231.40</td>
<td>276.7</td>
<td>4</td>
</tr>
<tr>
<td>Mean Mod.</td>
<td>4.65</td>
<td>226.60</td>
<td>275.5</td>
<td>6</td>
</tr>
<tr>
<td>Mean Bias</td>
<td>-0.88</td>
<td>-4.80</td>
<td>-1.20</td>
<td>0.72</td>
</tr>
<tr>
<td>Mean Error</td>
<td>1.97</td>
<td>51.76</td>
<td>4.09</td>
<td>2.67</td>
</tr>
<tr>
<td>Agree Index</td>
<td>0.74</td>
<td>0.74</td>
<td>0.71</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table 1. NCEP/eta/DREAM baseline performance before and after EO data assimilation for a Dec. 2003 dust storm.Italic values are before EO data assimilation; other values are after assimilation. For the equations, M = modeled; O = observed.

5.2 Dust cloud detection and movement

For dust cloud detection and movement V&V analyses use the growing archive of model runs. The 72-hour rolling dust forecast alerts team members to impending dust events, but receipt of the AIRNow data needed for statistical comparison typically lag a few days behind the model runs. When the in-situ data become available, model outputs are compared to observed PM$_{10}$ and PM$_{2.5}$ in a hind-cast mode. The measures of greatest interest for monitoring human exposures to dust are: dust concentration; dust episode duration; and, hour of peak concentration.

One dust episode occurred in January 2007. A severe wind and dust storm near Barstow, California caused traffic accidents...
killing two and leaving others with severe injuries. Wind continued to interrupt traffic, freeways were congested, and several large trucks toppled or jack-knifed. High winds spread across the southwest eventually including parts of Texas. This dust event was investigated using DREAM model hind-casting.

Data from seven AIRNow monitoring stations were used for the analysis. Four were located in Southern California (Burbank, Riverside, Palm Springs, Indio) and three in Texas (El Paso, Mission, Selma). Figure 1 shows a 72-hour plot for each station (January 4-6, 2007) and illustrates the dust event that occurred around 2300 UTC on January 5th at most stations. The stations are plotted geographically west (on the left) to east (on the right). Southern California was affected most by this event. Both the observed and modelled data show a strong dust gradient from mild in the east to more severe in the west, with the exception of Riverside, where virtually no significant dust was recorded by the ground station data. Of particular note in Figure 1 is the difference in dust concentrations between model run 15a and 20a. Dust concentration is estimated in the model by partitioning particle sizes into four bins. PM$_{10}$ is extracted from parts of 2 bins. Therefore the modelled dust concentration can be higher or lower depending on how highly refined the extraction process is. Comparison between run 15a and 20a shows a slight decrease in dust concentration at several stations (Burbank, Riverside, and Palm Springs). This difference was obtained by refining the bin size algorithm in run 20a to use a narrower bin size.

Figure 1. Modelled and observed PM$_{10}$ concentrations at seven AIRNow stations across the southwest for January 4-6, 2007.

A statistical analysis that included the seven sites using the latest version of the model (20A) is shown in Table 2.

<table>
<thead>
<tr>
<th>N (seven sites)</th>
<th>443 obs / 443 mod</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>29.2 obs / 26.3 mod</td>
</tr>
<tr>
<td>Mean bias</td>
<td>2.8</td>
</tr>
<tr>
<td>Mean error</td>
<td>26.0</td>
</tr>
<tr>
<td>Normalized mean bias</td>
<td>10.8</td>
</tr>
<tr>
<td>Normalized mean error</td>
<td>76.2</td>
</tr>
<tr>
<td>Fractional bias</td>
<td>12.1</td>
</tr>
<tr>
<td>Fractional error</td>
<td>88.1</td>
</tr>
<tr>
<td>Index of agreement</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Table 2. Statistical analysis of seven test sites, Jan 4-6, 2007.

The timing correlations for two test cases are shown in Figure 3. The X-axis is a 72-hour event clock showing the observed peak hour concentration. The Y-axis shows the modelled peak hour concentrations during the event. Several sites had more than one peak hour during the three-day event. A plot of daily peak hours for each of the seven sites would yield 21 data points. Occasionally, however, no peak hour was evident, particularly on January 4. These results ($R^2 = 0.95$) for model version 20a show an improvement over previous versions of the model published in earlier work ($R^2 = 0.76$, Yin et al., 2005).

Figure 2. Magnitude correlation for seven sites during the Jan 4-6, 2007 event ($N = 443$).

Figure 3. Timing Correlation ($N=18$ peak hours, seven sites) for the Jan 4-6, 2007 dust event.
Another event was modelled for February 23-25, 2007. Very strong and gusty westerly winds caused blowing dust over a large area of eastern New Mexico and northwest Texas on the afternoon and early evening of February 24. A huge dust cloud was blown easterly across much of the eastern half of the state on the 25th and then stagnated over parts of central, southeast, and south Texas on the 26th and 27th. PM$_{10}$ levels in parts of the southern Panhandle were hazardous according to EPA’s Air Quality Index (AQI) scale.

Figure 4 shows the 72-hour plot for each station and illustrates the dust event that occurred around 00:00 UTC on February 24 at most stations. The stations are plotted geographically west (left) to east (right). Two versions of DREAM are plotted (15a and 20a). The DREAM model under-predicted the event at Palm Springs, over-predicted the event at Indio, but performed well at the Texas sites, particularly at El Paso. Observed data from Selma and Mission, TX indicated a minor event and the DREAM model outputs were in fairly good agreement for these sites.

Figure 5 illustrates the magnitude correlation between modelled and observed data for the February 23-25, 2007 test case. Correlations for both model versions were poor for this test case ($R^2 \sim 0.1$), due primarily to the Palm Springs and Indio data discrepancy. In spite of this, the timing correlation (Figure 6) again shows excellent agreement between observed and modelled peak hour.

The same statistical analyses that included seven sites using version 20a of the model are shown in Table 3 for the February storm. The statistics indicate that the model had a negative bias, or under-predicted the event. The January event (Table 1) had a positive bias and a much better index of agreement (0.63 vs. 0.42).


<table>
<thead>
<tr>
<th>N (seven sites)</th>
<th>346 obs/346 mod</th>
<th>341 obs/59.3 mod</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-25.0</td>
<td>56.0</td>
</tr>
<tr>
<td>Mean bias</td>
<td>-42.4</td>
<td>67.7</td>
</tr>
<tr>
<td>Mean error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm. mean bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractional error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index of agreement</td>
<td>0.42</td>
<td></td>
</tr>
</tbody>
</table>

The two events (January and February 2007) indicate that the model can accurately predict the timing of dust events, but overestimates their severity (magnitude). They also suggest that the model is sensitive to minor alterations of input parameters, in this case a slight change in particle size bin widths.

6. SUMMARY AND CONCLUSIONS

Remote sensing of the environment is critical in advanced systems to warn of imminent, life-threatening sand and dust storms and to reduce risk of exposure to mineral dust concentrations that contribute to respiratory and cardiovascular disease. MODIS data improve identification of active mineral dust sources, and thus, numerical model simulations and forecasts of dust generation, entrainment, and downwind dispersal and deposition.

An advanced numerical dynamical model of dust generation and entrainment (DREAM), driven by operational, validated, weather forecast models of the U.S. National Weather Service (eta) initialized with MODIS landscape information can forecast the timing of an advancing dust storm verifiably to meet the needs of many users. While the dust forecast system...
simulates and predicts the three-dimensional size-concentration characteristics of the dust cloud, verification of model output requires on-going verification and validation. V&V of airborne particulate concentrations rely primarily on a regionally sparse network of in-situ sampling stations for statistical comparison with DREAM-generated PM$_{10}$ and PM$_{2.5}$ concentrations. These sampling networks are concentrated in large, densely-populated urban areas that include PM$_{10}$ and PM$_{2.5}$ anthropogenic as well as atmospherically generated concentrations. There are too few speciated particle sampling sites available to identify natural vs. human-generated sources.

Products designed specifically with the end user in mind are being evaluated in state health offices with operational health and air quality responsibilities. These products will be modified as needed, and further V&V will play a large role in adapting/adopting the new technology developed under PHAiRS for public health services.

REFERENCES


ACKNOWLEDGMENTS

This contribution has been sponsored by the United States National Aeronautics and Space Administration (NASA) under Agreement # NNSO4AA19A, John Haynes, Project Manager. The Authors thank all PHAiRS team members for contributing materials that made this paper possible. These team members, in alphabetical order by affiliation, are Shirley Baros, Karl Benedict, Thomas Budge, and William Hudspeth, all of the University of New Mexico; Brian Barbaris, Patrick Shaw, and William Sprigg, all of the University of Arizona; Dazhong Yin, formerly at the University of Arizona.