



## Project Members:

University of New Mexico, Earth Data Analysis Center Karl Benedict4; Amelia Budge5; Thomas Budge6; William Hudspeth7; Peggy Allison8; Matt Gagnon9; Huiping Sheng10 4EDAC Director; 5Project manager; 6Image processing specialist; 7Data processing & information; 8Metadata; 9Air quality observer coordinator University of New Mexico Health Science Center, MPHS Program Huiping Sheng, Metadata & health data statistics University of Arizona, Department of Atmospheric Sciences Brian Barbaris10; Patrick Shaw11; Slobodan Nickovic12; Goran Pejanovic13; Ana Vucovic13 10Model verification & validation; 11Aerosol Modeling; 12DREAM model design; 13DREAM model performance & testing University of Arizona High Performance Computing & Operations Center Lucy Carruthers; Jimmy Ferng; & Cyrus Jones Universities Space Research Association Maudood Khan, CMAQ Modeling; MOD/AOD & CALIOP Data Analysis





In 2003 NASA's Office of Earth Science Applications inaugurated a systems approach toward research and development (R&D) projects across 12 societal benefit areas, among them air quality and health. The program focused on Federal requirements for documenting *inputs*, *outputs*, *outcomes* and *impacts*. In the lower left are Earth observations derived from appropriate space-acquired data from national and international sensors and platforms. These data sets were used as inputs to selected models designed for specific purposes. Outputs from the data and models were first verified and validated against independent (ground-based) observations before being configured into products that met operational agency requirements. In the *partner area* (green column) researchers and users interacted to design specific outputs that impacted the desired agency outcomes—in this case a decision support system that would supports actionable decisions for a state-wide environmental public health tracking system (EPHTS) and a national environmental public health tracking network (EPHTN). The values and benefits to society would be in the form of early warning of air quality events and the knowledge gained from new space-based products would represent the primary *impacts* shown on the right.

This presentation shows results from two NASA projects: Public Health Applications in Remote Sensing (PHAiRS) and Environmental Public Health Applications (ENPHASYS)



The model domain for PHAiRS had its domain center at (109°W, 35°N). It used a horizontal semi-staggered Arakawa E grid, with a horizontal grid spacing 1/3° (ca. 50 km). Weather conditions at the boundaries of the domain define the initial and boundary conditions for model runs. A higher resolution domain centered on the Four Corner states of Arizona, Utah, Colorado, and New Mexico is shown on the right. It was used in ENPHASYS for higher resolution CMAQ modeling of PM<sub>Fine</sub> and aerosols in the Four Corners. The model grid cell size in this domain is 7.5km.



## Definitions

<u>1. Fusion</u>: The process of including categorized **pictures of EO image products** (at any of several levels of processing) into a GIS architecture in such a way that the datasets, both vector and raster, are geospatially registered at a specified projection and scale. This usually requires sub-setting, re-projection and rescaling of the data.

2. Assimilation: The process of **replacing selected static parameters in an Earth system model with near real-time digital pixel values** from Earth observation data sets to improve the model's performance and convert it into a more dynamic (forecasting) form without changing the model's intended purpose.

Basic objective

Improve model output without altering the validity of the model's original function

## Simplified work flow

(1) convert the model into a dynamic forecast

(2) select satellite sensor data sets based on highest benefit for enhancing model output

(3) replace selected analog data with regularly refreshed EO digital data for "terrain," "surface conditions," and "atmospheric" parameters that drive DREAM/eta

(4) compare model outputs and iterate with successive EO inputs

(5) measure improvements at each stage and document overall performance improvement

(6) create metadata for resulting model attributes and EO inputs (e.g. measurement units; x,y,z; resolution; temporal frequency; projection; file formats; validity & accuracy; error & error propagation.



The triptych shows three generations of model improvements for a dust storm across New Mexico and Texas on 15-16 December, 2003. (left) the baseline model performance **before** satellite data were included; (middle) **after** satellite data replaced baseline parameters; (right) the same storm modeled by a higher resolution, weather forecasting model.

Issues of resolution for sensing and forecasting dust include: Design parameters of forecast model; Model domain---initial & boundary conditions; Horizontal & vertical resolution of sensors; Detector sensitivity of sensors; Temporal resolution---sensor field-of-view and revisit cycle; Precision and accuracy of collateral ground data; Cold vs. Hot starts; and many more.



This animation illustrates how individuals in different localities can compare dust forecasts. The chart on the right gives a forecast for Wilcox Playa (blue curve) and Silver City (red curve). Each of the locations has a peak hour storm separated by a 2-4 hour time delay. There is a precursor, lower concentration storm at both locations, but Wilcox Playa is eventually hit with 3 episodes over the forecast period. It is interesting that Silver City, at a higher elevation and surrounded by forested terrain, recorded the higher concentration over a longer period than did Wilcox Playa.



APS prefers an AQ alert system that obviates referring to websites. It favors a daily written synopsis of dust and air quality conditions across the APS District. The Asthma Action Plan has three categories of symptoms from mild (green) that are addressed by the classroom teacher; moderate (yellow), that are addressed by a school nurse or other medical provider; and severe (red), that are addressed by emergency response units. A similar three-color scheme could provide the basis for a written brief of the daily air quality forecast that could be sent via email to Asthma Registry nurses, Asthma Allies, ALA, NM, Project ECHO, and the print and broadcast media. In compliance with the ENPHASYS Transition Plan, EDAC proposes to initiate a dialogue to explore this possibility, and at the same time plan on a similar approach for aerosols and ozone from the Community Multiscale Air Quality Model (CMAQ) having 4-km spatial resolution. It is hoped that ozone and aerosol forecasts from CMAQ can be added into EPHTS within 12-15 months.

Using the Wilcox/Silver City example from the previous slide, a synopsis of dust forecasts might look as follows:

<u>Wilcox vicinity</u>: For January 6-8, expect moderate windblown dust late in the evening on the 6<sup>th</sup>, dissipating gradually through the night but increasing and peaking in concentration between 12N and 3PM on the 7<sup>th</sup>. There is a chance for moderate dust between 6 and 8AM on the 8<sup>th</sup>.

<u>Silver City vicinity</u>: For January 6-8, expect conditions as in Wilcox on the 6<sup>th</sup>. For the 7<sup>th</sup>, expect a sharp rise in concentration of dust between noon and 7PM and remaining high until after midnight. Expect diminishing dust on the 8<sup>th</sup>.





When severe weather systems form, or pass over the ENPHASYS model domain, sand and dust are lifted into the atmosphere from thousands of dust sources. Some of these are visible in the image; those that are not so readily visible are pin-pointed by white dots. The number and distribution very likely to change from one storm to another depending on local conditions of soil temperature, moisture content, changing land uses, and other factors.



Because of their small size, unbound surface silt and clay particles are potentially entrainable as dust and can be transported relatively long distance from their source. Larger particles are deposited more quickly. Bi-monthly NDVI dust up-dates add further information to the model about the changing locations of potential dust sources. Depending upon the model grid resolution employed it becomes possible to model the ability of these sources to produce dust. Ginoux's *S function* is a universal "preferential source" approach based on topographic position for detecting areas characterized by fine soil textures that have high dust entrainment potential (Ginoux et al., 2001).



In the United States quantitative sand, silt, and clay fractions are available from NRCS in geospatially accurate USDA County soil survey reports. Assimilating these data into the DREAM/eta model system would almost certainly improve model performance.



## NDVI Processing Steps:

Steps 1 & 2 The cropland class and barren class from the MOD12 land cover classification (classes 12 and 16 respect.) were put in separate files by themselves. They are shown here together with cropland in yellow and barren in red. The cropland class has been compiled from all 12 months of a given year so that it includes fields that were vegetated at one time of year but possibly fallow in another.

<u>Step 3</u> An NDVI product was obtained for July 20, 2008 from the EOS Data Gateway. Now available from WIST (Warehouse Inventory Search Tool - https://wist.echo.nasa.gov/api/).

<u>Step 4</u> By analyzing the irrigated lands and the land cover features around them it was determined that a gray density value of approximately 2500 separated irrigated land from barren. The NDVI image was then modified to show only the 2501 and greater values with the barren values turned to black (0).

<u>Step 5</u> The resulting modified NDVI image was then masked against the cropland class from step 1 & 2 to create a new image which shows cropland only for July shown in yellow.

<u>Step 6</u> This image was then subtracted from the original cropland class (step 1 & 2) and the result in yellow shows cropland in July that was actually fallow. As you can see there is almost no difference between the two images.

<u>Step 7</u>The July fallow class (shown in cyan) is then added to the original barren class (shown in red) from steps 1 & 2 to create a new barren class which is then run through the DREAM model.

<u>Step 8</u> An NDVI product was obtained for February 20, 2008 from the EOS Data Gateway. Now available from WIST (Warehouse Inventory Search Tool - https://wist.echo.nasa.gov/api/). Note the very dark area in the upper right portion of the image. This is snow cover. On its own the DREAM model cannot differentiate between barren ground that is a potential dust source and snow cover which is also "barren" but not a dust source. This will be addressed in the final slides.

<u>Step 9</u> By analyzing the irrigated lands and the land cover features around them it was determined that a gray density value of approximately 2500 separated irrigated land from barren. The NDVI image was then modified to show only the 2501 and greater values with the barren values turned to black (0).

Step 10 The resulting modified NDVI image was then masked against the cropland class from step 1 & 2 to create a new image which shows cropland only for February shown in yellow.

<u>Step 11</u> This image was then subtracted from the original cropland class (step 1 & 2) and the result in yellow shows cropland in February that was actually fallow. As you can see there is a significant difference between the two images compared to the difference in Step 6 which is logical since more fields should be fallow in the winter months than in summer.

Step 12 The February fallow class (shown in cyan) is then added to the original barren class (shown in red) from steps 1 & 2 to create a new barren class which is then run through the DREAM model. The next step was to create another version of this new barren class that could account, in part, for the large snow covered area in the upper right.

<u>Step 13</u> The original February NDVI image was used to create a polygon around the snow cover. This was easily done manually since the snow/no snow line was very distinct. This polygon was then overlaid onto the new barren image created in step 12 and everything inside the polygon was then removed from the barren class. The resulting new barren image can then be run through the DREAM model and the differences the snow cover makes in the output can be evaluated.



The procedures described for identifying variable patterns of dust sources on MCD12, seven sites were selected for monitoring 16-day NDVI changes between Feb/Mar 2008--Jan/Feb 2011. The image shows trends for four categories of change in agricultural cropping systems in seven areas in the Southwest for the period Feb/Mar2008 to Jan/Feb2011. The blue curve shows the percentage of land that was in production from one month to the next. Intuitively this should be highest during the Apr-Oct months in the northern hemisphere. During the winter season, there was a variable amount of land in production depending on cropping system and latitude. The winter wheat belt from Texas to southern Nebraska, and the winter vegetable crops of Southern CA to AZ are among the agricultural systems represented here. The lower amount of land in crops from Nov '09 to Mar '10 compared to the same period in '08-'09 may reflect water shortages or other economic conditions. For dust forecasting, the question is how much do relatively small agricultural fields add to, or diminish, the entrainment of atmospheric and fugitive dust as modeled by either DREAM/eta or CMAQ? Evidence from the data indicate that every 2 weeks approximately one percent (1%) of the grid cells in each NDVI up-date in each category change location in the model domain; but that in cropped areas, higher percentages are observed that add to dust entrainment potential. It is possible that a severe dust storm that overwhelmed the area northeast of Lubbock, TX on 17OCT2011 involved fugitive dust from barren agricultural fields more than atmospheric dust. DREAM/eta simulated light dust in that area, but failed to simulate the high concentrations actually observed.



The blue curve shows the percentage of land that was in production from one month to the next. Intuitively this should be highest during the Apr-Oct months in the northern hemisphere. During the winter season, there was a variable amount of land in production depending on cropping system and latitude. The winter wheat belt from Texas to southern Nebraska, and the winter vegetable crops of Southern CA to AZ are among the agricultural systems represented. The lower amount of land in crops from Nov '09 to Mar '10 compared to the same period in '08-'09 may reflect water shortages or economic conditions, among other explanations. For dust forecasting, the question is how much do relatively small agricultural fields add to, or diminish, the entrainment of atmospheric and fugitive dust as modeled by either DREAM/eta or CMAQ? Evidence from the data indicate that every 2 weeks approximately one percent (1%) of the grid cells in each NDVI up-date in each category change location in the model domain; but that in cropped areas, higher percentages are observed that add to dust entrainment potential. It is possible that a severe dust storm that overwhelmed the area northeast of Lubbock, TX on 17OCT2011 involved fugitive dust from barren agricultural fields more than atmospheric dust. DREAM/eta simulated light dust in that area, but failed to simulate the high concentrations actually observed.

The length of the growing season is longer for the ABQ-Estancia area and shorter for the San Luis and Llano Estacado areas. The percent of month-to-month changes in crop-to-crop (C to C) land reaches 80+% for ABQ-Estancia, and nearly 100% for San Luis and Llano Estacado. As one would expect, during this time period the percentage change for no crop-to-no crop (NC to NC) is at its lowest (around 14% for ABQ-Estancia; near 0% for San Luis and Llano Estacado). For these areas the highest percent of no crop-to-crop (NC to C) change in agricultural practices occur in spring as new crops are being planted (~40% for ABQ-Estancia and San Luis; 26% for Llano Estacado); then decline to around 0% during the growing season. Crop- to-no crop (C to NC) changes are practiced during the winter months when fewer crops are being cultivated; about 30% for ABQ-Estancia and San Luis, and 41% for Llano Estacado. These cropping patterns suggest that the potential for fugitive dust is during September through March.



Clearly, model grid cell size influences greatly the recorded dust loads. MCD12 NDVI values are recorded on a 1km grid, but modeling can be executed at a variety of grid cell sizes. The above graphic illustrates three modelling grids of dust concentrations ( $\mu$ g/m<sup>3</sup>). Top, left = coarse resolution grid (0.3°, 50km) showing patterns for 4-bins of particle sizes; top, right = the same model grid with 8 particle size bins; bottom, left = a high resolution grid (0.05°, 7.5km) with 8 particle sizes. Bottom, right shows the effective radii for both 4 and 8 bin models. The derivation of PM<sub>2.5</sub> and PM<sub>10</sub> is given for the 4-bin version and PM<sub>2.5</sub> for the higher resolution 8-bin version.



Hypothetical area showing the complex association of urban and/or agricultural land uses with varying anthropogenic (fugitive) dust and fine particulate levels, together with the over-print of an atmospheric dust cloud moving through the region. Both the PHAiRS-modified version of DREAM/eta and the CMAQ model working in tandem are necessary to better monitor dust dynamics.

Simulating both windblown dust and anthropogenic air pollution events is intensive computationally in dusty regions. The diagram at right is a schematic of that complexity. The diagram is a hypothetical region within a model domain that has a low level of ambient dust and aerosols. In this region there are five AIRNow monitors (circles) and three AERONET Stations (stars). Each of these sites has a shade of gray indicating its observed concentration of dust and aerosol measured at a specific time, as contributed by its respective local population and economic activities. Each of the population centers (irregular shapes) has a shade of gray to indicate its ambient level of anthropogenic dust and aerosols at the moment of measurement. A dust cloud (pale brown) is shown moving toward the northeast that will add an atmospheric dust component to the ambient anthropogenic signal. The population center in the upper left of the domain shows a particularly heavy concentration of atmospheric dust mixed with the ambient anthropogenic dust/aerosol load. One can imagine that levels of air contaminants will rise in the population center near the center of the image as the dust cloud passes through. Dust concentrations modeled by DREAM/eta can therefore consist of three different phenomena: atmospheric dust only; the ambient level of urban dust only; or a combination of both. To separate the atmospheric from the ambient concentrations, one must know the dust loads of each. DREAM/eta supplies the atmospheric component; CMAQ provides the anthropogenic component. It is therefore desirable to evolve an efficient model system that combines both.



WRF model simulations were conducted using USGS and NLCD datasets. The model configuration employed in these simulations was arrived at through a cursory literature review of recent modeling projects undertaken in support of air quality management activities. Future studies must investigate the impact of LULC classification on dust generation. Daily performance statistics for the 4-km modeling simulation using USGS and NLCD datasets.





Simulations were conducted at 36, 12 and 4-km grid resolution using the 24-category USGS and 2001 National Land Cover Dataset (NLCD). The performance of the system was evaluated against observations at National Weather Service (NWS) stations. The simulated meteorological fields were first processed using the Meteorology Chemistry Interface Processor (MCIP), and then used to process emissions from mobile, area, non-road, electricity and non-electricity generating units, as well as biogenic emissions. The Sparse Matrix Operator Kernel Emissions (SMOKE) processor was used. Air quality model simulations were performed using CMAQ v4.6. Simulated model concentrations were compared against observations recorded at monitoring stations operated by US/EPA.



CMAQ is a Eulerian air quality model that solves the discretized form of the Advection-Diffusion Equation. Meteorological fields for CMAQ are generally obtained from dynamical data assimilating meteorological models (also referred to as mesoscale models). A processor called Meteorology Chemistry Interface Processor (MCIP) was used to create input files for CMAQ. Its main function is to read meteorological fields simulated by a mesoscale model, compute dry deposition velocities and other variables that CMAQ needs but are not available from the meteorological model, and output data in Models-3 IOAPI format. MCIP is capable of processing meteorological fields simulated by WRF.

Daily average, gridded emissions processed of PM<sub>10</sub> (left panel), PM<sub>2.5</sub> (center panel), and PMFINE (right panel) using EPA's 2002 Emissions inventory (top) with Source Classification Codes (CCS); and bottom panel without SCCs (bottom). This process aids in identifying localities generating fugitive dust

CMAQ science modules used in ENPHASYS		
Physical process	Module name	Description
Horizontal & vertical advection	HYAMO and VYAMO	Global mass-conserving scheme of Yamartino to calculate horizontal & vertical advection
Horizontal diffusion	MULTISCALE	Use diffusion coefficient based on local wind deformation
Vertical diffusion	ACM2	Calculate vertical diffusion using the Asymmetric Convective Model version 2
Gas-phase chemical mechanism & solver	EBI_CB05CL	Use the Euler Backward Iterative solver optimized for the Carbon Bond-05 mechanism with air toxics, mercury, and chlorine
Aerosol chemistry	AERO4	Fourth-generation modal CMAQ aerosol model with extensions for sea salt emissions and thermodynamics
Gas and aqueous- phase chemical mechanism	CB05CL_AE4_AQ	Carbon Bond-05 mechanism with Chlorine, with fourth-generation modal CMAQ aerosol model with extensions for sea salt emissions and thermodynamics
Dry deposition	AERO_DEPV2	Use second-generation CMAQ aerosol deposition velocity routine
Cloud dynamics	CLOUD_ACM	Use ACM cloud processor that uses the ACM methodology to compute convective mixing

Model predictions were compared against recorded observations with the help of a software tool METSTAT, developed by ENVIRON corporation. This tool computes surface statistics for temperature, wind speed, wind direction, and humidity. The metrics include: Bias Error (B), Gross Error (E) and Root Mean Square Error (RMSE), Systematic Root Mean Square Error (RMSE<sub>s</sub>), Unsystematic Root Mean Square Error (RMSE<sub>u</sub>) and Index Of Agreement (IOA).



Caution: Summary statistics are useful in making a general assessment about the adequacy of meteorological fields. For example, daily-mean performance statistics are likely to conceal important hour-to-hour variations. Also, summary statistics depend upon the number of observation-prediction pairs and generally improve with larger sampling sizes and longer averaging periods. This is because the probability of statistics being affected by extreme values is high in smaller sample sizes. These and other concerns have lead US/EPA to recommend that benchmarks should not be used in a "pass/fail" model, but only as a means of assessing general confidence in meteorological model output.



CMAQ is installed on the EDAC system and has been tested by running a 2-day simulation over the 4-km grid. WRF will be installed on the system in October as well. This WRF-CMAQ system will be used for retrospective simulations of past dust events. It could be initialized with dust concentrations from DREAM/eta model, different soil moisture datasets, and NLCD land cover. Our initial intention was to initialize a CMAQ simulation using dust concentrations from the DREAM/eta model. However, the DREAM/eta simulation for Feb 2006 was never conducted; so, when we compare IMPROVE data (observed), we cannot answer the question on how much of the "modeled" pollution is natural versus anthropogenic.



CALIOP data provide total attenuated backscatter (TAB) at 532 nm. Data from these overpasses can be mapped onto a 17 layer CMAQ grid over the US for selected dates. The data cover a ground swath of essentially 0- meters between 60°N latitude (left side) to 8°N latitude (right side) and 90°W longitude (left side) to 104°W longitude (right side). The narrow swath means that data represent an instantaneous vertical "snap-shot" of the atmospheric layers (including aerosols), but the very low temporal resolution (revisit time), obviates tracking their concentrations or movement across landscapes. The map below shows the geographic coordinates on this descending node of CALIPSO.





This vertical feature mask shows widespread aerosol along track from latitude 33.36°N to minus 15° latitude (left side) to 15.43°S latitude and along longitude -96.95°W to -107.73°W. The image below follows the track of the CALIOP sensor. Data represent atmospheric conditions from -5km to +30 km. Total attenuated backscatter (TAB) values range from  $1\times10^{-4}$  (bottom, deep blue) through  $1\times10^{-3}$  to  $6.0\times10^{-3}$  (greens through reds). For quick interpretation, these data are often processed into a "vertical feature mask", as shown below. The vertical bar on the right shows seven categories: 1 = blue (clear air); 2L = red (low/no confidence cloud); 2 = light blue (cloud); 3L = brown (low/no confidence aerosol); 3 = orange (aerosol); 4 = yellow (stratospheric layer); 5 = green (surface); 6 = gray (subsurface); 7 = totally attenuated). Note the "plume" of aerosols (orange) rising over a mountain range and fumigating a large area south/southwest across the Sierra Madre Oriental into the valley of Mexico and beyond to the Pacific Ocean.





ENPHASYS has interlocking elements each having unique processing steps to generate products needed by the New Mexico Environmental Public Health Tracking System (EPHTS), part of the evolving Environmental Public Health Tracking Network (EPHTN) being developed by the National Centers for Disease Control & Prevention (CDC). The objective of ENPHASYS was to engineer NASA's Earth science results so that they contribute essential inputs to specific health tracking needs. The needs addressed by ENPHASYS are dust concentrations, both atmospheric and anthropogenic (fugitive) dust, speciated aerosols (specifically NO<sub>2</sub>, SO<sub>2</sub>, NO<sub>3</sub>, CO<sub>2</sub>, and ozone (O<sub>3</sub>). Such engineering recognizes that each input has individual image processing requirements employing a suite of software packages (in this case DREAM/eta and CMAQ), and that outputs all have associated attribute uncertainties. Each output product must be documented using metadata standards so that data sets can be discovered, retrieved, interpreted, and converted into statistical data or visualizations. These outputs could then be used with confidence by clinicians to forecast near-term air quality conditions, or by epidemiologists to assess longer-term health outcomes in a population subject to chronic exposure.