INTEGRATING EARTH OBSERVATIONS DATA INTO GEOSPATIAL DATABASES THAT SUPPORT PUBLIC HEALTH DECISIONS

S.A. Morain and A.M. Budge

Earth Data Analysis Center, MSC01-1110, University of New Mexico, Albuquerque, New Mexico, USA, 87131-0001
(smorain, abudge)@edac.unm.edu

Commission VIII, WG VIII/2

KEY WORDS: Earth observation, System integration, Internet GIS, Public health/medical, Environmental monitoring, Spatio-temporal modelling, Spatial data modelling, decision support

ABSTRACT

There are several reasons why public health communities do not use information from Earth observations routinely. Most notable among them are: (1) they need science results that verify, validate, and benchmark the statistical and economic benefits from these exotic inputs; and, (2) they lack the systems that would deliver such reliable information economically and swiftly in their already heavy workloads. The Public Health Applications in Remote Sensing (PHAiRS) project is engineering elements of an enhanced disease surveillance system for dust-related respiratory diseases in the desert southwest of the United States. Several Earth observations (EO) data sets are replacing parameters traditionally used in dust forecasting models to improve simulations of particulate matter entrainment, timing of entrainment, concentrations, and subsequent movement. Output from the enhanced dust forecasting model is nested within a regional version of the National Centers for Environmental Prediction (Eta version) model (NCEP/Eta). Simulations using the enhanced dust model have been compared to actual dust episodes recorded by NCEP/Eta's atmospheric patterns and by dust data from a ground-based Continuous Air Monitoring System (CAMS). The simulations were re-run after replacing the land cover layer with land cover classes derived from MODIS (MOD-12). For the CAMS test cases, this resulted in a significant improvement in dust episode patterns. Additional EO data assimilations investigate whether further improvements can be gained by replacing the topographic layer with higher resolution digital elevation data from SRTM, dust-generating areas derived from MOD-15's Fraction of Photosynthetically Active Radiation (FPAR) and Leaf Area Index (LAI) data, soil moisture data from AMSR-E, and aerodynamic surface roughness. Ongoing simulations aim eventually to measure hourly, daily, and weekly model improvements from EO data replacements that are refreshed on a weekly, seasonal, or inter-annual basis. The overall aims are to: (a) combine the measured improvements from several EO data series that optimize dust forecast scenarios for public health authorities; (b) benchmark each step in the process to document the benefits of EO data inputs into respiratory health care; and (c), develop retrospective and forecast statistics from model runs that boost system reliability and user confidence.

1. INTRODUCTION

Medical communities, governments, and societal best interests, demand new disease surveillance systems that link health care reporting at the individual and community levels (disease detection and transmission) with robust relational databases (socioeconomics, demographics), and with global and regional Earth observations data (weather forecasts, disaster monitoring). Modern medical diagnoses, treatment, reporting, and containment in the face of potentially devastating emerging infectious diseases like avian flu (H5N1), require readily available historical information, as well as forecasts of future conditions that might accelerate an epidemic. There is widespread interest in disease surveillance systems covering a wide array of communicable diseases and medical conditions that allow health care professionals to query databases for similar cases reported locally, regionally, and globally. These new systems represent the enabling technology for and protecting, the lives of people, and thereby the economies and resources of nations.

1.1 Global Agenda

No emerging infectious diseases have received more attention from governments, international bodies, or the medical and scientific communities than SARS and Avian Flu. They have the ability to become global epidemics at lightening speeds. Other diseases like malaria, polio, and HIV/AIDS are of immense concern; and relative new-comers like West Nile Virus are receiving close scrutiny. What they all have in common is that their transmission is related to both social and environmental factors, some of which are well known or suspected, but few of which, until recently, have figured prominently in medical planning, preparedness, or mitigation.

As evidence for global climate change mounts, and as the global economic and medical consequences of natural disasters become better understood, it is clearly in the best interests of humanity to share health information in such ways as to connect even the most remote areas of the world. It is now commonly accepted among people everywhere that enormous societal benefits will accrue from the combined capabilities of remote sensing, geographic information systems, and position, navigation, and timing (PNT) technologies, especially those applied to disease surveillance. Toward this end, the Group on Earth Observation (GEO) has been joined by many concerned nations, international bodies, and associated professional societies to serve as a nexus for addressing these common needs in context of geospatial analysis and Earth observations. These needs encompass, among many others:

- vector- and raster-based technology;
- system and data interoperability;
- surveys of user needs for geospatial data

- mechanisms for more rapid data and information sharing; and,
- plans for disaster and public health response.

1.2 PHAiRS

Most of the rich literature describing the roles and benefits of EO data in public health is based on anecdotal inferences derived from traditional image interpretation. Public health communities cannot rely on this evidence because they need science results that verify, validate, and benchmark the statistical and economic benefits from these exotic inputs; and because they lack surveillance systems that can deliver reliable information economically and swiftly. An effective infrastructure for data discovery and sharing does not yet exist.

To begin developing an infrastructure, the *Public Health Applications in Remote Sensing* (PHAiRS) project is engineering an integrated vector- and raster-based system that can be accessed as a web-based service by health care professionals. It is initially designed for the desert region of the American southwest where several respiratory conditions and diseases occur. PHAiRS has three parallel thrusts: (1) assimilate EO data into an atmospheric dust model; (2) iterate model inputs to optimize model outputs; and, (3) undertake a beta-test program with public health authorities to assess relationships between dust episodes and increased respiratory complaints.

Project documentation and results to date can be found in Morain and Budge, 2006; Morain, 2006; Budge et al., 2006; Morain and Sprigg, 2005. This paper describes the general strategy for geospatial and EO data assimilation, and the experimental design for assessing quality of model outputs.

2. DATA ASSIMILATION

2.1 Definitions

Data assimilation and data fusion are different processes for integrating or enhancing geospatial databases with EO data. Provisional definitions for the PHAiRS project are:

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

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

2.2 Models

Most models available for Earth system science, are not designed for EO data. Compatibility issues arise, among which are: (a) measurement units, (b) x,y,z,t resolution, (c) map projection and ease of re-projection to fit model requirements, (d) file formats, (e) error and error propagation, and (f) validity of the data set as a replacement input. These issues must be reconciled before assimilation of each candidate data set can be performed, and before iterations of multiple data sets can be statistically evaluated. It is assumed that overall improvements in model outputs will result from accumulated incremental improvements from each assimilated data set (Morain, 2006; Morain and Budge, 2006).

PHAiRS employs the National Centers for Environmental Prediction (Eta version) regional weather forecasting model (NCEP/Eta) for mapping weather events in America's desert southwest. Within this model, the project has nested the Dust Regional Atmospheric Model (DREAM) developed by Nickovic et al. (2001), Janjic (1984), Mesinger et al. (1988) and Janjic (1994). Both models have been adapted for use in the region, and their performance has been tested and validated using observed weather patterns and dust events (Morain, 2006, Morain and Sprigg, 2005; and Yin et al., 2005). DREAM is a desert dust cycle model consisting of two modules: (a) an atmospheric simulator; and (b) a dust cycle simulator. The atmospheric simulator parameters include land surface processes, turbulent mixing, convection, large-scale precipitation, lateral diffusion, and radiation.

The dust cycle module simulates dust production, advection and turbulent diffusion, and dry and wet deposition (Nickovic et al., 2001; Shao et al., 1993; Georgi, 1986). The module consists of three static surface parameters.

- soil texture classes at $2'$x$2'$ resolution (Cosby et al., (1984);
- vegetation cover at $10'$ resolution (Olson World Ecosystems); and,
- elevation at 1x1 km resolution (GTOPO30).

Two of these, land cover and elevation, are candidates for replacement by EO data products assimilated into DREAM

2.3 Process

The disease surveillance system in PHAiRS is designed for the stated needs of epidemiologists, school nurses, doctors, and veterinarians, among others (Figure 1). The objective is to augment traditional medical data and information (health questions) with web-based services that provide a geographical and environmental context for the broader implications of reported cases that might otherwise be static, individual case records. PHAiRS is modelling primarily mineral dust and respiratory diseases, but water-borne and vector-borne diseases like malaria and typhoid can be modelled also. In concept, the web-based service system would not only provide doctors and clinicians with a rapid response capability at the case level; but, would also provide public health authorities with longer range forecast capabilities that protect the public at large.

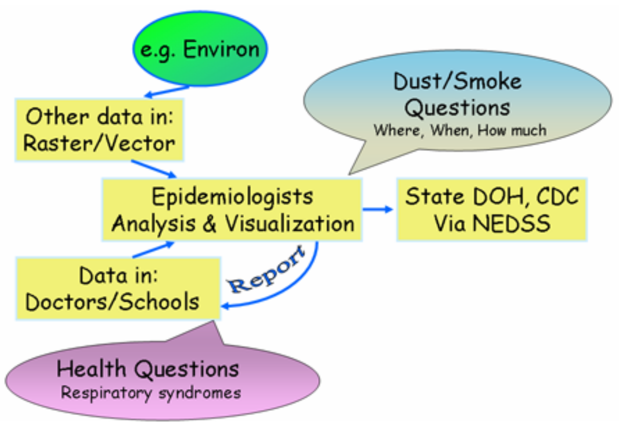


Figure 1. Disease surveillance system concept

The process is familiar to geospatial analysts. Conceptually, the aim is to select candidate parameters from the array of data types that drive DREAM and to physically replace them with higher spatial and temporal resolution data derived from space sensors (Figure 2). The goals are to:

- replace selected trays in the rack with regularly refreshed EO digital data from the "terrain," "surface conditions," and "atmospheric" parameters that drive DREAM;
- improve model output without altering the validity of the model's original function; and,
- convert the model to a more dynamic forecast.

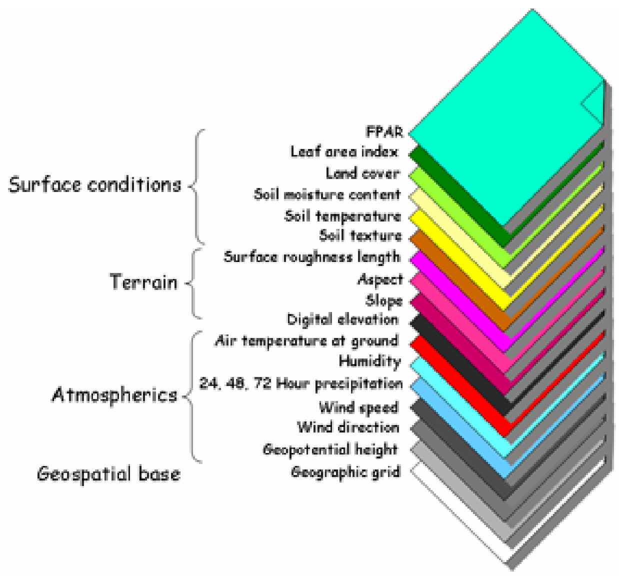


Figure 2. Concept for data assimilation into models

2.4 Sample Model Runs

The process just described has challenges aside from the data quality and compatibility issues described in Sec. 2.2. Included are: (a) how to measure small incremental model improvements using single parameter replacements; and, (b) how to measure improvements using sequences of model parameters. Table 1 shows a series of model runs using different sets of assimilated data. Replacement parameters were: MOD-12 (Terra MODIS-12), 1km spatial resolution land cover; SRTM (Shuttle Radar Topography Mission), 1km spatial resolution elevation; NASA z_0, aerodynamic surface roughness (Blonski et al., 2005); FPAR (Fraction of Photosynthetically Active Radiation), class 253, 1km spatial resolution barren or sparsely vegetated; and AMSR-E (Advanced Microwave Sounding Radiometer-E), 0.26 degree spatial resolution soil moisture.

Run #	MOD-12	SRTM	NASA z_0	FPAR	AMSR-E
Baseline	No assimilated EO data				
Run 2c	✓				
Run 4a	✓	✓			
Run 5a	✓	✓	✓		
Run 5b	✓	✓	✓		
Run 6a	✓			✓	
Run 10a	✓	✓	✓		✓
Run 15a	✓				✓

Table 1: NCEP/ETA-DREAM model runs

Run 10a (***bold/italic***) was used to generate the statistics given in Table 2. Their graphic representation is given in Appendix A. Table 2 compares three parameters in DREAM before and after EO data assimilation. 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 higher index values improve the ability of the model to forecast dust entrainment.

Metrics	Wind Speed (m/s)	Wind Dir (°)	Temp (K)	Definition
Mean Obs.	5.53	231.40	276.74	$\frac{1}{N}\sum_{i=1}^{N} O_i$
Mean Mod.	*4.65* 4.37	*226.60* 230.38	*275.56* 277.48	$\frac{1}{N}\sum_{i=1}^{N} M_i$
Mean Bias	*-0.88* -1.16	*-4.80* -1.02	*-1.20* 0.72	$\frac{1}{N}\sum_{i=1}^{N} (M_i - O_i)$
Mean Error	*1.97* 2.03	*51.76* 47.85	*4.09* 2.67	$\frac{1}{N}\sum_{i=1}^{N} \lvert M_i - O_i \rvert$
Agreement Index	*0.74* 0.75	*0.74* 0.76	*0.71* 0.95	$1 - \frac{\sum_{i=1}^{N} (M_i - O_i)^2}{\sum_{i=1}^{N} \left(\lvert M_i - \bar{O} \rvert + \lvert O_i - \bar{O} \rvert \right)}$

Table 2. DREAM performance before and after EO data assimilation; values in italic are before EO data assimilation; other values are after assimilation. For the equations M = modeled; O = observed

Appendix A is a series of charts comparing agreement indices for surface wind speeds, wind directions, temperatures, PM_{10} concentrations, and $PM_{2.5}$ concentrations as modelled for two dust storms over New Mexico and Arizona on December 8-10 and December 15-16, 2003. The charts show for example, that *run10a* had similar model performances in surface wind speed, wind direction, and temperature as the other model runs having one or more NASA sensor data sets included. For case 2, *run10a* (which included all of the assimilated EO data products except FPAR) showed better performance in surface wind than *run6a* or *run15a*. However, *run10a* did not show better performance in PM_{10} and $PM_{2.5}$ compared to other model runs that used fewer EO data products. The $PM_{2.5}$ model performance of *run10a* was not as good as *run2c*, which only replaced the static land cover parameter with a higher resolution MOD-12 product. It is apparent that non-linear interactions of data sets with model performance exist. Consequently, a thorough experimental design is required to understand the interactions.

3. EXPERIMENTAL DESIGN

PHAiRS has adopted an experimental design to address non-linearities and to assess quantitatively its model runs. Among the challenges in this design are that:

- newer versions of EO data products are constantly being released, most without adequate metadata to assess the impacts of their use in a given model, or to determine whether the older version should be retained;
- models, themselves, are being upgraded and refined as progress in data inputs improve their performance;
- new sensor and platform technologies lead to previously unavailable data sets whose candidacy for assimilation must be cycled in.

To date, PHAiRS has focused on assessing six EO data products as candidates for assimilation into NCEP/Eta-DREAM. These are MOD 12 (land cover); SRTM-30 (elevation); AMSR-E (soil moisture); aerodynamic surface roughness (derived from MOD12); MOD-11 (soil temperature); and AMSU-A (humidity). When these products are adequately prepared for assimilation, they will be run sequentially in non-duplicating combinations to measure their affect on the performance of the nested model. This will require six sets of model runs as follows

- Iteration I: 6 model runs, 1 parameter at a time;
- Iteration II: 15 runs, 2 parameters at a time;
- Iteration III: 10 runs, 3 parameters at a time;
- Iteration IV: 6 runs, 4 parameters at a time;
- Iteration V: 3 runs, 5 parameters at a time; and,
- Iteration VI: 1 run, all six parameters.

4. CONCLUSIONS

The need for geospatial data in disease surveillance systems has been established at the highest levels in public health communities. There is growing interest also in linking EO data and products to these systems as a means for up-dating local, regional, and global environmental conditions, especially short-lived phenomena that bear directly on specific disease transmission mechanisms. Research results are making progress in developing web-based services that augment rapid data and information sharing, but it will take additional research effort to bring these capabilities to operational status.

It is necessary to convince public health authorities that disease surveillance systems enhanced with geospatial data provide useful, accurate, and verifiable information. Without statistical evidence and validation, information derived from relational databases, including those employing satellite observations, is still too novel to warrant more than research interest. Practicing medical and health professionals are still reluctant to make decisions on the basis of geospatial technologies because the risks of making bad decisions currently out-weigh any perceived benefits. Public health organizations will more than likely need to add qualified information technologists to their staff to take full advantage of emerging digital communication systems.

5. REFERENCES

Blonski, S., D. Holland, and V. Henderson, 2005. NASA soil moisture data products and their incorporation in DREAM. Contractor Report to UNM Earth Data Analysis Center.

Cosby, B.J., G.M. Hornberger, R.B. Clapp, T.R. Ginn, 1984. A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resources Research,* 20 pp. 682-690.

Budge, A.M., K.K. Benedict, and W. Hudspeth, 2006. Developing Web-Based Mapping Services for Public Health Applications. In: The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences, Goa, India, Vol XXXVI, Part 4, pp. ??-??.

Georgi, F., 1986. A particle dry-deposition parameterization scheme for use in tracer transport models. *JGR*, 91 pp. 9794-9806.

Janjic, Z. I., 1984. Non-linear advection schemes and energy cascade on semi-staggered grids. *Monthly Weather Review*, 118, pp. 1234-1245.

Janjic, Z. I., 1994. The Step-mountain Coordinate Model: Further Developments of the Convection, Viscous Sublayer and Turbulence Closure Schemes. *Monthly Weather Review*, 122. pp. 927-945.

Mesinger, F., Z.I. Janjic, S. Nickovic, D. Gavrilov and D.G. Deaven, 1988. The Step-mountain Coordinate: Model Description and Performance for Cases of Apline Lee Cyclogenesis and for a Case of an Appalachian Redevelopment. *Monthly Weather Review.* 116. pp. 1493-1518.

Morain, S., 2006. Replacing model parameters with Earth observation data to improve atmospheric dust forecasts and public health responses. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Haifa, Israel, Vol. XXXVI, Part ??, pp. ??-??

Morain, S. and A.M. Budge, 2006. Science data products for public health decision support. In: *IGARSS, 2006*, Denver, CO, July 30-Aug.2.

Morain, S. and W. Sprigg, 2005. Initial benchmark report for public health. NASA Cooperative Agreement NNS04AA19A. Sep. 30. 36 pages.

Nickovic, S., G. Kallos, A. Papadopoulos, and O. Kakaliagou, 2001. A Model for Prediction of Desert Dust Cycle in the Atmosphere," *JGR*, 106(D16). pp. 18,113-18,119.

Shao, Y., M.R. Raupach and P.A. Findlater (1993), "Effect of Saltation Bombardment on the Entrainment of Dust by Wind," Journal of Geophysical Research, 98: 12719-12726.

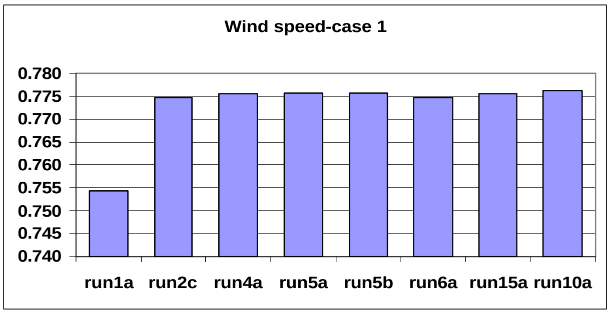
Yin, D., S. Nickovic, B. Barbaris, B. Chandy and W. Sprigg, 2005. Modeling Wind-blown Desert Dust in the Southwestern United States for Public Health Warning: a Case Study. *Atmospheric Environment*, 39 pp. 6243-6254.

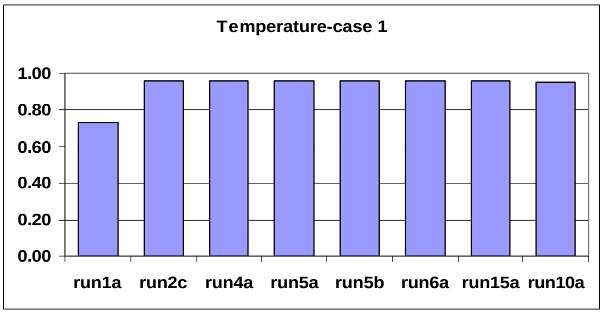
6. ACKNOWLEDGMENTS

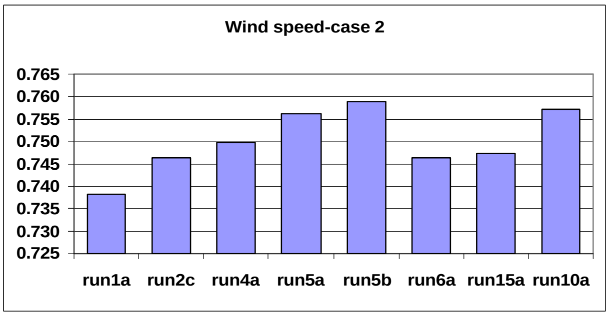
The authors thank all PHAiRS team members for contributing materials that made this paper possible. These team members, in alphabetical order by affiliation, are Karl Benedict, Thomas Budge, Gary Sanchez, and William Hudspeth, all from the University of New Mexico; and William Sprigg, Brian Barbaris, Christopher Cattrall, Beena Chandy, and Dazhong Yin, all from the University of Arizona.

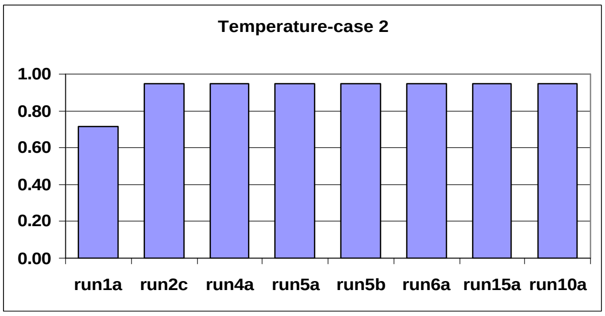
This research was funded by the U.S. National Aeronautics and Space Administration under Agreement NNSO4AA19A.

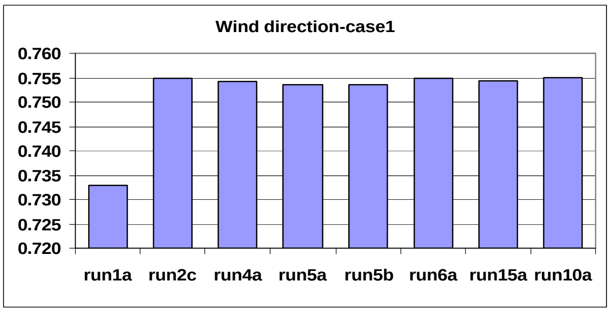
7. APPENDIX A: GRAPHIC RESULTS OF MODEL RUNS FOR TWO OBSERVED DUST STORMS

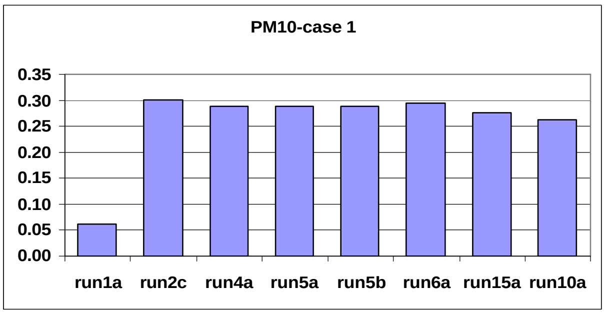

Wind speed-case 1
0.780
0.775
0.770
0.765
0.760
0.755
0.750
0.745
0.740
run1a run2c run4a run5a run5b run6a run15a run10a

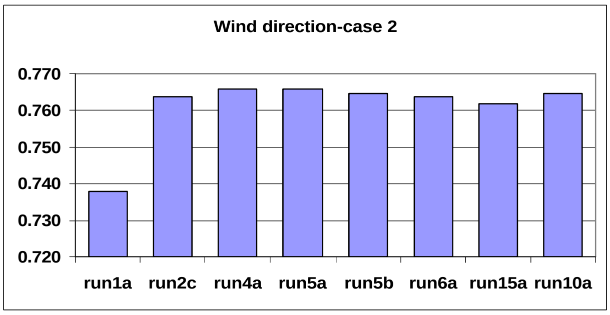

Temperature-case 1
1.00
0.80
0.60
0.40
0.20
0.00
run1a run2c run4a run5a run5b run6a run15a run10a

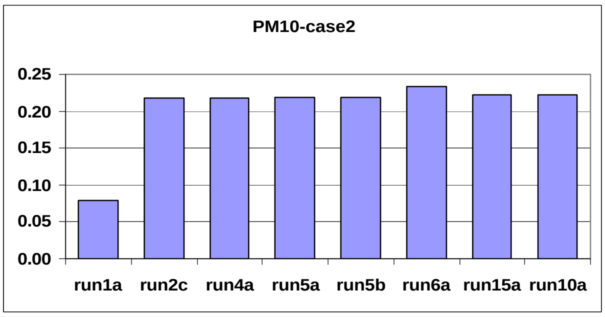

Wind speed-case 2
0.765
0.760
0.755
0.750
0.745
0.740
0.735
0.730
0.725
run1a run2c run4a run5a run5b run6a run15a run10a

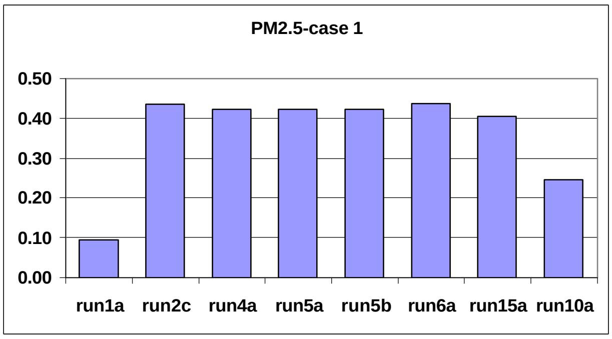

Temperature-case 2
1.00
0.80
0.60
0.40
0.20
0.00
run1a run2c run4a run5a run5b run6a run15a run10a

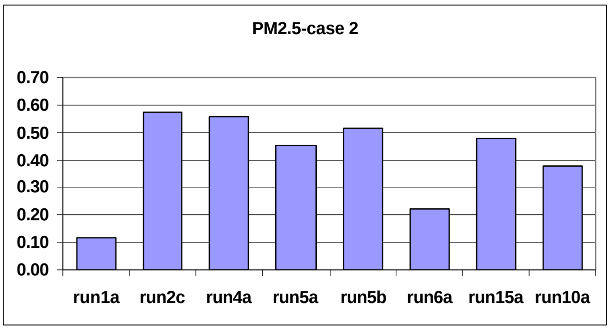

Wind direction-case1
0.760
0.755
0.750
0.745
0.740
0.735
0.730
0.725
0.720
run1a run2c run4a run5a run5b run6a run15a run10a


PM10-case 1
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
run1a run2c run4a run5a run5b run6a run15a run10a


Wind direction-case 2
0.770
0.760
0.750
0.740
0.730
0.720
run1a run2c run4a run5a run5b run6a run15a run10a


PM10-case2
0.25
0.20
0.15
0.10
0.05
0.00
run1a run2c run4a run5a run5b run6a run15a run10a


PM2.5-case 1
0.50
0.40
0.30
0.20
0.10
0.00
run1a run2c run4a run5a run5b run6a run15a run10a


PM2.5-case 2
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
run1a run2c run4a run5a run5b run6a run15a run10a