Atmospheric Sciences and Informatics
EarthCube Driver Whitepaper: Use Cases

Beth Plale[1], Rich Clark[2], Craig Mattocks[3], Keith Brewster[4], Rebecca Barthelmie[5], Kelvin Droegemeier[6], Dennis Gannon[7], Sara Graves[8], Scott Jensen[1], William Mahoney[9], Sara Pryor[5], Rahul Ramachandran[8], Robert McDonald[10], Mohan Ramamurthy[11], Bob Wilhelmson[12], Ming Xue[4], Sepi Yalda[2]

[1] School of Informatics and Computing and Data To Insight Center, Indiana University
[2] Dept. of Earth Science, Millersville University
[3] Rosenstiel School of Marine & Atmospheric Science, University of Miami
[4] Center for Analysis and Prediction of Storms (CAPS), University of Oklahoma
[5] Dept. of Geography, Indiana University
[6] School of Meteorology, University of Oklahoma
[7] Microsoft Research
[8] Information Technology and Systems Center (ITSC), University of Alabama
[9] National Center for Atmospheric Research (NCAR)
[10] Indiana University Libraries and Data To Insight Center, Indiana University
[11] Unidata
[12] Dept of Atmospheric Science, University of Illinois and NCSA

1. Introduction

In response to the EarthCube call for transformative concepts and approaches to create the next generation of integrated data management infrastructures across the Geosciences, the above team of atmospheric, climate, emergency management, transportation, and informatics researchers has assembled a set of science scenarios that are described in this document. *We have also described a technical infrastructure, and this appears in an accompanying whitepaper by the same set of authors.* The authors draw in part from the NSF large ITR Linked Environment for Atmospheric Discovery (LEAD) (2003-2009)(Droegemeier et al. 2005). LEAD was one of the pioneers of the Science Gateway, a portal serving a community of researchers and educators that was one of the first to bring high performance computing resources into the hands of users in an on-demand way. The science gateway concept became so successful that in 2010, it supported 30% of all TeraGrid users[1]. While the LEAD cyberinfrastructure was very successful in demonstrating key advances in technology, it is not being proposed as a solution here. This whitepaper does not advocate specific technologies per se. It represents new thinking, new people, new research questions drawn from our vast experience in data driven science, largely in the atmospheric sciences, but relevant to other geosciences as well.

2. EarthCube Use Case Scenarios

Hurricane Preparedness. It is the year 2011. Hurricane Irene (Irene 2011) is forecast to make landfall along the east coast of the United States and track northward along the coastline. The potential for heavy rainfall amounts and devastating flooding is 48 hours away. Emergency response is already in effect and new smart phone applications allow individual users to view evacuation routes and traffic density. But the applications are static, i.e. they are not informed by

---

[1] Personal correspondence, 2010.

in-situ observations or updated forecasts of weather, coastal water levels, or road weather conditions.

It is the year 2014. EarthCube has been deployed. Hurricane Isabella is forecast to make landfall along the east coast of the United States and track northward along the coastline. The potential for heavy rainfall amounts and devastating flooding is 48 hours away. Emergency response is already in effect. Individual smart phone users can invoke an application that will provide up-to-the-minute in-situ guidance for emergency evacuation. Invoking the application sends a query to a cloud-hosted service that initiates a workflow on the user's behalf to provide the best possible evacuation route as a function of time from 48 hours before to well after the hurricane has made landfall. The workflow uses a spider to crawl relevant data sources and retrieve pertinent data including historical evacuation data from an national emergency management repository, surface data from automated surface observing systems (ASOS), in-situ road condition observations from IntelliDrive systems (Intellidrive 2011) that are collected and curated by state-wide road weather information systems (RWIS 2011), watershed and flooding data from the national hydrological database, and numerical weather prediction, hurricane prediction, storm surge, and rainfall/runoff models, (e.g. WRF, SLOSH, Nat'l Weather Service River Forecast System (NWS-RFS)). Wind forecasts combined with power grid applications predict areas of likely power outages. Some of this information may be coalesced by cloud-based data aggregators. The application workflow creates a decision tree based on the forecasts and provides the user with suite of possible evacuation routes and nearby destinations likely to have power, fuel and services. As the hurricane approaches, updated forecasts and observations dynamically prune the decision tree of possible evacuation routes, considering storm track probabilities, mile-marker specific road conditions, traffic congestion, alternative routes, and whether contraflow (traffic lane reversal to aid in evacuation) should be activated.

Energy-Environment-Economy-Education Application. The City of Baltimore's Office of Sustainability (Baltimore 2011) has hired a team of scientists, analysts, and urban planners and policy-makers to scope out Baltimore in 2025, an imperative that will enhance the economy, create jobs, and ensure urban sustainability by creating a clean and green transportation system and architectural infrastructure, while mitigating the impact of climate change on the environment. Everything is on the table. The team maps out the imperative in terms of energy, environment, economy, and education. To lower the energy footprint of the city, the team considers the fractional area of vegetated roofs that will be needed to reduce the sensible heat flux by 10 percent and the storm water runoff by 25 percent. A recharging network will be installed for electric vehicles (EVs), curbside pay-to-ride rental EVs will be available for short distance commutes and taxis and mass transit will be required to transition to EV technology over the 15-year project timeline. The stretch goal is to have 50 percent of all vehicles in Baltimore be fully electric or hybrids. Urban green areas will be expanded, including green ribbons along urban corridors to reduce the city's heat footprint and increase CO_2 uptake. Solar panels and high albedo roofing materials and vegetative cover will be installed on the rooftops of public buildings and tax incentives will be available to encourage business, industry, and private citizens to follow suit. Special attention is given to protecting the Inner Harbor, the Port of Baltimore, and other economic strongholds from unwanted sea level encroachment as a consequence of climate change. The team is charged with developing a roadmap to 2025 with multiple pathways and outcomes that are reasonable and can come to fruition at a cost that is manageable within the projected revenue stream.

Tangible Geospatial Modeling combined with GIS analysis tools (such as GRASS) will be used to import and merge multiple layers of information (topography, population, streets, traffic flow, power, 3D buildings, albedo, radiation measurements, etc.) to characterize the urban landscape.

These disparate sources of data must be minable and the results able to flow unimpeded into the analysis. Physical radiative-adaptive models initialized with thermal infrared satellite imagery and surface observations will determine the urban energy budget and map the thermal imprint. New and renovated buildings with vegetated rooftops and solar panels will be added to the TanGIS modeling system with GRASS, and diagnosed with radiative transfer modeling for impact on thermal imprint reduction. Microeconomic models will be applied to cost/benefit scenarios, social/economic impacts, improving in social welfare, growth, trade, and jobs. Global and regional climate model output from models such as WRF and SLOSH will be used in predictive mode to characterize sea level changes and run scenarios for sea level rise and storm surge inundation. This information will be used to assess the resilience of surrounding communities, adjust development strategies to mitigate risk, promote sustainable development, determine whether to move/rebuild infrastructure or harden existing developments, revert developed areas to protective wind/wave energy-absorbing wetlands, and preserve/purchase land to ensure the successful migration of people, habitats and wildlife to higher ground.

Data Mining Ensemble Forecasts. Each spring since 2007 the Center for Analysis and Prediction of Storms (CAPS, University of Oklahoma) has been utilizing TeraGrid to generate multi-model storm-resolving forecasts (4-km and 1-km grids) for use in the development of new forecasting tools for the National Weather Service and other entities (Clark et al., 2011, Xue et al. 2011). Forecasters and researchers view and evaluate output from these models in real-time in the Hazardous Weather Testbed, a collaborative project among OU, the National Weather Service and the National Severe Storms Laboratory (NSSL). The three-dimensional data from these forecasts comprise hundreds of terabytes of data saved in the mass store system at the Pittsburgh Supercomputing Center (PSC) and the National Institute for Computational Sciences (NICS). Two-dimensional slices of key fields from these model runs, a much smaller dataset, are housed at CAPS at OU and at NSSL. Some data are also stored for evaluation at the Developmental Testbed Center. A large amount of research and number of publications has emerged from these datasets in the evolving field of ensemble forecasting techniques, yet the use of these fields for science discovery is limited by the vast size and cumbersome access of the original data, and by the limitations of the smaller data set, because decisions had to be made *a priori* about which fields and levels to extract.

The operational models that are archived at the NOAA National Climatic Data Center (NCDC), and available via NOMADS (Rutledge et al., 2006) along with quasi-operational model runs from the National Center for Atmospheric Research (NCAR), the University of Washington (Mass et al., 2010), and other universities represent other large data sets available for research.
Through EarthCube a researcher should be able to easily discover the existence of these datasets, learn about the data sets from the metadata (e.g., models and configurations that were included, available variables, etc.), gain fast access to the 2-dimensional datasets, run queries on those datasets at the source looking for certain forecasted features or characteristics of the ensemble, then identify specific times and locations for 3D exploration. This will aid discovery by weather data mining tools such as recent work by Gagne et al. (2009), McGovern et al. (2010) and Liu et al. (2008). An EarthCube tool will then access the PSC mass store for the relevant 3D model files and extract the required subdomain and variables. This is ideally done at the source to minimize network bandwidth utilization. The subdomain is delivered as a set of 3D fields in a form, such as netCDF, that will allow for 4-D mining for specific relationships associated with observed weather, such as tornadoes, damaging winds, flash floods, etc. The tools would also be able to interface with verification data, such as the NCDC Storm Data, documenting storm damage, and the NEXRAD radar data archive to guide searches and verify hypotheses.

Useability of EarthCube Data. One concern expressed by many scientists is “loose standards”. Having dealt with a variety of surface and radar data and trying to make them all work in the OU-CAPS modeling system, there is always a certain amount of human examination of the data files needed to try to piece-together what is in the file and the structure of the data. Because data providers are never going to completely follow community standards, there must be some sort of smart interface layer that can sort these things out and possibly reformat the data to a community standard. Perhaps this can be done by an automatic scan of the data, then an interactive tool that identifies any unusual features and requests human input to develop a mapping from the provided data variables and layout to the desired variable names and storage organization. For example, by hand one atmospheric research scientist sorted several netCDF file formats of surface, profiler and radiosonde data and wrote code to convert it, filter it (by time and location), and output it in a simple ASCII format that is needed for our analysis program, ADAS. This process could be more automated to allow quicker adaptation of new and unusual data sources

There should be multiple layers of metadata. So we might have metadata about the 2D fields and then metadata about coherent features found in the 2D fields, perhaps from previous searches. Finally, information must point to the original source files from which the 2D files were extracted. If multiple copies exist, mirror file servers can also be indentified to permit fast access.

Surface Transportation. It is the year 2020. Millions of passenger vehicles are now equipped with wireless communication devices (e.g., cell, Wi-Fi, Dedicated Short-Range Communications DSRC) allowing them to function as Lagrangian weather and road condition sensors and enabling data communication to and from the vehicle. The driver starts the car and enters his/her trip plan. The vehicle seeks the latest navigation information, maps, weather hazards, road condition, and traffic information along the route from the EarthCube cloud. The data are processed to provide the safest and most efficient route. As the vehicle progresses along its route, vehicle probe data (e.g., ambient air temperature and pressure, wiper settings, stability control, headlamp, ABS brake activity, pavement temperature, ice, ponding of water, speed, time and position) (Chapman et al. 2010, Drobt et al. 2010) are sent from the vehicle every minute to the EarthCube and shared among public and private weather service and transportation management organizations. These data, anonymized for personal privacy, are combined with other traditional weather datasets gathered from the EarthCube by the weather enterprise (e.g., radar, satellite, surface observations, etc.) and fused to diagnose 1-km road segment weather and road condition hazards. Hazards specific to the travel route are then communicated back to the vehicle navigation system and driver (Mahoney et al. 2010). The vehicle recommends an alternate route to avoid snow-covered roads and the driver accepts the change. The new route information is then communicated to EarthCube cloud data aggregator service so that it can be factored into updated traffic prediction algorithms, shared by other travelers and manage the road network by controlling traffic signal timing, for example. The driver used in the example arrives safely and the weather, road condition, and vehicle data are utilized by other drivers to help them avoid hazards and congestion. This live data detection and feedback scenario is repeated tens of millions of times per day across the nation contributing to a reduction in crashes and improved mobility for private and commercial vehicles (Petty and Mahoney 2008). Overall system efficiency is improved and fuel consumption and pollution are reduced. The vehicle probe datasets are captured by the cloud to support thousands of other applications and research and development activities.

Renewable Energy. It is the year 2020. The nation is progressing with its goal of reaching 20% wind energy generation by 2030, but it still needs to improve its ability to accurately predict wind energy generation for all wind turbines from 1 to 36 hours into the future to reduce the variable generation integration costs by anticipating sudden changes in wind speed (wind ramps) and dangerous high wind events that require feathering of the turbines (Cheng et al. 2011, Johnson et al. 2011). In order to analyze the current state of the atmosphere, millions of tropospheric and land surface characteristics measurements are required, particularly in the lowest 300 m of the atmosphere. Disparate datasets need to be acquired, quality checked, and assimilated into numerical weather prediction models.

The EarthCube data network is utilized for this purpose as it provides access to local and global land surface and atmospheric datasets via a distributed data server. A wind energy prediction system is invoked to seek all relevant datasets to optimize the forecast skill. Data are gathered from wind turbines, tall instrumented towers, regional and commercial aircraft, lidar and Doppler radar wind profilers, insolation, Doppler weather radars, surface sensors, commercial and passenger vehicles, rawinsondes, satellites, vegetation leaf fraction, sea surface temperatures, and snow cover. The system is able to identify high-quality data sources by tracking performance over time, marking with metadata, and filters the data accordingly. As the EarthCube data repository expands, the system is able to identify the most useful new data sources and adjusts its processing accordingly (via self-learning) resulting in constantly improving forecast skill for the renewable energy industry. The EarthCube enables rapid improvements in forecast skill keeping the adoption of renewables on a healthy track.

References

Baltimore City Office of Sustainability, 2010. Baltimore City 2010 Annual Sustainability Report, Baltimore, MD, 52 pp. http://cleanergreener.highrockhosting2.com/uploads/files/AnnualReport.2010.ForWeb.pdf, www.baltimoresustainability.org

Chapman, M.C., S.D. Drobot, T. Jensen, C. Johansen, W.P. Mahoney III, P.A. Pisano, and B. McKeever, 2010: Using vehicle probe data to diagnose road weather conditions – results from the Detroit IntelliDrive (SM) field study. Transportation Research Record, No. 2169, Maintenance Services and Surface Weather, pp. 116-127.

Cheng, W., Y. Liu, B. Mahoney, M. Politovich, T. T. Warner, K. Parks, and J. Himelic, 2011: Improving the 0-3 hour wind forecast through wind farm data assimilation in the NCAR/ATEC WRF RTFDDA. Second Conference on Weather, Climate, and the New Energy Economy. 23-27 January 2011, Seattle, WA, AMS, Boston, MA.

Clark, A. J., S. J. Weiss, J. S. Kain, I. L. Jirak, M. Coniglio, C. J. Melick, C. Siewert, R. A. Sobash, P. T. Marsh, A. R. Dean, M. Xue, F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, J. Du, D. R. Novak, F. Barthold, M. J. Bodner, J. J. Levit, C. B. Entwistle, T. L. Jensen, and J. James Correia, 2011: An Overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. Bull. Amer. Meteor. Sci., Accepted. (pre-publication document).

Drobot, S.D., M.C. Chapman, E. Schuler, G. Wiener, W.P. Mahoney III, P.A. Pisano, and B. McKeever, 2010: Improving road weather hazard products with vehicle probe data – the Vehicle Data Translator quality-checking procedures. Transportation Research Record, No. 2169, Maintenance Services and Surface Weather, pp. 128-140.

Droegemeier, K., K. Brewster, M. Xue, D. Weber, D. Gannon, B. Plale, D. Reed, L. Ramakrishnan, J. Alameda, R. Wilhelmson, T. Baltzer, B. Domenico, D. Murray, A. Wilson, R. Clark, S. Yalda, S. Graves, R. Ramachandran, J. Rushing, E. Joseph (2005). Service-oriented environments for dynamically interacting with mesoscale weather, *Computing in Science and Engineering*, IEEE Computer Society Press and American Institute of Physics, Vol. 7, No. 6, pp. 12-29.

Gagne II, David J. and McGovern, A. and Brotzge, Jerry. (2009). *Classification of Convective Areas Using Decision Trees,* Journal of Atmospheric and Oceanic Technology. Vol 26, Issue 7, pages 1341-1353. [pdf (1.1 MB)]

Hurricane Irene (2011). http://en.wikipedia.org/wiki/Hurricane_Irene_(2011) - hundreds of references to storm listed at bottom of web page.

IntelliDrive (accessed 2011). http://www.its.dot.gov/its_program/its_factsheets.htm

Johnson, D. B., B. Mahoney, Y. Liu, G. Wiener, W. Myers, and K. Parks, 2011: An overview of NCAR's advanced wind forecasting system for integrating wind resources into the new energy economy. Second Conference on Weather, Climate, and the New Energy Economy. 23-27 January 2011, Seattle, WA, AMS, Boston, MA.

Li, X., B. Plale, N. Vijayakumar, R. Ramachandran, S. Graves, and H. Conover, Real-time Storm Detection and Weather Forecast Activation through Data Mining and Events Processing *Earth Science Informatics,* H.A. Babaie, Ed., Springer Berlin/Heidelberg, Vol 1, No 2, 2008.

Mahoney, B, S. Drobot, P. Pisano, B. McKeever, and J. O'Sullivan: Vehicles as Mobile Weather Observation Systems. Bulletin of the American Meteorological Society, September 2010, 1179-1182.

Mass C.F., M. Albright, D. Ovens, R. Steed, M. MacIver, E. Grimit, T.Eckel, B. Lamb, J. Vaughan, K. Westrick, P. Storck, B. Colman, C. Hill, N. Maykut, M, Gilroy, S. A. Ferguson, J. Yetter, J. M. Sierchio, C. Bowman, R. Stender, R. Wilson, and W. Brown, 2003: Regional Environmental Prediction Over the Pacific Northwest, *Bull. Amer. Met. Soc*., **84**, 1353-1366.

McGovern, Amy; Rosendahl, Derek H; Brown, Rodger A; and Droegemeier, Kelvin K. (2011) Identifying Predictive Multi-Dimensional Time Series Motifs: An application to severe weather prediction. Data Mining and Knowledge Discovery. Volume 22, Issue 1, pages 232-258

Petty, K.R., and W.P. Mahoney, 2008: Enhancing Weather Information Through Vehicle Infrastructure Integration. Transportation Research Record, No 2015. Transportation Research Board of the National Academies, January 2007, Washington, D.C., pp. 132-140.

RWIS (accessed 2011). U.S. Dept. of Transportation, Federal Highway Administration, Road Weather Management Program - http://ops.fhwa.dot.gov/Weather/