Exploitation of Coincident Earth Observations

Catherine Slesnick, Stephan Kolitz, Natasha Markuzon, Mark Abramson
The Charles Stark Draper Laboratory, Inc.
Cambridge, MA 02139

Abstract:

In the current age of rapid climate change, analysis of coincident Earth Science data taken from different observing platforms will provide critical knowledge both to scientists and to policy makers. We are pursuing efforts aimed at overcoming challenges in several different phases of this process. 1) We have developed an architecture for optimized, coordinated dynamic tasking of asynchronously planned and distributed sensor systems, on both satellites and aircraft. Implementation of this architecture allows for an increased likelihood of coincident observation between heterogeneous instruments. 2) We have developed the framework for a web portal to help scientists and decision makers more easily discover which Earth systems are observed as a function of time and space. The web portal would allow users to visually determine coincident data availability within a specified space-time query, thus enabling an investigator to efficiently identify and analyze all available observations without knowing *a priori* which assets were monitoring his/her domain of interest. 3) We are using data mining techniques to develop models that use coincident Earth observations combined with a range of expected climate change parameters to determine statistically likely event scenarios for extreme disturbance events (e.g., fires, landslides, spread of infectious disease) in the near- and long-term future.

I. PLANNING FOR COINCIDENT EARTH OBSERVATIONS

Draper Laboratory's Earth Phenomena Observation System (EPOS) was primarily developed in a number of NASA ESTO AIST projects (under AIST-99, AIST-02, and AIST-05). The fundamental EPOS concept of operation is that of optimized dynamic replanning and execution. Sensor data and model forecasts are inputs to a closed-loop decision-making system. In collaboration with users, EPOS monitors the input, and when appropriate, replans and executes a new plan that optimizes the tasking of available sensing assets to gather data. The functional architecture for EPOS is illustrated in Figure 1. Included in this architecture are three primary components.

Situation Awareness: Situation Awareness provides estimates of current world and system states.

Situation Assessment: Situation Assessment takes Situation Awareness output and uses Information Exploitation technologies, e.g., pattern recognition and data mining, to support monitoring, diagnosis and prediction of world and system states.

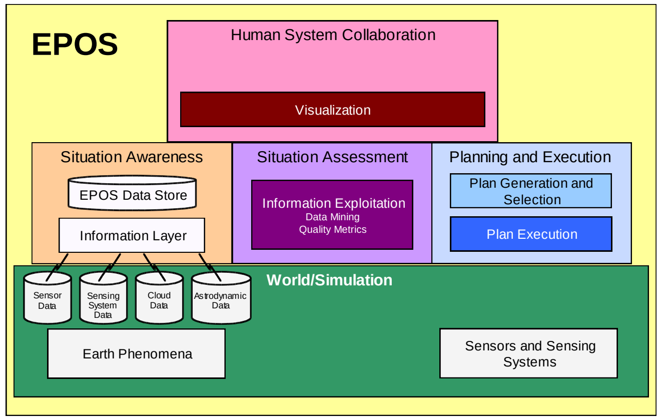

Figure 1: EPOS Functional Architecture

Planning and Execution: Planning and Execution is optimization-based technology that will support a full range of operational autonomy, from manual operator control to full autonomy. Recently, the Coordination Manager (Abramson, Kahn and Kolitz 2010; Herold, et al. 2010) has been integrated into this EPOS function. The Coordination Manager addresses the problem of asynchronously planned stovepipe legacy single mission Earth observation systems. The benefits achieved by introducing a Coordination Manager into this situation are described in the remainder of this section.

More efficient use of resources: The Coordination Manager knows about all the capabilities of Earth observation systems and when they are available. The Coordination Manager can optimally balance timeliness, quality, and availability factors for a request. Previously difficult to accomplish coordinated observations can be routinely scheduled.

Enterprise-wide value is a factor in collection decisions: By definition, Collection Manager decisions include global optimization criteria, e.g., quality and benefit of collected data, timeliness, and cost. During each of its planning cycles, the Collection Manager will make the most effective use of all Earth observation resources to maximize collection of requests.

Highly dynamically responsive: The Coordination Manager can get requests assigned to the earliest available and appropriate Earth observation system through knowledge of the planning phases and capability of all the systems.

Increased opportunities for enhanced science value through coincident observation collection: The Coordination Manager can assign coincident or near-coincident joint collection of observations of the same location easily given its knowledge of the planning phases and capability of all the Earth observation systems. This makes joint collection scheduling a real-time capability and increases the potential use of this feature.

Joint collection of data is useful in a variety of applications, e.g., calibration, validation, science campaigns. The Coordination Manager can plan coincident observations, but still remaining is the challenge of subsequently retrieving the data from the disparate Earth observation systems.

II. RETRIEVING COINCIDENT EARTH OBSERVATION DATA

Whether through a capability such as the Coordination Manager or through constellation design (e.g., the A-train), there are coincident or near-coincident joint collections of observations of the same location by multiple platforms that produce geoscience data. These platforms include a combination of (i) space-based satellites aimed at global or large-area mapping, (ii) airborne experiments designed to make regional or site-specific observations, and (iii) ground-based observatories that monitor local events. Each of these platforms may host several instruments. From each instrument, and each combination of instruments, dozens of data products can be created, and the spatial extent of these data products is constantly changing as the observing platforms shift their views. The fact that most of these observations are not yet coordinated in an optimized method (e.g., using tools such as the Collection Manager described in Section I) adds an additional layer of complexity to the system.

To organize the enormous cumulative volume of data from these products, dozens of data access services have been created to help users find and understand data of interest. These services are usually organized around individual data types or instruments. For example, snow data is kept at the National Snow and Ice Data Center, and data specific to Global Positioning system instruments is available at UNAVCO. This stove-pipe system of data availability comes at a high cost to users, and is ultimately inefficient for the whole community. Currently, for each service, a user must learn a different navigation system, not only to retrieve data, but also to discover if the desired data exist. This situation is aggravated further by the fact that the same data set may be available through multiple data access points. In the current state of Earth Science data access, investigators spend a great deal of time and money identifying, collecting and sorting data from disparate distribution services, rather than spending available resources pursuing science goals. The heterogeneous system of data availability poses a particular challenge to current investigators pursuing research that requires spatially and temporally coincident observations from multiple earth observing instruments. The required process is inefficient, and demands *a priori* knowledge of many different data access methods.

We have developed the software framework for a new web portal, **Vi**sualization of **C**oincident **E**arth **O**bservations (ViCEO), that will provide a solution to this problem. ViCEO will enable users to know *a priori* what earth observing assets were monitoring the location and time of interest prior to an extensive search to retrieve the data. ViCEO will allow users to visually determine data availability within a specified space-time query, and will further provide detailed information about each dataset, thus enabling an investigator outside his or her area of expertise to efficiently identify and analyze the most appropriate data. Details of the envisioned user interface (UI) are shown in Figures 2 and 3.

Advantages of ViCEO for the user - ViCEO will:

- provide a central portal from which to efficiently identify coincident datasets.
- help users to discover data that are potentially pertinent to a specific investigation, but that may otherwise have been overlooked.
- make documentation for each dataset easily discoverable, with clear information regarding data formats and required analysis software.
- provide a central portal for researchers to stay up to date on data availability.

Advantages of ViCEO for existing data access services and repositories - ViCEO will:

- allow existing data access sites and repositories to be more easily discovered.
- advertise those sites to users to whom they were previously unknown.
- provide descriptions of how to navigate each site, thereby increasing the likelihood of data being downloaded and used.

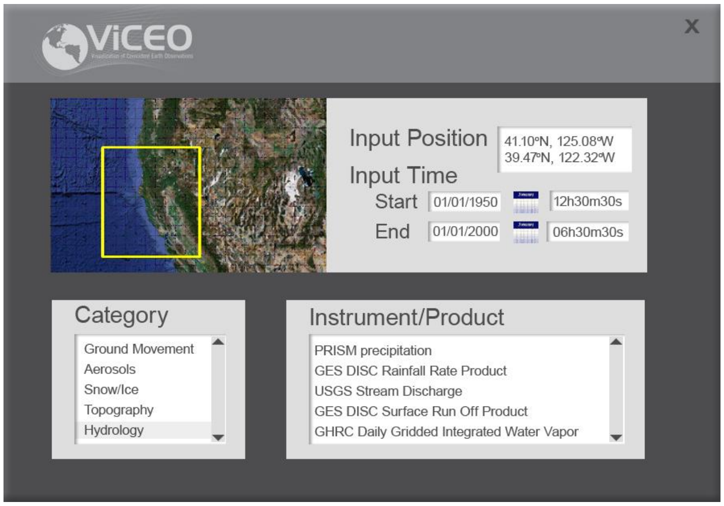

Figure 1: The user will select a desired spatial box in latitude and longitude, and a temporal window. Optional science categories may be selected to narrow a search.

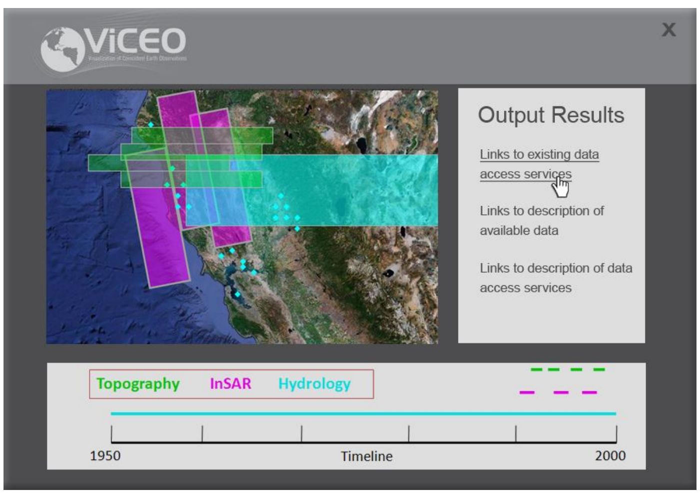

Figure 2 - A Google map will be displayed with overlays of available data, and a timeline will be created to help the user visualize temporally coincident data.

ViCEO is designed to be used both by Earth Scientists looking for easier access to a desired data set, and researchers who have a general idea of what they would like to achieve but who may not yet know all of the details. Examples include investigators seeking to incorporate new types of data into existing research, decision makers and scientists who desire to know how earth systems interact, and students starting out on a research project.

While currently several web resources exist that visually bring together data in time *or* space (e.g., OpenTopography/GEON Catalog, EarthScope, CUAHSI Hydrologic Information System), ViCEO represents a unique approach to Earth Science data access in that it will, for the first time: (1) fuse data in both time *and* space, and (2) fuse completely disparate data types into a common portal.

The architecture of the ViCEO framework integrates commercial off the shelf (COTS) web-based products with custom software to provide a simple and efficient search mechanism to query multiple data sources concurrently. Most of the technologies that will be used are cutting edge standards in the web community. The UI layer of the framework is based on the fundamentals of the Stripes technology which is a presentation framework for building web applications. The software will implement Apache Tomcat, a Java Servlet and Java Server Pages (JSP) technology that enables rapid development and maintenance of information-rich, dynamic web pages that are platform agnostic. JSP technology separates the UI from content generation, enabling changes to the overall page layout without altering the underlying dynamic content. The servlet software framework provides a logical separation of function and data transport across the software architecture allowing for additional data sources to be integrated without large scale changes to the core. This functionality provides an open-ended service system. Thus, the ViCEO implementation of this technology will have the flexibility to add additional data sources in the future with little or no necessary framework modification.

Once developed, the ViCEO web portal will help to revolutionize the way investigators retrieve Earth Science data. The ease with which ViCEO will allow access to coincident Earth observations will enable research studies that may not have been previously possible due to limited time and monetary resources.

III. Science with Coincident Earth Observation Data

Once implemented, the web portal described above, together with several other newly developed data access sites, will allow investigators access to unprecedented amounts of information about spatially and temporally coincident events. With this enormous influx of data comes new challenges for how best to ingest and evaluate the observations.

Data mining, or data-driven modeling, is capable of pattern evaluation, normalcy modeling and anomaly detection, future state prediction, and determination of dominant variables in a decision process. This technique offers several advantages for assessment of Earth observations, including: 1) the ability to combine large volumes of heterogeneous data into a single model; 2) the development of data-driven models when no physical models exist; and 3) the assessment of both spatial and temporal system components in concert. Data mining also allows the investigator to explore non-linear relationships between parameters without making prior assumptions, thereby enabling discovery of new relationships between variables. For these reasons, data-driven modeling can provide information complementary to traditional analytical models, and is becoming increasingly used to predict the effects of climate change on Earth systems such as cloud-ceiling-height forecast (Bankert & Hadjimichael 2007), wildfires (Cheng & Wang 2008; Markuzon and Kolitz 2009), landslides (Lin et al. 2008; Nefeslioglu et al. 2009; Xianmin & Ruiqing 2009; Zia et al. 2006), and vegetation variability (Whitea et al. 2005). The basic principles of this methodology are shown in Figure 4.

We are using data mining techniques applied to coincident Earth observations to better understand shifts in extreme disturbance events, such as fires, landslides, or floods, under the influence of climate change. Unforeseen changes in frequency, intensity, pattern or duration of these events could lead to catastrophic economic and human losses as well as near-term and long-term shifts in anthropogenic and natural land cover, and is one of the most important challenges facing society today. To accurately monitor and predict the risks of future events requires a thorough understanding of the coupled system of observational analyses and model development. Our study integrates remote sensing data and existing analytical models of Earth systems with newly-developed data-driven models capable of producing high temporal- and spatial-resolution projections. The goal of these efforts is to demonstrate the ability of data mining algorithms to help forecast both extreme disturbance events and the ensuing consequences.

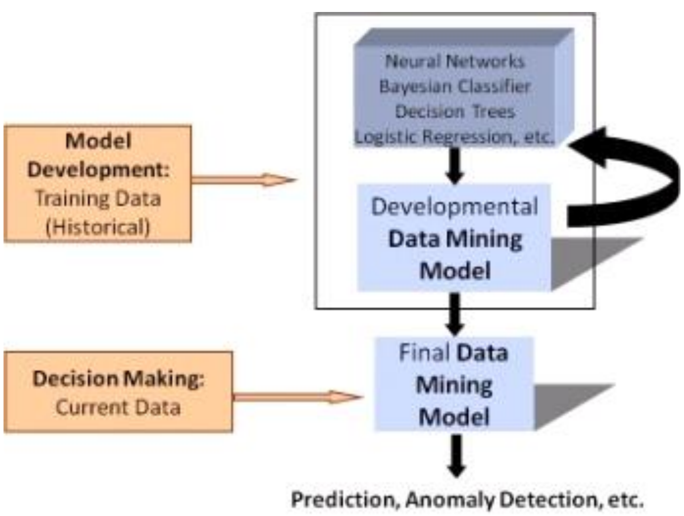

Figure 4 – Data-driven model development and evaluation.

For the first part of this study we have developed data mining models that combine thermal observations with land cover and weather information to determine the probabilistic risk of fire expansion and movement in the Southwestern part of the United States. We fused three datasets to create our model:

a. MODIS/Terra Thermal Anomalies/Fire Daily Level 3 sinusoidal grid product. These data represent thermal anomalies and global daily fire information at a pixel resolution of 1km.
b. hourly NOAA weather information including temperature, relative humidity, wind speed, wind direction, wind gusts collected from weather stations across the US. We approximated weather conditions by using information from the closest to the fire weather station. Weather data were smoothed to daily resolution to match the MODIS fire observations.
c. high level 2001 National Land Cover Database (NLCD). This database includes 9 classifications, in particular water, developed areas, barren, forested upland, shrubland, non-natural woody, herbaceous upland, planted/cultivated, and wetlands. These data were used as a proxy for the amount of combustible material.

Fires tend to move, making it important to monitor fire progression daily. In all, 17 features represented each day of a fire, including the geometry of the fire, surrounding weather, and land cover data. We observed fires for three days to predict its future growth.

We introduced a binary variable that was equal to one if the fire became “large” on the day following three days of observations or the other day, and zero otherwise. We considered fire to be “large” if its area covered 6 or more contiguous pixels. We tested several algorithms that have a history of good performance, including Random Forest Decision Trees Bayesian Networks, and K Nearest Neighbor. None of the classifiers showed a significantly superior performance over the others. The resulting models are capable of predicting threats up to two days in advance, and can be used to aid in the allocation of disaster recovery resources.

This study is the first of many studies we intend to carry out applying data mining techniques to Earth observations. Other areas of interest include: 1) assessment of how climate change will impact extreme precipitation and landslide hazards, and what risks those events will pose for natural and human systems in the future. 2) use of coincident Earth observations to help predict the spread of infectious diseases.

Acknowledgments

Part of this material is based upon work supported by the National Aeronautics and Space Administration's Earth Science Technology Office (ESTO) AIST-05 program, under Grant Number NNX06AG17G. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration

References

Abramson, M., A. Kahn and S. Kolitz, (2010). Coordination Manager – Antidote to the Stovepipe Antipattern, Proceedings AIAA Infotech@Aerospace

Cheng, T., and J. Wang, (2008). Integrated Spatio-temporal Data Mining for Forest Fire Prediction. Transactions in GIS 12-5, 591-611.

Herold, T.M., Abramson, A. Kahn, S. Kolitz, H. Balakrishnan (2010). A synchronous, Distributed Optimization for the Coordinated Planning of Air and Space Asset Operations, Proceedings AIAA Infotech@Aerospace

Lin, Y.B., Y.P. Lin, D.P. Deng, and Chen (2008). K.W. Integrating remote sensing data with directional two dimensional wavelet analysis and open geospatial techniques for efficient disaster monitoring and management. Sensors, 8, 1070-1089.

Markuzon, N., and S. Kolitz (2009) Data Driven Approach to Estimating Fire Danger from Satellite Images and Weather Information, 38th IEEE Applied Imagery Pattern Recognition Workshop

Nefeslioglu, H.A., E. Sezer, C. Gokceoglu, A.S. Bozkir, and T.Y. Duman (2009). Landslide susceptibility mapping of a landslide-prone area by data mining, Geophysical Research Abstracts, Vol. 11, EGU2009-13499

Xianmin, W., & Ruiqing. (2009) Spatial Forecast of Landslides in Three Gorges Based On Spatial Data Mining. Sensors 9, no. 3: 2035-2061.

Whitea, A.B., P. Kumara., and D. Tcheng, (2005). A data mining approach for understanding topographic control on climate-induced inter-annual vegetation variability over the United States. Remote Sensing of Environment. v. 981-1, 1-20.