Computational Sciences: At the Intersection of Science and Engineering - Case Study for Academic and Research Programs

Menas Kafatos  
*Chapman University*, kafatos@chapman.edu

P. A. Becker  
*George Mason University*

Z. Boybeyi  
*George Mason University*

Follow this and additional works at: [http://digitalcommons.chapman.edu/scs_articles](http://digitalcommons.chapman.edu/scs_articles)

Part of the [Computer Sciences Commons](http://digitalcommons.chapman.edu/computer_sciences_commons) and the [Engineering Commons](http://digitalcommons.chapman.edu/engineering_commons)

**Recommended Citation**

Computational Sciences: At the Intersection of Science and Engineering - Case Study for Academic and Research Programs

Comments
This article was originally published in *Turkish Journal of Electrical Engineering & Computer Sciences*, volume 14, issue 1, in 2006.

Copyright
David Publishing
Computational Sciences: At the Intersection of Science and Engineering–Case Study for Academic and Research Programs

Menas KAFATOS, Peter BECKER, Zafer BOYBEYİ
School of Computational Sciences, George Mason University
Fairfax, VA 22039, U.S.A.
e-mail: mkafatos@gmu.edu

1. Computational Sciences in the 21st Century

Computational Science can be simply defined as the performance of science using computers. However, this elementary definition masks some of the important subtleties that make this new field a hallmark of interdisciplinary science in the 21st century. Computational Sciences are at the intersection of science and engineering. Computational scientists often work in large interdisciplinary programs that span not just one or more scientific fields, they also work closely with computer scientists and engineers. The present article demonstrates the applied nature of the field (as in engineering), the scientific focus (as in scientific disciplines) and the cross cutting or interdisciplinary nature of computational sciences. Faculty and researchers can possess a science degree or an engineering degree. However, unlike engineering and computer science fields, the overall emphasis and driver is science and relevant applications.

Perhaps it is best to begin the discussion with some examples. A classic example at the cutting edge of computational science a decade ago is the simulation of quantum mechanical systems, with the goal of understanding the fundamental behavior of matter. Such small-scale simulations offered little real world benefits because they focused on highly symmetrized, relatively simple systems. Computational science has evolved at a fantastic rate since then, so that in the current epoch, this simple example from the realm of quantum physics has blossomed into the simulation of complex, macroscopic systems including solid-state surfaces covered with exotic structures such as nanotubes [1]. The simulation of these relatively large systems allows researchers to carry out “virtual experiments” that would be either impossible or extremely expensive to conduct using actual physical procedures [2],[3]. This example serves to illustrate one of the most compelling reasons for doing computational science in the modern era. Quite simply, computational techniques allow researchers access to experimental procedures, simulated using computers, which would be either prohibitively expensive or perhaps even impossible to conduct in reality.

Expanding on the theme of virtual experimentation, modern computational science includes applications as diverse as the simulations of aircraft and automobile aerodynamics, as well as the hydrodynamics of ships at sea [4]. These applications are central to the modern design process for vehicles of all sorts, up to the largest aircraft carriers. Computational science has also led to an improved world through the reduction of the environmental damage resulting from inefficient coal-burning power plants, and the testing of nuclear
weapons. Both power-plant design and, amazingly, nuclear detonation, can now be well simulated using computational techniques. Computational Science also plays a major role in the space program, in the general context of space-based communication and data transmission, as well as support for the state-of-the-art artificial intelligence required by autonomous robots such as the Mars Rovers [4].

Computational Science is an interdisciplinary area that draws upon the traditionally distinct areas of computer science, applied mathematics and statistics, and one or more of the natural sciences. It is more than any of its component fields, and therefore it does not fit within any of the traditional subject areas. An example of a research endeavor that would be considered part of Computational Science is the development of a large-scale computer simulation code to investigate a scientific question that goes beyond the use of canned (commercial or freely available) software. The scientific component in this case may be fluid dynamics, astrophysics, physics, or chemistry. Another example is the development of a large-scale Earth Observing data archive and query system that facilitates the analysis and validation of global climate models. This requires not only an understanding of the science, but also insight into a number of specialized aspects of the simulation, such as computer architecture or algorithms, that would ordinarily be considered to lie inside a “black box” from the viewpoint of the scientific user. This often results in much higher performance, efficiency, or scalability than a generic “brute force” technique is likely to achieve, resulting in an unprecedented degree of fidelity in the results of the simulation.

There are many international groups working to promote the continued development of computational science. Examples of government-funded entities in the United States include the National Center for Supercomputing Applications (funded by the National Science Foundation), the National Center for Computational Sciences (at Oak Ridge National Laboratory), the National Computational Science Institute (hosted by the non-profit, NSF-supported Shodor Education Foundation), NASA’s Computational Sciences Division (at Ames Research Center), the Mathematical and Computational Sciences Division of the National Institute of Standards and Technology), the Mathematical, Information, and Computational Sciences Program in the Department of Energy [1]-[6].

2. School of Computational Sciences

In addition to these federally funded research organizations listed above, there are also many universities around the world engaged in computational science research and education. An important example is provided by the School of Computational Sciences (SCS) at George Mason University (GMU), which founded the first computational sciences doctoral program in the United States in 1992. The mission of SCS is to provide quality graduate education, research, and service emphasizing the central role of computational methodologies in the biological, physical, mathematical, and data sciences [7]. The educational and research programs of SCS are highly interdisciplinary, with an emphasis on theoretical science, computer simulation, data studies, and hardware design and development. The objective of SCS is to provide Virginia, and the nation as a whole, with world-class resources for attacking the interdisciplinary research problems that characterize the challenges faced as we move into the new millennium. The inclusion of theoretical, laboratory, and data science components results in a unique community of scholars ideally suited for the challenges of interdisciplinary research in the years ahead.

SCS serves as the primary academic unit within George Mason University providing state-of-the-art scientific and applications content to GMU’s information technology focus. This content includes applications in the biological, physical, mathematical, and data sciences. Along with other units, SCS also contributes
to the university’s focus on educational and research programs related to the environment and homeland security, as well as to areas involving fundamental sciences. The research and teaching activities of SCS reflect the recognized role of computation as part of a triad with theory and experimentation, leading to a better understanding of nature than can be obtained via classical theory or experimentation alone. The triad is represented graphically by the following figure, produced by the Shodor Education Foundation.

![Triad of Theory, Computation, and Experiment](image)

**Figure 1.** Triad of Theory, Computation, and Experiment.

Through its interdisciplinary and multidisciplinary activities, SCS provides a perfect environment for the education of Fellows in areas of national need involving the Computational Sciences. Specific examples of documented areas of national need that fall within the purview of SCS academic programs are the following: Computational Sciences and Informatics; Bioinformatics; Climate Dynamics; Earth Systems and Geoinformation Sciences; Neuroscience; and Computational Social Science. In particular, the doctoral program in Computational Sciences and Informatics offered by SCS provides students with many opportunities to conduct interdisciplinary research in a number of areas that are at the forefront of modern scientific advances.

We now turn our attention to two representative areas in computational sciences, one involving compute-intensive applications; the other involving data-intensive applications. Both broad areas deal with real-life applications and necessitate the interaction of computer specialists, data specialists, as well as discipline scientists such as atmospheric scientists and remote sensing scientists. Although housed in the School of Computational Sciences at George Mason University, they cut across different disciplines and even are utilizing multi-university and agency teams. The first describes work in the atmospheric modeling program area, led by one of us (Zafer Boybeyi). The second focuses work in the remote sensing data area also led by one of us (Menas Kafatos).

### 3. Numerical Modeling

Numerical modeling applies to different scale atmospheric science, from global to regional or even local. One needs to discuss both the tools and the science areas. Here we concentrate on numerical tools and applications in atmospheric science. For the tool, the Operational Multiscale Environment model with Grid Adaptivity (OMEGA) is used for numerical simulations. OMEGA is an operational real-time atmospheric
modeling system. The modeling system consists of: 1) a fully non-hydrostatic, three-dimensional prognostic model. The model is based on an adaptive, unstructured triangular prism grid that is referenced to a rotating Cartesian coordinate system (see Figure 2). The model uses a finite-volume flux-based numerical advection algorithm. The model also contains a Lagrangian dispersion algorithm embedded into the model for atmospheric transport and dispersion simulations; 2) routines to maintain and manage real-time weather data feeds from NCEP, 3) world-wide datasets for surface elevation, land/water, vegetation coverage, soil type, land use, deep soil temperature, deep soil moisture, and sea surface temperature at varying resolutions; 4) an integrated Graphical User Interface that provides a user-friendly method for rapid model re-configuration; 5) a grid generator that accesses the surface datasets and creates grid and terrain files; 6) a meteorological data pre-processor that ingests gridded terrain, gridded meteorological analyses and forecasts, and raw observations, and performs a detailed quality control of the ingested data, followed by an Optimum Interpolation data assimilation to produce initial and boundary conditions; and 7) several graphical post-processing tools that enables the user to display model output as two-dimensional slices, skewT-logP profiles for any location, and animations.

![Figure 2. The OMEGA grid structure: (a) an OMEGA grid element, and (b) the coordinate system and vertical alignment of the grid.](image)

The unstructured grid of the model provides flexibility to facilitate the gridding of arbitrary surfaces and volumes in three dimensions. In particular, unstructured grid cells in the horizontal dimension can increase local resolution to better capture the topography or important physical features of atmospheric flow and cloud dynamics (e.g., initial location of a hurricane). This grid feature provides smooth transition from high resolution where needed to low resolution elsewhere and hence effectively removes the wave reflection and wave dispersion problems that are common in other grid techniques. The model grid can also adapt dynamically to a variety of user-specifiable adaptivity criteria (e.g., precursors to convection, eye of a hurricane). In other words, the model focuses the model’s horizontal grid resolution during the run on regions of complex and critical phenomena to improve the overall quality and efficiency of simulations.

The modeling system has been successfully used and validated for scales ranging from local scale (a few kilometers), to large severe events such as hurricanes (hundreds of kilometers). A full description of the system can be found in Bacon et al. [8],[9] and the details of validation studies can be found in Boybeyi et
4. Example Simulations

Example simulations of OMEGA model are presented in this section.

4.1. Dust storm simulation

To demonstrate the ability of OMEGA to track the dust through the atmosphere, a simulation was made of a Saharan sandstorm. A massive sandstorm blowing off the northwest African desert blanketed hundreds of thousands of square miles of the eastern Atlantic Ocean with a dense cloud of Saharan sand. The massive nature of this particular storm was first seen in a SeaWiFS image (see Figure 3a) on Saturday, 26 February 2000 when it reached over 1000 miles into the Atlantic. An OMEGA simulation using a specified source region produced the simulated plume shown in Figure 3b. These storms and the rising warm air can lift dust 15,000 feet or so above the African deserts and then out across the Atlantic, many times reaching as far as the Caribbean where they often require the local weather services to issue air pollution alerts as was recently the case in San Juan, Puerto Rico. Recent studies by the U.S. Geological Survey have linked the decline of the coral reefs in the Caribbean to the increasing frequency and intensity of Saharan Dust events. Other studies suggest that Saharan Dust may play a role in determining the frequency and intensity of hurricanes formed in the eastern Atlantic Ocean.

Figure 3. A close-up view of (a) the SeaWiFS images of a sandstorm blowing off the northwest African desert that blanketed hundreds of thousands of square miles of the eastern Atlantic Ocean with a dense cloud of Saharan sand on Saturday, February 26, 2000. The OMEGA simulation results are shown in (b).

4.2. Hurricane simulation

Accurate forecast of hurricanes is a high priority topic of research, due to their potential large economic impact as well as public safety issues. (Note that only economic impact will be addressed in our discussions,
as human lives lost we consider priceless.) Hurricanes rank among the most destructive and costly of natural phenomena. For example, in 1992, hurricane Andrew caused more than $30 billion in direct economic losses in south Florida. The most recent hurricanes of 2004 (Charley, Frances, Ivan and Jeanne) that spun toward and crossed Florida, caused over $20 billion in losses. Untold damage is still being assessed for hurricane Katrina (August, 2005).

Even more alarming, total loss estimates for a category-4 hurricane striking Miami, Florida were estimated at $60 billion. The losses from Katrina may exceed $100 billion. This trend in damages highlights the importance of and need for better TC track and intensity forecasting. The model forecast skills can be improved by enhancing the accuracy of initial conditions of numerical weather prediction (NWP) models. Figures 4, for example, shows one such study in which temperature and dew point temperature soundings from the GOES satellite were assimilated to improve the model initial condition [11]. The results show that by reducing uncertainties in the steering polar jet entrance flow region at initial time, the track error was improved by about 30%. Lorenz had reported that in a NWP model, the initial uncertainty is mainly associated with the low resolution of the observations. This problem is amplified further in hurricane initialization due to lack of observations over the ocean regions.

5. Virginia Access Middle Atlantic Geospatial Information Consortium: A Data-Intensive Interdisciplinary Project

5.1. Description of the project

VAccess/MAGIC is a consortium of Middle-Atlantic universities promoting the usage of NASA earth observing and related data led by Menas Kafatos. The work focuses on two primary aspects: The usage of such
data for national priorities; and the dissemination of such data through interoperable information systems, that couple to NASA’s systems and promote open source solutions and standards.

The two-fold aspects manifest in all MAGIC products: Environmentally-important applications are developed, the stakeholders are identified, relationships with partner federal agencies are developed and when appropriate, local and even international relationships are promoted. Applicable value-added data and information products are developed, adding to NASA’s mission. Moreover, decision support systems are identified as applicable to these application areas.

Table 1. Priorities and themes of the MAGIC program

<table>
<thead>
<tr>
<th>National Priorities</th>
<th>Major MAGIC Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecological Forecasting</td>
<td>Wetlands</td>
</tr>
<tr>
<td>Disaster Preparedness &amp; Water Management</td>
<td>Local Impacts of Global Climate Phenomena</td>
</tr>
<tr>
<td>Agricultural Efficiency</td>
<td>Agriculture</td>
</tr>
<tr>
<td>Carbon Management</td>
<td>Forestry</td>
</tr>
<tr>
<td>Public Health</td>
<td>Public Health</td>
</tr>
<tr>
<td>Invasive Species</td>
<td>Ground Truth Data</td>
</tr>
<tr>
<td>Air Quality &amp; Coastal Management</td>
<td>Pollution</td>
</tr>
<tr>
<td>NASA Program Needs/Alignment</td>
<td>Cross Cutting Techniques/Technology</td>
</tr>
<tr>
<td>Education</td>
<td>Education Outreach &amp; Workforce Development</td>
</tr>
<tr>
<td>Cross Cutting &amp; Collaboration</td>
<td>Information, Data Dissemination &amp; DSS</td>
</tr>
</tbody>
</table>

The second aspect, data dissemination is of equal importance: Interoperable technologies, coupling to NASA’s systems, will be utilized, following accepted standards, including metadata. The MAGIC federated solutions utilize open source and existing solutions. In this way, MAGIC assists in dissemination to its various stakeholders NASA and related data. The table summarizes activities that align with NASA priorities and needs.

6. Application Areas

Here we describe a few key areas of research and applications in the MAGIC project.

6.1. Wetlands

In MAGIC Phase 3 a cross-correlation analysis of wetlands throughout the Chesapeake Bay watershed using Landsat TM and National Wetland Inventory (NWI) maps was undertaken. In MAGIC Phase 4, we propose to systematize the evaluation of the wetland change results obtained in Phase 3 to enable users to extract the level and precision of information necessary for specific applications. The system will be developed in consideration of the MAGIC team’s Web Mapping Portal for the Mid-Atlantic Region to provide Internet access and interoperable, geospatial visualization and analysis services. The proposed research will develop a decision-support system based on machine-enabled analysis of wetland change probability maps. The user will be able to interact with the system to obtain a range of levels of detail and tuning to their area of interest. The entire Chesapeake Bay Watershed, consisting of parts of six states and Washington, D.C., will be included. Steve Prince, University of Maryland is the lead of this project.
6.2. Global climate phenomena and local impacts

It is projected that temperature, stream flow, and sea level will increase in the Middle Atlantic Coastal region in response to atmospheric CO$_2$ increasing and global warming. Climate change is now understood to play an important role by increasing frequency and severity of extreme weather events. It is suggested that the unique combination of geography, aging infrastructure, rapid population growth and mixed land use over the mid-Atlantic region has imposed considerable stress on its local ecosystems making them vulnerable to extreme weather events. In this study, we focus on potential ability of a mesoscale atmosphere system coupled with land-surface hydrology to aid better diagnosis/prognosis of severe weather events with the conjunction of satellite data available from NASA and NOAA sources [12]. We propose to combine MM5 (eventually WRF) and satellite observations to monitor and study severe weather events, such as hurricanes and hail storms according to the NASA’s application plans outlined for disaster management plans. Since ingesting satellite data to modeling system improves the model’s forecast ability further, MODIS surface products are being used to provide surface characteristics for better surface forcing, while it is planned to ingest GOES imager [12] and sounder data together with MODIS cloud products to better represent cloud parameterization in model atmosphere. Donglian Sun (GMU) and Ismail Yucel (Hampton Univ.) are the leads.

![Figure 5](image_url)

**Figure 5.** Time series of total tropical storm number, and numbers of hurricane and severe hurricane (categories 3, 4, and 5), for the period from 1995 to 2003.

6.3. Agriculture

One of the major focuses of the National Aeronautics and Space Administration (NASA) agricultural efficiency program is to integrate remote sensing data into crop yield forecasts utilizing crop models [13]. CERES and CROPGRO crop model series have been widely used in the US and worldwide. We propose to utilize TRMM rain rate as an input to these crop models for spatial consistency for Virginia state crop yield estimates, and further application to the regions where no gauge rain rate is available. MODIS/LAI is a major input parameter for these models; however, data are available only for 3 crop seasons (2001-2003). We estimate LAI for the years prior to 2001 based on MODIS/LAI and AVHRR/NDVI, via regression analysis, to be used as an input to the crop models for historical yield estimates and validation. Chai Lim (GMU) and Ayman Suleiman (Hampton) are the leads.

6.4. Forestry and carbon management

This component will develop or refine remote sensing applications and related Earth system models needed for forest carbon management. Algorithms, models, and protocols are tailored for use of data from NASA
and commercial sensors to support forest carbon management decision support systems used by a wide variety of partners and stakeholders, including the coal industry, NGOs, forest industry, state natural resources agencies, the USDA forest service, NWI, and DOE. The tools and data that we are developing in partnership with NASA will improve resource management and policy decision support in the Carbon Management National Priority Application, with secondary emphasis on Agricultural (Forest) Efficiency and Ecological Forecasting. We focus upon advancing compatibility between current needs of federal and state managers, to include well-designed web gateways, user-friendly software, effective documentation, and tailored workshops. This project implements the SEEDS vision and the NewDISS paradigm, by serving needs of our user community, while advancing interoperability with ESIP Federation data systems. The solutions we are developing leverage the substantial public- and private-sector investments in remote sensing technologies and lead to measurable enhancements to resource management, and economic growth. Randolph H. Wynne (Virginia Tech) leads this project [14],[15].

6.5. Public health, mosquito vector-borne infectious diseases

This research extends, validates, and disseminates tools for epidemiologic surveillance of mosquito vector-borne infectious diseases and coastal environmental health using NASA ESE resources [16],[17]. The project supports NASA’s Public Health Applications Program through state and local public health activities. This scale is chosen for nurturing practitioners of surveillance and control of mosquito-borne infectious diseases (e.g., Eastern Equine and LaCrosse Encephalitis, malaria, and West Nile Virus.) The project also supports NASA’s Coastal Management Application program via relevance to environmental health of coastal bays (chronic coastal water quality problems such as harmful algal blooms, hypoxia, eutrophication, and bacteria.) This project uses ESE satellite observations (Landsat TM/ETM+, Terra Aster and MODIS, and TRMM) and commercially available IKONOS imagery for producing informational output and decision support to collaborating end-users in local to state-level public health and environmental management agencies.

6.6. Hyperspectral spectral analysis for chemical detection in the Chesapeake bay

In studies of oil spill detection in Chesapeake Bay, an analytical approach, spectral image analysis and a spectral library, was used [18]. In our work, field measurements and AVIRIS data will help us understand the relationships between water surface reflected spectra and water pollutant concentrations, as well as the change of water quality at different locations and on different dates. Analysis of two water quality parameters, chlorophyll-a (Chl) and suspended substance (SS), will measure the ratio of reflected spectra from the water surface. From a hyperspectral AVIRIS image with 224 channels from 400nm wavelength to 2500 nm wavelength, we can measure the reflectivity spectra of water surface parameters. By statistical correlation, the spectral analysis for absorption and reflection features, [of water quality spectra and water quality parameters], can be established for different pollutant concentration ranges.

Chemical detection techniques, using mid-wave infrared or long-wave infrared hyper-spectral imagery (HSI), generate massive data sets that complicate the detection and analysis of chemical plumes. In addition, HSI data sets tend to require long periods of time to extract useful information from a scene. In order to benefit from non-proliferation activities, new signal processing algorithms are needed to detect and classify target chemicals quickly in these massive data sets. This project develops two methods for digital signal processing algorithms that can provide substantial improvements in the identification and quantification
of chemical compounds from effluent plumes, using hyperspectral data. Foudan Salem (GMU) leads this project.

*Hyperspectral Image Analysis for Chemicals and Suspended Sediments Contaminants*

![Image](image.png)

Figure 6. Preliminary results for K-mean classification show the water polluted zone. The classification shows highly polluted water (appeared in red and dark blue) with vegetation damaged by chemicals and petroleum contaminants (appeared in yellow).

6.7. Information technology, interoperability

We are building the requisite information infrastructure by adopting information technology and geospatial interoperable standards to support the VA-MAGIC objectives, especially to support the NASA Geospatial Interoperability Office, following engineering practices. This will be accomplished through transparently, integrating NASA earth science gateway data [19],[20] and information into the research, education, and development process of VA-MAGIC as well as with VA-MAGIC user communities. Ruixin Yang and Phil Yang (GMU) lead the overall infrastructure.
The overall system design is illustrated in Figure 8, showing 3 relevant dimensions: The first dimension is the interoperable information and processing layer including the data inside the database, metadata for maintaining data and information, and specialized information processing components, such as decision support system, spatial analysis, data mining, geospatial information interoperable and other middleware components. Different portals and decision support systems will be built based on these components. The second dimension is the computing infrastructure dimension including computing resources such as CPUs, computers, storage, and the GRID that will be built. The third dimension is an application layer consisting of different applications such as those for a severe weather information center, agriculture, forest fire, and public health.

Building on GRID

An effort to transparently link all the processing, services, information, and data will be conducted and serve as a test-bed for integrating middleware with standard-based geospatial information processing toolkits. Intergraph Corp’s Web Mapping Service (WMS) and Web Feature Service (WFS) will be used as standard interfaces for generating images in specific formats (jpg, geotiff, png), and for generating features in GML - a standard format for point, line, and polygon datasets. In this way, we address the domain-specific challenges of geospatial inoperability [cf. 21,22]. Our system is based on existing middleware, Globus toolkit and SRB from San Diego Supercomputing Center (SDSC). In another project, GMU has been developing software for an Earth Science Data Grid System based on SRB. We have developed a complete data Grid system consistent with an SRB server and metadata catalog server (MCAT). The SRB server supports seamless access of data from distributed archives with a series of federated SRB servers.
Figure 8. The VA-MAGIC system with three dimensions: the application layer, computing resource layer and interoperable information and processing layer.

Figure 9. Operational MAGIC Portal provides geospatial information services for the Mid-Atlantic region.

WMS and WFS will be added to provide seamless access to geospatial information within the Grid system. GeoConnect from Intergraph will be added to provide a universal semantic, spatial, and temporal query mechanism. Simulation models will obtain preprocessed inputs from the Grid and will utilize the massive computing resources linked by the Grid. WMS and WFS kits from Intergraph will provide access services to the public through the VA-MAGIC portals and decision support applications, and also through the national geospatial data portals, (GeoData.gov), the National Map, FGDC Clearinghouse Node, and earth science gateway.
Acknowledgments

We acknowledge generous support for VAccess from NASA’s Earth Sciences Applications Directorate, John C. Stennis Space Center.

References


