IMAGe strategic plan

Introduction

The power of mathematical science is that similar methods and models can be used to solve scientific problems in very different contexts. The Institute of Mathematics Applied to Geosciences (IMAGe) was formed in October 2004 to develop mathematical tools, methods, and models that can address NCAR's fundamental science problems. IMAGe is also active in introducing the mathematics community to new problems that are posed by geophysical processes and observations. Some important vehicles that support these interdisciplinary activities are the Theme-of-the-Year and publically available software for numerical and statistical methods.

Strategic plan structure

This strategic plan is written to address IMAGe activities in a time frame of 1-5 years although most projects will have scientific impact and vitality well beyond this horizon. The format of this plan has been adopted to match a broader planning effort across NCAR structured around the following *themes*:

- *Imperatives*: Core activities that define the group.
- *Frontiers*: Important new activities for growth and emphasis.
- *Fabrics*: Attributes of a group that make it successful.

Each of the items in these themes has associated with it specific *actions* that can be measured or evaluated. Within NCAR measurable or well-defined progress on an action is termed a *metric* and where possible metrics for IMAGe actions are mentioned in this plan.

Connection to the laboratory plan

IMAGe is a division within the Computational and Information Systems Laboratory (CISL). The identity of CISL is a synthesis of three essential functions: a *computational* laboratory that provides services; a *scientific* laboratory that conducts research; and an *educational* laboratory that trains and mentors. Accordingly, IMAGe has an important role for scientific leadership and education within CISL and it is useful to coordinate this strategic plan with respect to the CISL plan.

The CISL strategic plan includes many specific actions that are associated with IMAGe but the imperatives, frontiers and fabrics are necessarily broader to encompass the entire laboratory. To create a crisper plan that outlines this Institute's specific vision and mission, the imperatives, frontiers and fabrics have been formulated specific to IMAGe and are not in one-to-one agreement with the CISL strategic plan. However, IMAGe based actions from the CISL plan have been migrated into the IMAGe themes and so a direct reference between these actions in both plans is maintained. It is also our intent to have these actions be the backbone for annual summaries of IMAGe such as the annual reports and the annual program operating plan.

Themes for IMAGe

With this introduction to the format, listed below are the guiding themes for IMAGe

- **Imperatives:**
    1) Develop new mathematical and numerical approaches that improve models for the Earth-Sun system.
    2) Develop mathematical methods for the interpretation and synthesis of geophysical data and model experiments.
    3) Engage the applied mathematics community in scientific problems of the Earth-Sun system including the interaction of human activities with the Earth's weather and climate. Engage the geoscience community in adopting mathematical tools and models that will improve models and the interpretation of data.
- **Frontiers:**
    1) Integrate multiscale modeling, numerics and computational science to meet the challenge of petascale computing.
    2) In partnership with the NCAR science laboratories build an earth system data assimilation, analysis and prediction system.
    3) Develop strategies to exploit the information in petascale data volumes.
- **Fabric:**

    Enhance the international scientific stature of the Institute through fundamental research in applied mathematics, statistics and physics.

These topics with introductions and specific actions are developed in the next sections. Actions that are drawn from the CISL plan imperatives and frontiers are references following the codes: C=computational, S=Scientific, E=Education.

IMAGe Imperatives

1) Develop new mathematical and numerical approaches that improve models for the Earth-Sun system.

Simulating geophysical processes with numerical models is critical to understanding complex relationships in the Earth system. Model development depends on efficient numerical methods, and as models become more complex, the underlying algorithms must be continually improved. In IMAGe, numerical methods research topics are chosen to address outstanding or anticipated problems in geophysical modeling. Much of this research is built on test-beds: codes or software environments for studying numerics on increasingly realistic problems.

Action 1 (S2.1): *Further develop the Geophysical and Astrophysical Spectral element Adaptive Refinement (GASpAR) code. GASpAR* is a research framework for solving partial differential equations using 2D and 3D scalable adaptive spectral element methods. Here *adaptive* means the numerical methods will modify the resolution and order of the solver based on the structure in the solution. Algorithmic innovations are planned that will enable the GASpAR platform to be used to study problems as diverse as: a) the planetary boundary layer in the day/night transition—in collaboration with the NCAR Meso- and Micro-scale Meteorology (MMM) Division, or b) the 3D Parker problem in astrophysics, a possible mechanism for solar coronal heating—via a proposed semi-implicit magnetohydrodynamics (MHD) solver. The key elements of these innovations are a) the development of a 2D and 3D Boussinesq solver that can automatically take advantage of the adaptive grid mechanics, and b) an optimized additive Schwarz preconditioner capable of working with non-conforming 2D and 3D elements used in the grid adaptation to accelerate simulation rates as well as enable simulations on staggered grids.

Action 2 (S2.3): *Further develop the High Order Method Modeling Environment (HOMME) dynamical core* for the Community Atmospheric Model (HOMME-CAM). HOMME is a highly scalable dynamical code that was developed by CISL/IMAGe as a foundation and test bed for a new generation of atmospheric general circulation models for the atmospheric science community. HOMME is also a vehicle to investigate using high-order-element-based methods to build conservative and accurate dynamical cores that efficiently scale to hundreds-of-thousands of processors, achieve scientifically useful integration rates, and can easily couple to community physics packages.

Plans for HOMME are to incorporate numerical schemes that conserve mass (e.g., discontinuous Galerkin) and evaluate this using the physical parameterizations and forcing functions from the NCAR Community Atmosphere Model (CAM). HOMME will also be used to compare other physical parameterization schemes within a common dynamical framework. One metric for HOMME will be its success in moving beyond a numerical code and making scientific contributions to the simulation of complex atmospheric processes. A companion goal is that HOMME will become one option for the atmospheric dynamical core in the NCAR Community Earth System Model. In complement, it is expected that HOMME will continue to serve as a test bed to engage computational scientists and applied mathematicians in migrating new numerical methods to a geophysical modeling context. These include issues of scaling, new horizontal discretisation techniques, implicit time stepping, regridding, and parallel implementations of model Input/Output. Part of its success will be to maintain a hierarchy of models that facilitate the testing of ideas from simple primitive-equation models up through an atmosphere coupled to land, ocean, and ice processes.

Action 3: *New approaches to numerical modeling based on adaptive bases.* While computer technology has advanced dramatically in recent years, numerical schemes currently used for climate and solar modeling fall drastically short of scientists' expectations. Spherical harmonics require large grids to resolve small features. Spectral element methods can resolve small features, but they require higher resolution near artificial boundaries to achieve high accuracy. Both methods are difficult or awkward to apply to irregular geometries. As a result, geoscientists and computational mathematicians are searching for new options. Radial basis functions (RBFs) offer the geosciences a novel numerical approach for solving time-dependent partial differential equations. We will focus on various areas of RBF mathematical development, from adaptive local node refinement to non-uniform fast Fourier transforms and filtering methods, to hybrid RBF-pseudospectral and RBF-finite difference schemes. Adaptive strategies for RBF solvers may benefit from statistical models where the adaptation is included as an additional numerical model and has tuning parameters, estimated from the current solution, that moderate the adaptively.

2) Develop mathematical methods for the interpretation and synthesis of geophysical data and model experiments.

Numerical models are used to make predictions and forecasts at many different geophysical scales. The research to support this activity involves methods for combining observational data with a numerical model, known as data assimilation, and statistical methods for summarizing and interpreting complex model output.

Action1 (C3.1): *Develop and support the Data Assimilation Research Test Bed (DART).* It is imperative that we maintain and improve the capabilities of the DART assimilation facility, particularly the interface to NCAR's weather and climate models. Existing capabilities will be ported to new generations of supercomputers and new algorithms developed by the research community for ensemble data assimilation will be incorporated. In addition, DART will be used to facilitate parameter estimation and serve as a computational platform for more the formal statistical analysis of large geophysical data sets and numerical models.

Action2: *Statistical models for interpreting spatial and spatiotemporal data*. Due to the heterogeneity, irregular sampling and scale dependence of geophysical observations it is usually difficult to make direct comparisons between observed processes and model simulations. Typically observations need to be extrapolated to regular grids or transformed through inverse modeling in order to provide process information that is easily interpretable. In addition, short numerical simulations can also produce variable climate estimates for intermittent process variables or rare events. Statistics plays a key role in managing the uncertainty when interpreting observational data or limited model experiments and often hinges on flexible models for spatial fields. Covariance models will be developed that blend the efficiency of spatial lattices with the flexibility of multi-resolution and/or compactly supported basis functions. One promising approach is to use a lattice model to induce dependence among the basis coefficients. Coupled with new spatial models is the need for efficient computational strategies to handle large data sets and work will be continued on tapering covariance functions and compactly supported covariance functions. Spatial models will be applied to regional climate experiments, paleoclimate reconstructions and the historical temperature and precipitation records. Adding a temporal component will focus on autoregressive models with spatially dependent innovations. This type of model will be explored in the context of the process layer in a Bayesian Hierarchical model for paleoclimate reconstructions.

The analysis of the distribution of extreme weather events will be continued using spatial statistic models to describe the dependence over space of parameters of the tail distribution. In addition, modeling of spatial extremes will be approached from a new viewpoint that considers discrete, extreme weather events that have a spatial footprint (for example, extreme precipitation from a particular storm system). The formal distribution of extremes is then interpreted as a compilation of the distribution of these space-time weather events. The practical goal is to handle short climate records and adjust for spatial dependence among the extreme values.

3) Engage the applied mathematics community in scientific problems of the Earth-Sun system including the interaction of human activities with the Earth's weather and climate. Engage the geoscience community in adopting mathematical tools and models that will improve models and the interpretation of data.

Each year IMAGe carries out a program, the Theme-of-the-Year (TOY) that focuses on a specific aspect of mathematics applied to the geosciences. It is designed to advance interdisciplinary research and education between the mathematical and geosciences communities in topics such as multi-scale modeling, data assimilation, turbulence and numerical methods.

Some metrics that will be applied to the TOY include:

- Number of first-time visitors to NCAR and number of early-career mathematical scientists who become engaged in NCAR scientific projects
- Success in recruiting prominent external co-directors
- Success in pursuing joint activities with the NSF Mathematics Institutes
- Development of other successful modes of interaction beyond workshops and lecture-based schools
- Participation in the TOY by other NCAR laboratories

The Geophysical Turbulence Program (GTP) is a longstanding but flexible activity that is administered through IMAGe but has a scientific membership that spans NCAR. The broad goal of GTP is to promote research, education and awareness of geophysical turbulence at NCAR and in the scientific community. This includes bringing visitors to NCAR, sponsoring workshops, and under some circumstances support for staff.

Two other programs administered and budgeted at the laboratory level but important to IMAGe are the Summer Internships in Parallel Computational Science (SIParCS) program offering graduate and undergraduate students significant hands-on opportunities in computational science, applied mathematics, and geostatistics and Research and Supercomputing Visitor Program (RSVP) allowing prolonged engagement and collaboration between CISL/IMAGe staff, the university community, and members of peer centers. RSVP has substantial flexibility and is open to students, early career scientists and established researchers.

Action1 (E1.1): *To put TOY on a secure financial footing over the next two years, TOY will operate at a smaller scale and will involve student training within the workshops rather than at a separate summer school.* This will include tutorial segments within the workshops, student presentations with peer feedback, and fielding several CISL summer internship projects. The TOY may also include visits to several graduate programs in applied mathematics, offering a mini-symposium to the students as insight into the numerical challenges specific to simulating geophysical processes.

Action2 (E1.3): *Solicitations for future TOY programs will encourage ever-broader participation of the mathematics community* through advertisements in mathematics and geophysical newsletters and also through emails to potential co-directors. Future TOY activities will be developed with an emphasis on improving interaction and learning through different formats. Some examples include multidisciplinary project teams, focused working groups, participation in field experiments, and short courses at universities.

Action3: *Expand the Geophysical Turbulence Program.* As geophysical models take advantage of larger computational resources and address processes at finer scales, turbulence will increasingly become an important issue in closures and parameterizations. It is expected that the role of GTP as a center wide focus on turbulence will become more important and along with the need to integrate fundamental research and geophysical modeling. In particular, it is important to create a GTP postdoctoral position that can serve as a bridge among GTP members and augment the existing shorter term GTP scientific visitors and workshops.

Action4 (E1.4): *Use the SIParCS, RSVP, and TOY programs as opportunities to better integrate CISL/IMAGe research and education opportunities.* SIParCS should add team research projects and projects with distributed mentorship—perhaps integrated with TOY themes or RSVP visits.

IMAGe Frontiers

1) Integrate multiscale modeling, numerics and computational science to meet the challenge of petascale computing

The capability to conduct numerical experiments at unprecedented scales provides an opportunity for mathematical tools and theoretical physics to be more directly involved in model development. Simulations that explicitly resolve geophysical processes or include detailed nonlinear dynamics will become the raw material for building accurate models of subgrid-scale behavior for medium-resolution models. Stochastic modeling will be important for representing the large number of degrees of freedom needed to capture the variability over space and time associated with nonlinear geophysical processes. Statistical analysis will also be important to streamline the series of model runs needed to develop parameterizations and closure schemes, and to summarize the complex output produced by multi-model experiments.

Action1 (F1.1): *Innovate statistical design and analysis techniques to improve the efficiency and accuracy of model development and testing.* The calibration and assessment of the complex computer models typical of the geosciences is a time-intensive and resource-consuming activity. By incorporating novel statistical design and analysis techniques, this scientific endeavor can be greatly enhanced through improved efficiency and by quantifying the uncertainty associated with estimated model parameters. There is need to develop test beds for these ideas as way to make the modeling community aware of the benefits of these ideas. Initial test beds may include a shallow-water version of CAM and the Lyon - Fedder - Mobarry model of the magnetosphere. Pilot studies will be conducted to assess the efficacy of recovering established parameter values using novel sequential designs and to assess techniques for choosing a limited but optimal collection of parameter values for test runs. As the research advances, this work will include higher-dimensional input parameter spaces and more complex model output fields.

Action2 (F1.2): *Apply stochastic models and direct numerical simulation to represent multiscale processes or multi-physics.* Multiscale physics is the hallmark of geophysical and astrophysical turbulent flows. The coupling among different scales in these flows must be modeled for accurate simulations of weather and climate. This is presently accomplished through, for example, nested regional climate modeling or through complex parameterizations. However, given the prospect of petascale and larger computational resources, there is the opportunity to use direct numerical simulations (DNS) as a numerical laboratory for studying the physics of scale interactions. Results from numerical experiments that resolve all relevant scales could serve as the basis for improving current parameterizations using physical theory but supported by statistical analysis and stochastic models. In particular, adding stochastic elements would provide a route to capture the nonlinear, temporal intermittency often associated with multiscale processes. Such modeling would be relevant to geophysical turbulent flows as they occur in the atmosphere and oceans of Earth, planets, and stars.

A starting point is modeling rotating (and stratified) flows such as in the Earth's atmosphere. For example, recent work suggests that the maximum intensity of a hurricane depends crucially on the (assumed) horizontal mixing length; this implies that an adequate treatment of the turbulence is essential in predicting important properties of a hurricane. Idealized cases are also useful for improving our understanding of these kinds of geophysical flows. The idea is that sufficiently high Reynolds number—sufficient multiscale interactions in turbulent fluids supporting inertial (or gravity) waves—is a desired ingredient for testing Large Eddy Simulation (LES) approaches. Two actions are contemplated to elucidate these issues:

- Analyze a high-resolution experiment of turbulence in a rotating flow for temporal behavior, scaling laws, structure formations, and for deciphering the wave-like versus eddy-like behavior. This will be used to justify the accuracy of a more efficient model using LES techniques. Stochastic variations of the deterministic closure will also be considered.
- Investigate stratified flow, in the presence of gravity waves, in bounded and unbounded domains, e.g., high-resolution simulation of the planetary boundary layer in the day/night transition as a stably stratified boundary layer becomes unstable and flow structures form. The energetics of these structures can be studied as a function of Richardson, Reynolds, and Rossby numbers. The statistical relationships of these

quantities will be used to motivate stochastic components in possible closures or parameterizations.

These two projects will create a basis of interaction between scientists in IMAGe, NCAR Geophysical Turbulence Program, and MMM that will be a fertile ground for training early-career scientists and provide a unique opportunity for new areas of collaboration.

Action3 (S2.2): *Enhance the Geophysical High Order Suite for Turbulence (GHOST) code.* A proven and highly scalable code for direct numerical simulations and modeling turbulence by solving the 2D and 3D Navier-Stokes and magnetohydrodynamics equations (with and without Hall terms) using the pseudo-spectral method. Several modifications are planned to enhance performance on petascale platforms and beyond. A pencil domain decomposition scheme will be adopted to more finely decompose the computational box for optimal processor distribution. Double-precision arithmetic will be implemented for the multidimensional FFTs and time stepping to allow accurate computation of small-scale fluctuations and spectra. In addition, experiments will be conducted with AltiVec™ technology to take advantage of high bandwidth that will be afforded by upcoming multicore systems. Finally, the output will be enhanced to provide a hierarchical data format required by the VAPOR package to analyze peta-scale datasets without time-consuming preprocessing.

Action4 (F2.3): *Accelerate applications algorithmically by developing new numerical methods, adaptive mesh refinement, new solvers, and new time integration schemes.* One approach to reduce the computational complexity of the simulation problem is by taking longer time steps or by using fewer grid points. To do this requires a paradigm shift in computations: instead of having the quality of a solution from a model driven by the uniform resolution achievable on a given computing resource, more flexible numerical methods will generate the best possible solution to a problem for any given computing resource.

The following innovations are needed to achieve this paradigm shift:

- To enable resolutions finer than about 5 kilometers for the atmosphere, non-hydrostatic governing equations coupled with an adaptive mesh need to be employed.
- The scale-dependence of physical parameterizations must be understood and represented.
- To integrate the model in time, procedures that are less dependent on spatial resolution will be required.

- Basic numerical research is needed to vary the grid resolution in an optimal manner. One promising area is uses the full adjoint of a model combined with an equi-distribution error estimator.

2) In partnership with the NCAR science laboratory build an Earth System data assimilation, analysis and prediction system.

Action (SF3.4): *Assimilate strategic, heterogeneous, and nonlinear observations into Earth System models.* Vastly expanded atmospheric observing systems and constantly improving numerical models have fueled a steady improvement in numerical weather prediction over the past five decades. Data assimilation—the combination of observations and model predictions to produce initial conditions for forecasts—has been the key to these improvements. To extend the successes of numerical weather prediction, data assimilation methods must be extended to deal with new challenges for Earth System model development and prediction. In particular, assimilating observations of the land surface, cryosphere, ocean, and biological systems will be essential. Earth System models and novel observations present challenges to traditional assimilation algorithms that assume linear and Gaussian relations between model variables and observations. Assimilation algorithms that can effectively extract information when this relationship is nonlinear and non-Gaussian are being developed by Data Assimilation Research Section scientists. These new assimilation techniques will not only lead to improved initial conditions for Earth System forecasts, but will also provide feedback for model development and the design of next-generation observing systems.

The first step will be developing algorithms that can deal with non-Gaussian distributions. This will improve assimilation of bounded observations such as atmospheric concentrations or Doppler radar reflectivity. A second step will allow nonlinear relations leading to improved assimilation of observations such as radiances that are related to model state variables in complicated ways. Although assimilation can be abstracted as determining the Bayesian posterior distribution of the system state given a prior distributions and observations, the practical implementation requires substantial approximations to be feasible for geophysical models and data. One challenge in this is to delineate how computational and discrete approximations match an optimal Bayesian solution.

3) Develop strategies to exploit the information in petascale data volumes.

Action1: *Analysis of numerical experiments based on multi-model ensembles.* Uncertainty in numerical models is often explored through the use of ensembles or collections of model runs that vary initial conditions, physics parameterizations, or even the models themselves. Identifying and quantifying the sources of uncertainty in these ensembles or delivering a unified product that essentially combines the information in these ensembles requires a deliberate statistical approach that will require new data analysis and computational tools to integrate the information in the ensemble. Recognizing that model output is typically a spatial, temporal, or spatial-temporal field, the functional analysis of variance approach currently being developed and based on Gaussian processes represents an excellent framework for in-depth analysis of the uncertainty in multi-model ensembles. This framework also allows for improvements in methodology for combining model output as well as probabilistic projections based on the ensembles. The focus on this work will include regional climate, in particular the North American Regional Climate Change Assessment (NARCCAP), as well as continued analysis of the model experiments submitted to Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). These multi-model ensembles comprise tens of terabytes of data and will drive new computational statistics approaches.

Besides lattice models, multi-resolution bases and reduced rank ideas there is the potential to exploit purely numerical algorithms as statistical approximations. One challenge on this frontier is to be ready for a large data volumes associated with the upcoming IPCC Fifth Assessment

Action2: *Highly scalable data assimilation.* Ensemble data assimilation algorithms consist of two alternating steps, first making a large set of short model forecasts, then adjusting the forecast model states using observations. The model advances are "embarrassingly parallel" since each forecast is independent of the others and can use all parallel features of the forecast model. Current algorithms for the adjustment step are moderately parallel, scaling well to perhaps 1000 processors. More scalable algorithms will be developed taking advantage of the fact that each state variable element can be updated independently given the increments generated for a particular observation. It is anticipated that these new implementations will scale to many thousands of processors. Novel algorithms in which observation increments can also be computed independently are being investigated but do not yet exist. Successful implementation would lead to assimilation algorithms that could scale to 100,000 processors or more.

IMAGe Fabric

The CISL strategic plan uses fabric themes to describe a workplace environment that promotes among other positive qualities collaboration, innovation and diversity. A different aspect in this plan is an acknowledgement that a national scientific center depends on its scientific staff being committed to basic and sometimes curiosity-driven scientific research as a complement to Institute and Center goals. In keeping this a core tenet IMAGe several different metrics are listed to measure its vitality.

Enhance the international scientific stature of the Institute through fundamental research in applied mathematics, statistics, and physics.

Action (S1.2 and S1.3): *CISL must also conduct applied mathematics research in geostatistics, numerical methods, geophysical turbulence, and data assimilation.* IMAGe's mission is to lead the mathematics and geophysical communities in ways that accentuate the contributions of mathematical methods and models to scientific progress in the geosciences. This role depends on IMAGe's scientific staff being individually recognized as experts and leaders in their own research areas. Most of the research activities in IMAGe will follow the lines of the current IMAGe sections; these activities will include directed and collaborative projects with other scientists plus independent exploratory research with less-immediate application.

To track the prevalence and quality of basic research among staff, IMAGe (and throughout CISL) will use more specific measures in addition to the NCAR-wide metrics used to assess scientific productivity such as publication quality, citations, and participation in professional meetings:

- The number and depth of collaborations of CISL staff with researchers outside NCAR.
- The extent that staff is invited to organize or participate in workshops and schools in applied mathematics and mathematical physics, and the frequency that CISL is solicited to host these synthetic activities between the mathematics and geosciences communities.
- CISL's impact on education, including number of students mentored, courses taught, and external degree committees supported; also, the number of CISL's students placed in multidisciplinary research positions.
- The extent that independent research projects eventually lead to transformative contributions within the geosciences.
- Identification of geophysical problems that motivate more general research in applied mathematics, statistics, and computer science.
- The extent that research modeling platforms are shared with and adopted by external groups.