Computational Challenges of the Applying Computationally Efficient Schemes for BioGeochemical Cycles (ACES4BGC) Project

Forrest M. Hoffman[§], Pavel B. Bochev[β], Philip J. Cameron-Smith[‡], Richard C. Easter, Jr.[α], Scott M. Elliott[*], Xiaohong Liu[α], Robert B. Lowrie[*], Donald D. Lucas[‡], Richard T. Mills[§], Timothy J. Tautges[†], Mark A. Taylor[β], Mariana Vertenstein[¶], and Patrick H. Worley[§]

[†]Argonne National Laboratory, [‡]Lawrence Livermore National Laboratory, [*]Los Alamos National Laboratory, [¶]National Center for Atmospheric Research, [§]Oak Ridge National Laboratory, [α]Pacific Northwest National Laboratory, and [β]Sandia National Laboratories

Project Goals and Objectives

- The goal of ACES4BGC is to advance the predictive capabilities of Earth System Models (ESMs) by reducing two of the largest sources of uncertainty, aerosols and biospheric feedbacks, utilizing a *highly efficient computational approach.*
- ACES4BGC will
  - implement and optimize new computationally efficient tracer advection algorithms for large numbers of tracer species;
  - add important biogeochemical interactions between the atmosphere, land, and ocean models; and
  - apply uncertainty quantification (UQ) techniques to constrain process parameters and evaluate feedback uncertainties.
- The objective of this SciDAC Partnership project is to deliver a second-generation ESM with improved representation of biogeochemical interactions at the canopy-to-atmosphere, river-to-coastal ocean, and open ocean-to-atmosphere interfaces.
- The resulting upgrades to the Community Earth System Model (CESM) will deliver new scientific capabilities, offer unprecedented accuracy in representing biogeochemical interactions, and yield *improved predictive skill and computational performance.*
- Significant computational challenges must be overcome to meet the scientific requirements for supporting large numbers of reactive tracers in CESM. The most significant challenge is presented by the advection of many tracers in the atmosphere and ocean component models.

Tracer Advection Using MOAB
Contact: Timothy J. Tautges (ANL)

- A computationally efficient and accurate tracer advection scheme is critical for transporting large numbers of reactive biogeochemical tracers throughout all component models of the ESM.
- The Community Atmosphere Model (CAM) utilizing the spectral finite element dynamical core from the High Order Method Modeling Environment (HOMME), referred to as CAM-SE, has unmatched parallel performance and is the preferred configuration for future CESM simulations (Taylor and Fournier, 2010; Evans et al., 2011; Dennis et al., 2011; Worley et al., 2011).

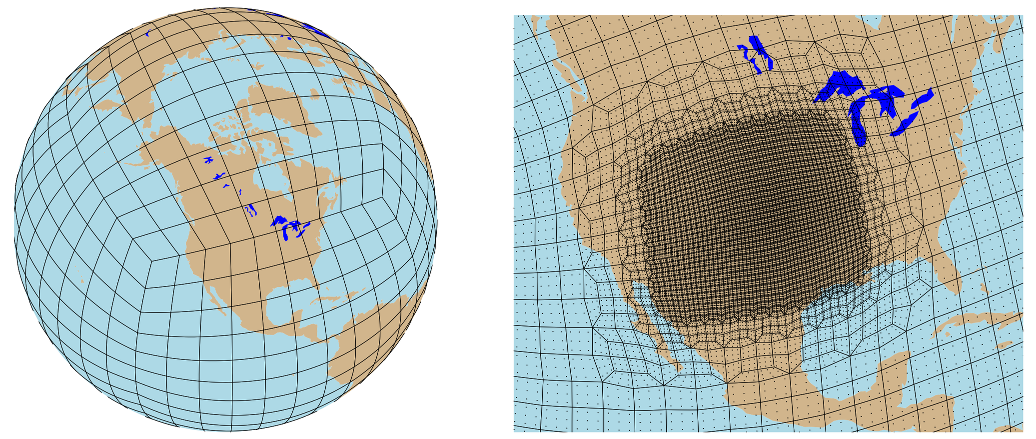

Figure 1: *An example of an unstructured CAM-SE variable resolution grid that increases 8-fold over the DOE Atmospheric Radiation Measurement (ARM) Program's Southern Great Plains (SGP) site.*

- The Model for Prediction Across Scales (MPAS) finite volume methods, utilizing Spherical Centroidal Voronoi Tessellations, offer dynamical cores for both future ocean and atmosphere component models (Thuburn et al., 2009; Ringler et al., 2010). The MPAS-Atmosphere is currently an option in CESM and the MPAS-Ocean is the preferred candidate for the next generate CESM ocean model.

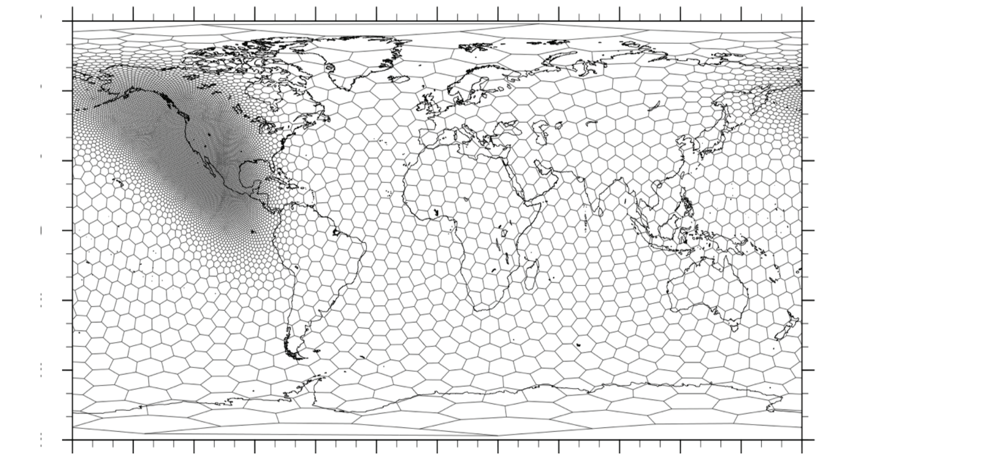

Figure 2: *A sample variable resolution grid for the MPAS Ocean component focused on the west coast of North America. Spatial resolution can be based on an arbitrary density function and focused along coastlines or in eddy active regions.*

- The backward-trajectory, semi-Lagrangian approach with conservative remapping of the Conservative Semi-Lagrangian Multi-tracer (CSLAM) method (Lauritzen et al., 2010) and the Characteristic Discontinuous Galerkin (CDG) method (Lowrie and Ringler, 2011) applied to unstructured grids offer promising techniques for computationally tractable advection.
- We will extend the work of collaborators J.-F. Lamarque and P. Lauritzen on CSLAM with new reconstructions to support unstructured grids like those shown in Figures 1 and 2.
- These methods allow for very large time steps, but require expensive geometric computation at each time step. However, this cost is independent of the number of tracers and is thus amortized over all the tracers included in the simulation.

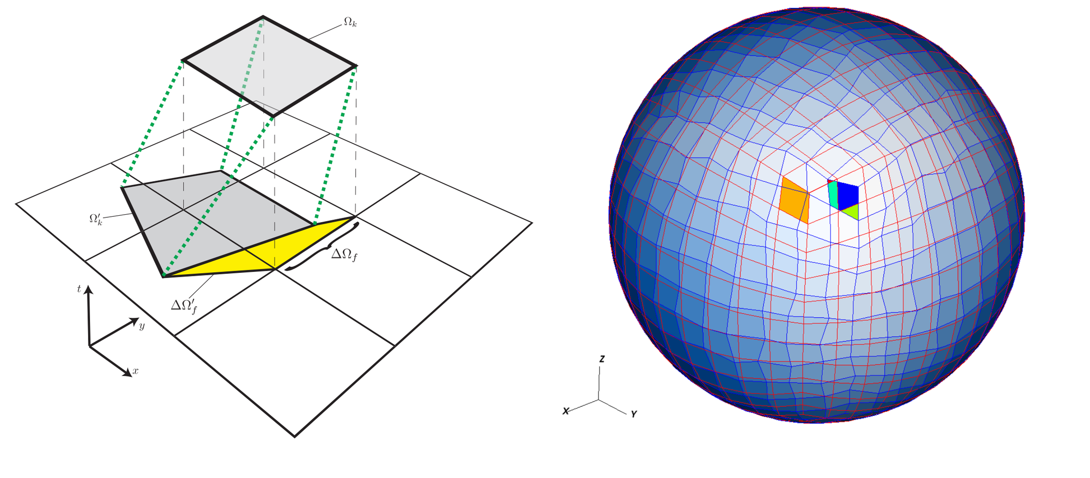

Figure 3: ***Left:*** *The spatio-temporal geometry of our characteristic approach used to update the tracer in cell Ω_k, over a time step. This example is for an underlying Cartesian grid, but the concepts extend to any grid topology. The region Ω_k' is the pre-image of the cell Ω_k, while $\Delta\Omega_f'$ is the region swept by a face $\Delta\Omega_f$.* ***Right:*** *A regular cubed-sphere grid (red) with a resolved pre-image grid (blue). Colored polygons show the intersection of a single cell on the two grids.*

- The *Mesh-Oriented datABase (MOAB)* technology, available through the **Frameworks, Algorithms, and Scalable Technologies for Mathematics (FASTMath) SciDAC3 Institute**, will provide parallel mesh infrastructure to extend CSLAM and CDG to unstructured, variable resolution grids while delivering efficient computational performance for the geometric computations required at each time step.
- Additional software engineering efforts in MOAB will provide a flexible interface and new features required for use by CAM-SE in the near term and by the MPAS-Ocean and MPAS-Atmosphere dynamical cores as the project progresses.
- Our goal is to produce a single tracer transport software module that is shared by MPAS and CAM-SE. CSLAM on the cube-sphere grid has already been implemented in CAM-SE, while MPAS is in the initial stages of testing CDG.

Verification, Validation & Uncertainty Quantification (UQ)
Contact: Donald D. Lucas (LLNL)

- The large number of biogeochemical-related parameters with uncertain values poses challenges to direct sampling of uncertainties and propagation of them through CESM simulations.
- Verification and validation (V&V) and uncertainty quantification (UQ) methods will be used to constrain model parameters based on observations and to understand the impacts of uncertainties on model projections for the terrestrial biosphere and for atmospheric chemistry.
- Standardized test cases and diagnostics employed by the CESM community will be adopted to quantify verification errors.
- For validating impacts on global water, energy, and terrestrial biogeochemistry, new metrics being developed by the International Land Model Benchmarking (ILAMB; http://www.ilamb.org/) project will be employed.

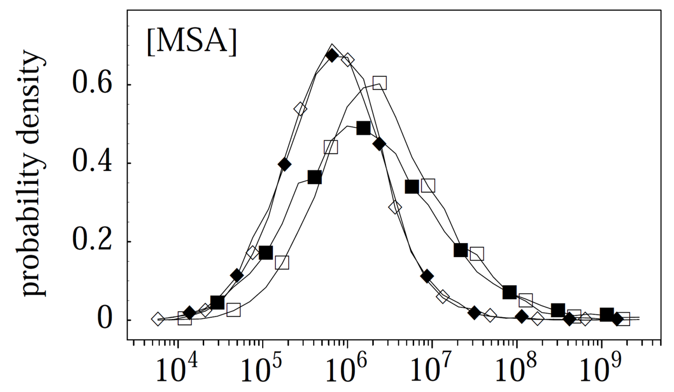

Figure 4: *PDFs of methane sulfonic acid (MSA) concentrations (molecules cm^{-3}) simulated with a detailed DMS chemistry model (open symbols) and polynomial chaos-based surrogate models (filled symbols) at two local times (squares = local noon, diamonds = 04:00). (From Lucas and Prinn, 2005)*

- In addition, we will apply advanced UQ methods to biogeochemical processes by
  - applying targeted schemes and utilizing the *DAKOTA Project tools*, developed by the **SciDAC Institute for Quantification of Uncertainty in Extreme Scale Computations (QUEST)** to sample parameter spaces,
  - decomposing and analyzing biogeochemical variances to produce probability distribution functions (PDFs) (*e.g.*, Figure 4),
  - performing dimensionality reductions,
  - constructing statistical surrogate models for biogeochemical processes, and
  - developing a model validation toolkit to optimize biogeochemical parameters using observational data sets (*e.g.*, DOE's Atmospheric Radiation Measurement (ARM) program, Next Generation Ecosystem Experiments (NGEE), and GoAmazon2014).

Software Engineering
Contact: Mariana Vertenstein (NCAR)

- ACES4BGC will follow the established software engineering standards for CESM development, coordinating with the head of the CESM Software Engineering Group (CSEG) at NCAR.
- New development will be performed on feature-specific code branches in the CESM revision controlled repository.
- CESM scripting will permit flexible and extensible incorporation of new biogeochemistry features, simplify testing of various model configurations, and support large numbers of UQ simulations.
- Working directly with CSEG staff, ACES4BGC will contribute all new model features to the CESM research community after they are tested, verified, validated, and reviewed.
- New model capabilities that meet with the approval of relevant CESM Working Groups and the CESM Scientific Steering Committee (SSC) will be offered for inclusion in future public releases of CESM.

Performance Engineering
Contact: Patrick H. Worley (ORNL)

- The ACES4BGC goal is to significantly improve model accuracy and skill by improving the representation of biogeochemical processes within CESM *without increasing the computational cost beyond practical limits.*
- Model development processes must include routine and accurate performance monitoring on relevant high performance computing systems, in particular DOE's Leadership Class supercomputers.
- ACES4BGC will leverage computer performance tools and technologies from the **SciDAC Institute for Sustained Performance, Energy, and Resilience (SUPER)**.
- We will monitor and optimize performance by
  - instrumenting code, deploying performance data bases and analysis tools, and establishing procedures for performance tracking;
  - routinely testing and tracking performance of new algorithms and model configurations;
  - developing optimized communications algorithms for new tracer schemes, particularly motivated by the expected large core count per compute node in the target systems; and
  - participating in end-to-end application testing and optimization for the next generation of CESM.

Project Summary

- ACES4BGC draws upon a diverse and multi-disciplinary team in the development of a second generation Earth System Model (ESM) supporting large numbers of biogeochemical tracers.
- In addition to the new model capabilities described above, other model improvements in the atmosphere, land, and ocean component models will add important biogeochemical interactions not presently captured in CESM. These aerosol and biogeochemical processes are outlined below.
- **Atmospheric Aerosols:** We will advance the representation of secondary organic aerosols (SOA) in CESM by
  - improving the treatment of SOA formation and aging based on the latest mechanistic understanding and evaluate against observation data (GOAmazon2014, GVAX, IMPROVE network, and the CAPT-aerosol capability);
  - implementing new mechanistic schemes for emission of volatile organic compounds (VOCs), POA, and other species;
  - apply UQ techniques to new schemes for OA to understand sensitivities and reduce uncertainties related to organics.
- **Atmospheric Chemistry:** We will improve the representation of organic chemistry by
  - calculating the rate of oxidation of VOCs into the condensable chemicals that form SOAs, which plays a key role in controlling aerosol and cloud droplet pH;
  - adding ammonia (NH_3), which plays a key role in controlling pH of aerosols and cloud droplets;
  - calculating the effect of emissions on the concentration of reactive greenhouse gases (CH_4, N_2O, HCFCs) and ozone depleting chemicals, which affect climate and air quality; and
  - constraining other model components through comparison with observations of related isotopic tracers (SF_6, ^{222}Rn, ^{210}Pb, OCS, and $CO^{18}O$).
- **Within the Canopy:** We will improve the representation of terrestrial biogenic emissions by
  - developing a canopy air space scheme supporting emissions of BVOCs and bi-directional fluxes of ammonia (NH_3);
  - developing and testing methods for reducing the range of uncertainty in BVOC emission factors, initially adding plant functional types (PFTs); and
  - evaluating emissions from dense woody vegetation against GOAmazon2014 observations under pristine and industrially polluted conditions.
- **Marine Chemistry:** We will improve the representation of marine organic chemistry by
  - identifying major classes of dissolved and particulate matter, and mapping compounds onto atmospheric species;
  - simulating dynamic distributions of chemical species across the surface ocean (due to grazing, ballasting, upwelling, photochemistry, heterotrophy, etc.);
  - providing OCS, NH_3, VOC, and aerosol emissions to the atmosphere; and
  - evaluating model performance using relevant data sets and traditional atmosphere-based kappa sensitivities.
- **River Transport and Ocean Coupling:** We will advance river-to-ocean biogeochemical cycles by
  - collaborating on development of tracer and nutrient transport schemes, building on a new two-way CLM/RTM coupling;
  - adapting ocean ecosystem dynamics to represent coastal zone processes;
  - combining CLM unstructured grid and variable resolution MPAS-Ocean to test river export and coastal zone biogeochemistry; and
  - evaluating model results against observations for the large Mississippi and Amazon basins.
- This five-year project is just starting, and we welcome additional collaborators and partners. Follow us on the web at http://www.aces4bgc.org/

References

J. Dennis, J. Edwards, K. J. Evans, O. N. Guba, P. Lauritzen, A. A. Mirin, A. St-Cyr, M. A. Taylor, and P. H. Worley. CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *Int. J. High Perf. Comput. Appl.*, 2011. to appear.

K. J. Evans, P. Lauritzen, S. K. Mishra, R. Neale, M. A. Taylor, and J. J. Tribbia. AMIP simulations with the CAM4 spectral element dynamical core. *J. Clim.*, 2011. under review.

Peter H. Lauritzen, Ramachandran D. Nair, and Paul A. Ullrich. A conservative semi-Lagrangian multi-tracer transport scheme (CSLAM) on the cubed-sphere grid. *J. Comput. Phys.*, 229(5):1401–1424, March 2010. doi: 10.1016/j.jcp.2009.10.036.

R. B. Lowrie and T. D. Ringler. The characteristic discontinuous galerkin method for tracer advection. to be submitted to *J. Comput. Phys.*, 2011.

D. D. Lucas and R. G. Prinn. Parametric sensitivity and uncertainty analysis of dimethylsulfide oxidation in the clear-sky remote marine boundary layer. *Atmos. Chem. Phys.*, 5:1505–1525, 2005.

T. D. Ringler, J. Thuburn, J. B. Klemp, and W. C. Skamarock. A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids. *J. Comput. Phys.*, 229(9):3065–3090, 2010.

M. A. Taylor and A. Fournier. A compatible and conservative spectral element method on unstructured grids. *J. Comput. Phys.*, 229:5879– 5895, 2010. doi: 10.1016/j.jcp.2010.04.008.

J. Thuburn, T. D. Ringler, W. C. Skamarock, and J. B. Klemp. Numerical representation of geostrophic modes on arbitrarily structured C-grids. *J. Comput. Phys.*, 228(22):8321–8335, 2009.

P. H. Worley, A. P. Craig, J. M. Dennis, A. A. Mirin, M. A. Taylor, and M. Vertenstein. Performance and performance engineering of the community earth system model. In *Proceedings of the ACM / IEEE Supercomputing SC'2011 conference*, 2011.