SCIENTIFIC COMPUTING

Computational Exposition

2018

Welcome to Computational Exposition 2018, an annual event that showcases the research of students in the Department of Scientific Computing (DSC). Although the research covers a broad spectrum of disciplines, they all concentrate on algorithm development and blend the computational and the mathematical to solve problems in the applied sciences. The innovation displayed is broad and remarkable. Our students make us proud!

The student posters reflect the breadth and depth of the research carried out in the DSC. They are the direct result of a fulfilment of our two most important missions: providing world-class interdisciplinary research and training in computational science.

Our graduate degree programs and the success of our students bolsters our confidence that we are becoming a premier institution for the training of the next generation of computational scientists. Indeed, the achievements of our current students over the past several years suggests that we are already there!

So, enjoy the presentation, interact with the students, challenge them, learn from them, and reflect on the fruits of their intelligence, skills, dedication and labor, and join us in thanking them for their contributions to the DSC, to FSU, and to science.

Gordon Erlebacher
Chair, Department of Scientific Computing

Cover: Quadrupedal robot LLAMA, courtesy of Mario Harper.

Presenting Researchers

Research

Brian Bartoldson
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Enhancing the Regularization Effect of Weight Pruning in Neural Networks

Artificial neural networks (NNs) may not be worth their computational/memory costs when used in mobile phones or embedded devices. Parameter-pruning algorithms combat these costs, with some algorithms capable of removing more than 90% of an NN's weights without harming the NN's performance. Removing weights from an NN is a form of regularization, but existing pruning algorithms do not significantly improve generalization error. We show that pruning NNs can improve generalization if pruning targets large weights instead of small weights. Applying our pruning algorithm to an NN leads to a higher image classification accuracy on CIFAR-10 data than applying the popular regularizer dropout. The pruning couples this higher accuracy with an 85% reduction of the NN's parameter count.

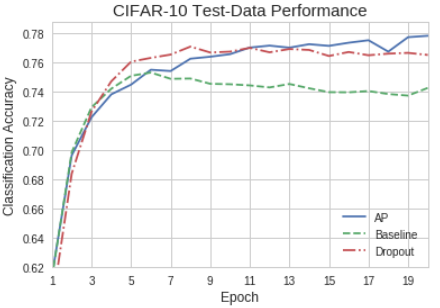


Figure 1. CIFAR-10 test-data accuracies for the baseline CNN, the baseline+dropout, and the baseline+AP. We plot averages of 10 runs for each CNN.

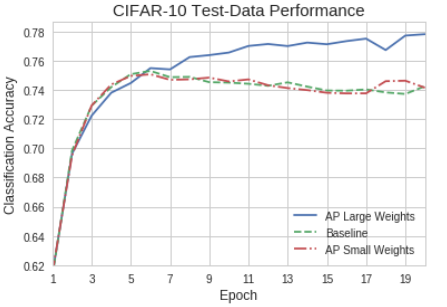
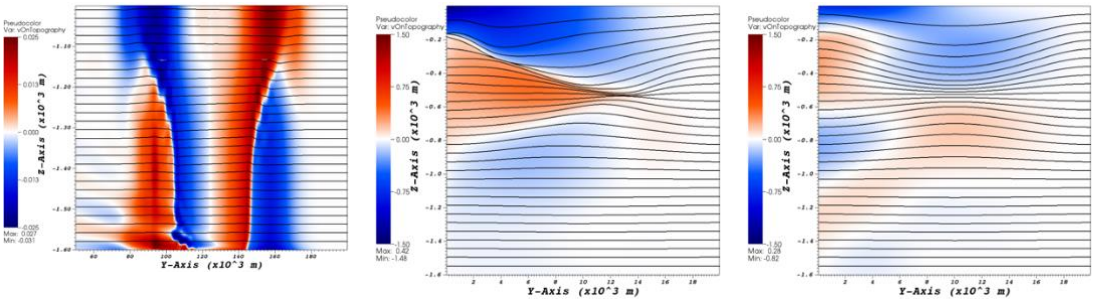


Figure 2: CIFAR-10 test-data accuracies are higher when AP targets the largest weights for pruning as opposed to the smallest weights.

Siddhartha Bishnu
Ph.D. in Computational Science
Advisor: *William Dewar*

Topographic Dissipation of Mesocale Energy

Energy is introduced into the oceans primarily at large scales by means of wind, tides and surface buoyancy forcing. This energy is transferred to the smaller mesoscale field through the geostrophic instability processes. The mesoscale field appears not to have accelerated appreciably over the last several decades, so we can assume that the mesoscale loses energy at roughly the same rate as it receives energy. Interestingly, how the mesoscale loses energy is not quite clear. We have been exploring topographic interaction as a pathway by which the mesoscale may lose energy to unbalanced forward cascading flows. To demonstrate this phenomenon, an approximate model theory is developed which consists of solving a reduced set of the momentum equations in density coordinates for any topographic configuration. The equations are solved using a high order spectral element technique and the results verified with already published MITgcm simulations.

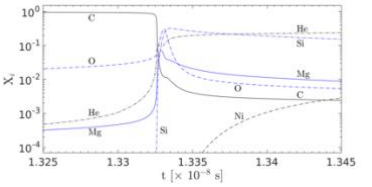


Figures: Topographic interaction modeling

Philip Boehner

Ph.D. in Computational Science
Advisor: *Tomasz Plewa*

Nucleosynthesis using Post-processing Techniques



Chemical evolution of stellar plasma is one of the most critical components of computational models in stellar astrophysics. Abundance distributions resulting from chains of nuclear reactions serve as a key comparison tool against observations, used to further constrain models. To that end, the present study focuses on improving the accuracy of model abundances. In most cases, abundances are obtained in the course of hydrodynamic simulations performed on Eulerian meshes. Unfortunately, those models are subject to the unphysical mixing of nuclear species due to numerical diffusion effects. For more reliable nucleosynthesis calculations, mass motions are described using passively advected Lagrangian tracer particles. These particles represent fluid parcels and allow to record their thermodynamic histories, which are subsequently used to drive detailed nucleosynthesis calculations performed with a large number of relevant isotopes.

Accuracy of nucleosynthesis calculations strongly depends on the accurate coupling between fluid represented on the Eulerian mesh and tracer particles. The coupling involves both interpolation of Eulerian data to particles as well as integrating equations of motion of particles. Both steps contribute numerical errors resulting in divergence of particle tracks from fluid streamlines. A number of particle advection schemes are discussed in terms of their accuracy and applicability towards particular needs.

Fig. 1, Above: Time evolution of nuclear mass fractions in a single zone model. At t=0, plasma is composed of carbon which, for the assumed density of ~4 x 10^7 g/cc and temperature of ~4 x 10^9 K burns explosively. Intermediate mass elements, such as Mg and Si, are produced at early times while He and Ni eventually dominate in the long term.

Ph.D. in Computational Science
Advisors: *Janet Peterson & Sachin Shanbhag*

A Contact-free Approach to Modeling Rigid Particle Suspensions

In everything from pharmaceuticals to ceramics, we often encounter particles suspended in viscous fluids. It is therefore very important to develop techniques that allow us to accurately and efficiently model these particle suspensions. Using boundary integral equations we can evaluate the hydrodynamic interactions between particles without meshing the fluid domain. By using fast summation techniques and efficient iterative methods with an appropriate preconditioner, we can solve the resulting equations in O(N) time. Once the hydrodynamic interactions between particles is known, the movements of the particles are governed by simple ODEs. When advancing these ODEs in time however, numerical error from either the boundary integral equation or the time stepper can cause nonphysical collisions between particles. Since spatial and temporal adaptivity can become expensive for dense simulations, repulsion forces are often used to avoid such collisions. We adopt a recent technique whereby the direction and magnitude of the repulsion force is determined by imposing a constraint guaranteeing that each time step is collision-free on the variational form of the Stokes equations. This allows us to simulate more dense suspensions without dramatically increasing the spatial discretization or reducing the time step size.

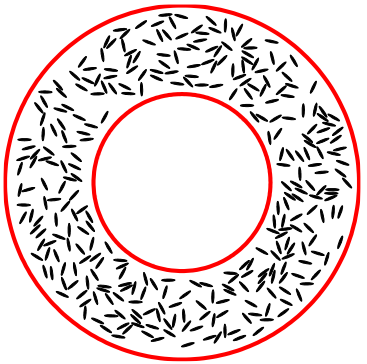


Figure 1, Left: Simulation of suspension of fibers inside a Couette apparatus.

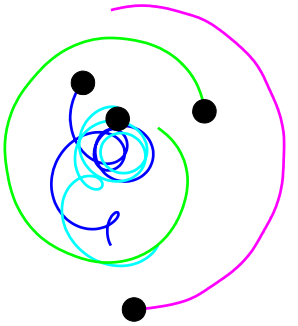


Figure 2, Above. Four rotors interacting hydrodynamically.

James Cheung
Ph.D. alumni in Computational Science
Advisor: *Max Gunzburger*

A Multifidelity Approach to Effectively Compute the Steady-State Flow of Iconic Solutions

We develop and demonstrate a new, hybrid simulation approach for charged fluids, which combines the accuracy of the nonlocal, classical density functional theory (cDFT) with the efficiency of the Poisson-Nernst-Planck (PNP) equations. The approach is motivated by the fact that the more accurate description of the physics in the cDFT model is required only near the charged surfaces, while away from these regions the PNP equations provide an acceptable representation of the ionic system. We formulate the hybrid approach in two stages. The first stage defines a coupled hybrid model in which the PNP and cDFT equations act independently on two overlapping domains, subject to suitable interface coupling conditions. At the second stage we apply the principles of the alternating Schwartz method to the hybrid model by using the interface conditions to define the appropriate boundary conditions and volume constraints exchanged between the PNP and the cDFT subdomains.

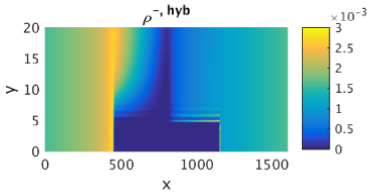


Figure 1, Above: The entire hybrid solution computed over the computational domain.

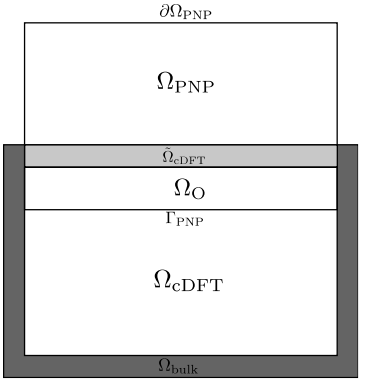


Figure 2, Right: A representative configuration for the coupled hybrid PNP-cDFT model.

Yu-Chieh Chi

Ph.D. in Computational Science
Advisor: *Chen Huang*

Efficiently Calculating RPA Energy Difference Between Two Similar, Large Systems Based on Atom-centered Correlated Sampling

Accurate prediction of binding energies between molecules and solids is one of the most challenging tasks in Kohn-Sham (KS) density functional theory (DFT). Commonly used local and semi-local exchange-correlation (XC) functionals often overestimate the binding energies due to the self-interaction error. They also cannot treat van der Waals interaction. Random phase approximation (RPA) correlation functional provides a promising route to improve the accuracy of binding energies by removing the self-interaction error by using the exact exchange, and, meanwhile, includes van der Waals interaction. Unfortunately, the computational cost of RPA is high and often scales with the fourth power of system sizes. We aim to develop a fast converging, stochastic method to directly compute the RPA correlation energy differences between two similar large systems. RPA correlation energy can be rewritten in terms of the density of states (DOS) of the product between Coulomb operator and the system's Kohn-Sham linear response function. The DOS is then stochastically evaluated by using the kernel polynomial method (KPM) by expanding it with Chebshev polynomials. The task to compute the RPA energy difference reduces to computing the difference of the moments between the two systems. In KPM, the DOS is rescaled to (-1, 1), and the maximum and minimum eigenvalues in the DOS is calculated using the conjugate gradient method. One basic operation is to multiply the KS linear response function with a vector, which is done with the density functional perturbation theory. We are in the progress of implementing this new method in the ABINIT program, an open-source, plane-wave based on DFT program. This method will be employed to obtain sufficiently accurate surface adsorption energies of molecules on solids to help us gain reliable understanding of heterogeneous catalysis with the atomic resolution.

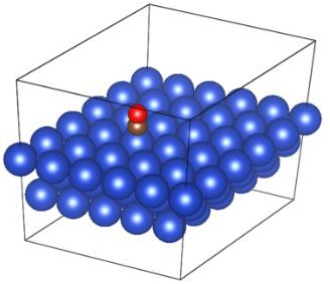


Figure 1: Method will be applied to obtain sufficiently accurate adsorption energies.

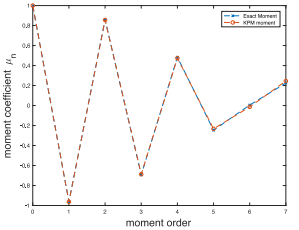


Figure 2: The moment result is obtained by using random vector with normalization.

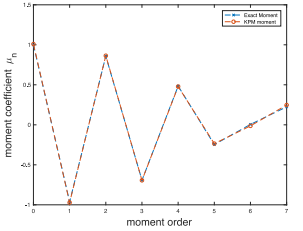


Figure 3: The moment result is obtained by using random vector without normalization.

Michael Conry
Ph.D. in Computational Science

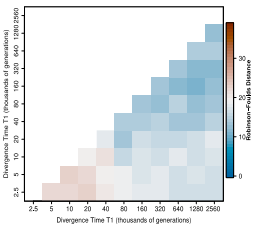
Advisor: *Alan Lemmon*

Determining Recombination's Impact on Phylogenetic Analysis

One of the central goals of evolutionary biology is to understand the evolutionary relationships among organisms by constructing phylogenetic estimates, commonly known as evolutionary trees. The accuracy of phylogenetic estimates can be strongly affected by the particular evolutionary processes that are taken into account during an analysis. One important process, genetic recombination, has been shown to lead to inaccurate phylogenetic estimates when ignored. Simulation studies measuring the accuracy of phylogenetic estimates in the presence of recombination have shown that when recombination is ignored, phylogenetic accuracy is reduced when divergence times are shallow and recombination is recent. Phylogenetic error due to recombination is expected to be greatest when all sequenced gene segments (e.g. exons) are simply concatenated and treated as one non-recombining unit, because the underlying history of these gene segments is not necessarily identical.

Although recombination within gene segments occurs and may negatively affect accuracy, most researchers choose to ignore intra-segment recombination due to the computational challenges with accounting for recombination within these segments. The question then arises: At what point is the advantage of having a longer gene segment outweighed by the negative effects of unaccounted for recombination within the segment? This question becomes especially important in phylogenetic studies relying on transcriptome data, in which the exons contained within each RNA transcript may be separated by long genomic distances across which recombination is very likely to have occurred. In these studies, researchers are faced with the question: when should exons be concatenated or treated as separate evolutionary units? This question is best addressed by a simulation study in which phylogenetic accuracy is studied in the context of the most relevant factors (i.e. recombination rate, gene segment size, population history).

Here, we describe a novel simulation study designed to determine the impact that genetic recombination has on phylogenetic analyses. In this simulation, exons that are exchanged between chromosomes and individuals are tracked in a set of populations that undergo recombination and speciation through time. Gene trees resulting from this simulation are compared to determine the degree to which recombination causes gene-tree species-tree discordance. Additional work will study this process in the context of exon length when gene trees are estimated from DNA sequence data that are simulated on the gene trees. Using this simulation, we will be able to provide recommendations to empirical researchers as to when it is most beneficial to treat exons as independent evolutionary units in phylogenetic analyses.



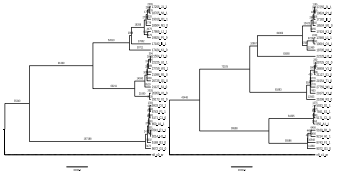


Figure 1, Far Left: This heat map shows the topological distance for 55 different datasets, each with varying divergence times. The cells in the matrix represent the average topological distance across all replicates at a recombination probability of 5.0e-4 per exon per generation.

Figure 2, Left: The two trees represent the varying evolutionary history for each exon undergoing recombination at a level of 0.5 per exon per generation.

Evan Cresswell
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Spatial Modeling of Intracellular Calcium Dynamics in Branched Astrocyte Processes

Several contemporary studies show that astrocytes, a type of glial cell, are fundamental to a variety of neural functions ranging from metabolic support to higher cognition such as recollection memory. This has led to the introduction of astrocytic dynamics into neural modeling. Most cellular functions in astrocytes are triggered by an increase or decrease in calcium concentration within the cytosol. Previous work treated astrocytic dynamics by representing calcium concentration as a point source or a completely spatial model in the cell. We now know that the role of the astrocyte takes many different perspectives. This work, which is inspired by in vivo recordings of
astrocytes in the ferret visual cortex, models the different levels of astrocytic calcium activity in the astro-neural system. In the model, we create a framework to enable the exploration of spatial calcium dynamics in astrocytes. Astrocyte processes are modeled as one-dimensional branches, over which we solve a system of reaction-diffusion equations for intracellular calcium dynamics. A branching structure, while not as general as a full 2D or 3D spatial simulation allows the study of astrocyte's cellular properties over extended regions of space. Studying a spatial representation of the processes will help investigate the role of astrocyte morphology in calcium signal propagation as well as developing intuition on the functional relationship between different levels of activity in the cell.

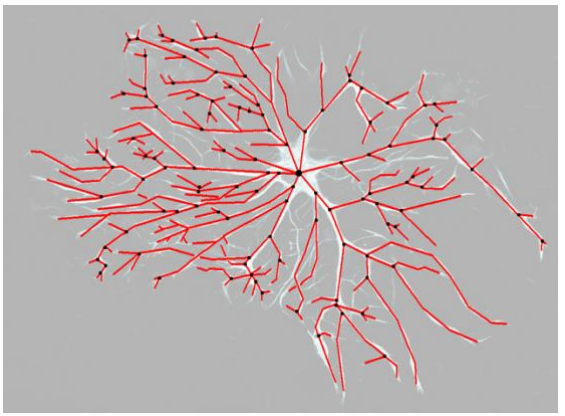


Figure 1: Branching Structure for Astrocyte.

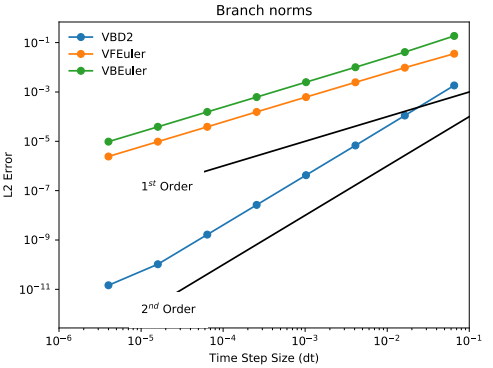


Figure 2: Code Order Verification.

Nathan Crock
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Correlating Brain State and Cell Function

Electroencephalogram (EEG) and Two-Photon microscopy (2P) data were simultaneously captured from the visual cortex of ferrets. The 2P data measures the calcium intensity of astrocytes and the EEG data measures the average neural activity (brain state) in the region. The objective of the experiment is to identify functional correlations between the brain state and astrocyte activity. Various unsupervised clustering algorithms were applied to the power spectra of the EEG waveforms to identify patterns of activity in the brain states. We started with naive algorithms such as k-means and hierarchical clustering, and moved to Gaussian Mixture Models (GMMs) and Bayesian Gaussian Mixture Models (BGMM). To determine the robustness of the clustering results we use the clustering output as ground truth and compare them with the successive trials of the same algorithm with different initial conditions. To quantify the similarity between trials we solved the assignment problem using the Hungarian Algorithm. To build intuition about each cluster's significance we reduced the dimensionality of using tSNE and interactively explored the projected data. We found two obvious clusters which we call the 'sync' and 'desync' state. There is evidence suggesting the existence of substates. We are continuing to explore this problem.

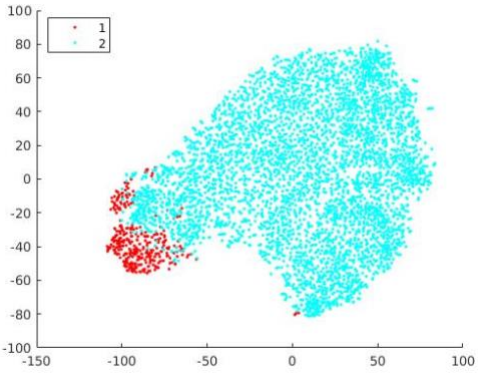
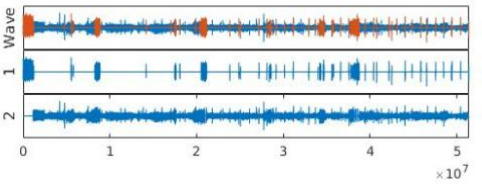


Figure 1, left: The data is first reduced from 5000 dimensions to 50 using principle components analysis. The data is further reduced to two dimensions using tSNE, for easy visualization. We can select nodes from this plot interactively (denoted in red) and visualize the actual 10 second snippet corresponding to each node.

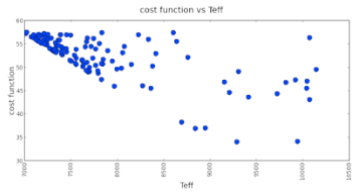
Figure 2, below: The 10 second snippet for each of the highlighted nodes in the tSNE projection is plotted in orange. This allows us to see if the clusters in the tSNE figure are representative of some clear physical property. The first thing noticed is that two clusters for 'sync' and 'desync' states are found.



Antigoni Georgiadou
Ph.D. in Computational Science
Advisor: *Tomasz Plewa*

Automated Parameter Fitting in Stellar Evolution

Stellar evolution presents one of the most challenging application areas of computational astrophysics. Computer modeling is extensively used to probe structure and evolution of stars and planets. Stellar evolution computations allow astrophysicists to connect the star formation process to astronomical observations of stellar objects. Because stellar evolution is a highly complex process, solving the inverse problem of matching the initial conditions to the observed objects requires substantial effort and expert knowledge. In this case, probing parameter space is done using a trial-and- error approach, which is inefficient, incomplete and prone to bias. Our aim is to identify the key model parameters, their most probable values, and their observationally-constrained ranges computationally in an automated way. We formulate a suitable constrained optimization problem, and combine select global optimization methods with the stellar evolution code, MESA. We demonstrate operation of the new stellar evolution optimization package in application to the evolution of intermediate and massive stars.



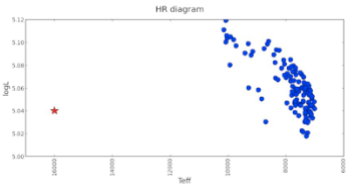


Figure 1, Far Left: The final cloud of a 5-parameter optimization run for massive stars using our parameter fitting tool.

Figure 2, Left: The HR diagram of the resulting population of stars after the optimization run.

Mario Harper
Ph.D. in Computational Science
Advisor: *Emmanuel Collins & Gordon Erlebacher*

Motion Planning for Dynamic Legged Robots and Unstructured Terrain Mobility

Successful motion planning requires understanding the relationship between the controllable robot inputs and the robot body velocity. In the case of fixed manipulators such as in a factory or assembly line, this mapping can be expressed by known mathematical equations. With mobile robots, such as wheeled and track-based platforms, the equations of motion governing the behaviors are still fairly accurate in a uniform structured environment; however, these equations start to degrade when operating in unstructured outdoor settings.

Dynamic legged robots traditionally are designed with speed and stability in mind. This has led to the development of stable gaits or locomotion patterns that allow the vehicle to move effectively. However, there are only a few cases where a mathematical model or simulation has given sufficiently accurate models of motion for planning purposes.

The goal of this research is to develop techniques to learn and adapt motion models to be effective in motion planning for dynamic legged robots and in difficult unstructured terrains. Developed technologies will be integrated on-board two different types of robots: the quadrupedal robot LLAMA (Simulated platform shown here, actual platform under construction) and the tracked vehicle RoMan (Robot Manipulator).

Figure 1: Quadrupedal robot LLAMA. This robot is being developed jointly by FSU, JPL and ARL through the RCTA program. This robot is capable of high speed locomotion and quasi-static stable motion.

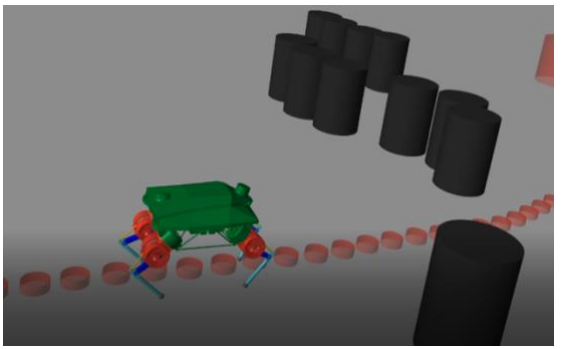


Figure 2: This autonomous track-based ground vehicle was developed by ARL, GDLS and JPL. FSU is currently building technologies to allow traversal of the robot through traversable obstacles.

Ryan Learn
Ph.D. in Computational Science
Advisors: *Tomasz Plewa*

Toward the Modeling of Astrophysical Magnetohydrodynamic Turbulence with Implicitly Modeled Braginskii Resistivity

In various astrophysical and high-energy density fluid flows, the evolution and behavior of the magnetic field can greatly influence flow morphology and result in dramatic transient phenomena. An example of such is the reconnection of magnetic field lines in the sun's corona, leading to solar flares and coronal mass ejections. Specifically in the case of reconnection, resistive destruction of the magnetic field is needed in order for the reconnection to take place, and thus the importance of accurately modeling these resistive effects cannot be understated. Many existing magnetohydrodynamic codes used in astrophysics often approximate this resistivity as being isotropic (as is the case in the Spitzer resistivity model), and then using an explicit time-marching scheme. Our work has been on constructing a computational model based on the anisotropic Braginksii resistivity formulation, that allows the magnetic field orientation and strength to effect the electrical conductivity and magnetic resistivity values throughout the plasma. For certain plasma configurations, the resistivity calculated through this approximation can become the dominant physical process at very small time scales, leading to a stiff system and necessitating and implicit solution to the magnetic induction equation. Outlined here are preliminary results of the Kelvin-Helmholtz interface instability under high-energy density conditions achieved through implicit solutions of our model.

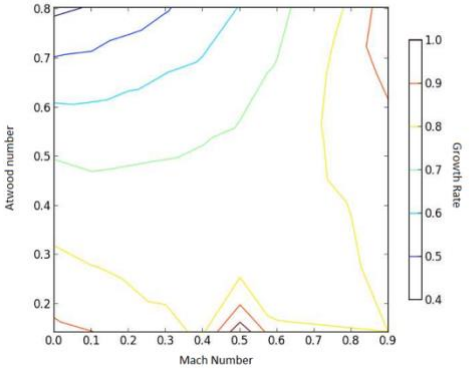


Figure 1: Normalized Growth Rates of the Kelvin-Helmholtz Instability as a function of stream velocity and density contrast

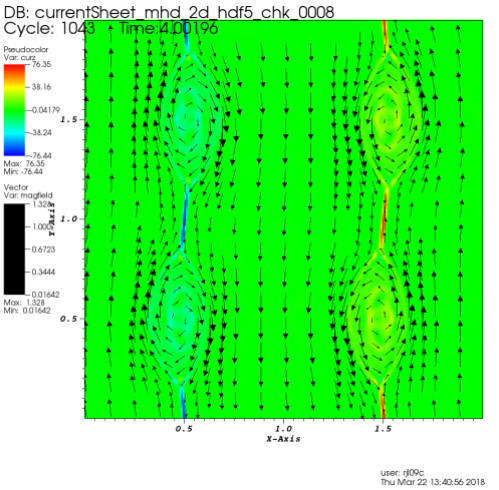


Figure 2: Magnetic field reconnection process in a plasma current sheet

Eitan Lees
Ph.D. in Computational Science
Advisors: *Sachin Shanbhag & Max Gunzburger*

Chemical Reaction Network Model Reduction with Applications to Nonlocal Corrosion

Many complex models involve robust chemical networks to capture the detailed behavior of a system. As a system grows in size there is a tradeoff between computational efficiency and model detail. Combustion models often contain hundreds of reactions and in the biological sciences they are sometimes trying to model thousands of interacting chemicals. When fast and slow reactions are included in the same model, the disparate timescales create stiff differential equations which are difficult to solve efficiently. Often these detailed systems are accompanied with large memory requirements. Methods to reduce model complexity while maintaining an accurate description of the system are of interest to many areas of research as well as industrial chemical engineers. We are working on methods for the reduction of complex chemical reaction networks. We are interested in approaching this problem from an optimization perspective. Future work includes building robust network reduction software and applying methods to previously studied corrosion systems.

Figure 1: Concentration profiles of a peridynamics corrosion model.

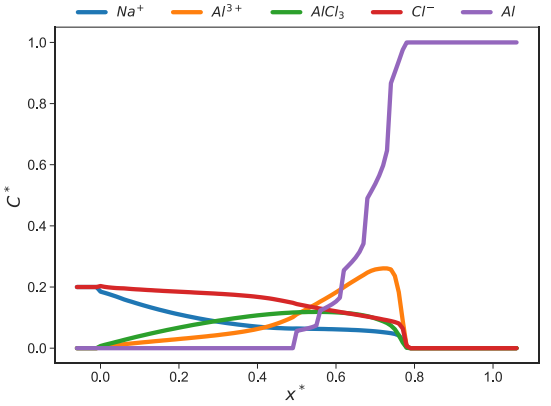
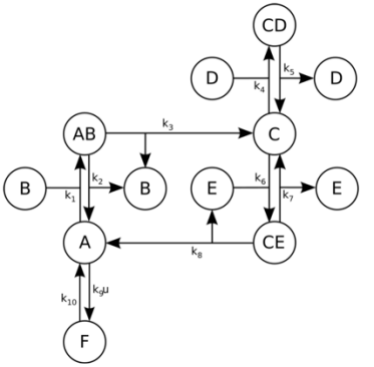


Figure 2: A network diagram for a nonlinear chemical kinetic system.



Isaac Lyngaas
Ph.D. in Computational Science
Advisor: *Janet Peterson*

Using RBF-Generated Quadrature Rules to Solve Nonlocal Models

In recent years, nonlocal continuum models have seen increased interest due to their ability to describe physical phenomena which are not modeled well by standard partial differential equation (PDE) models. These nonlocal models differ from traditional PDE models by using an appropriate integral operator instead of a differential operator. Two notable advantages nonlocal models have over typical PDE models are their ability to include discontinuous solutions in their formulations and their easy extendibility to anomalous diffusion models. Anomalous diffusion here refers to the case where the spatial spread of a diffusing quantity is not proportional to the square root of time as predicted by the heat equation.

In our work, we propose utilizing radial basis functions (RBFs) in order to derive quadrature rules for the integral operator in a nonlocal model. These quadrature rules are then used to derive a scheme for solving nonlocal models. This method is considered for its ability to easily extend to problems in higher spatial dimensional domains and its ease of use for problems that are defined on a scattered grid. In particular, we consider using this RBF quadrature method to solve nonlocal models that use an integral operator in place of spatial differential operators typically seen in PDE models. The models that we consider include nonlocal models for diffusion and anomalous diffusion equations, a nonlocal model for the advection-diffusion equation, and a nonlocal model for Burgers' equation. Numerical experiments solving these models are provided to show the accuracy of the RBF quadrature method.

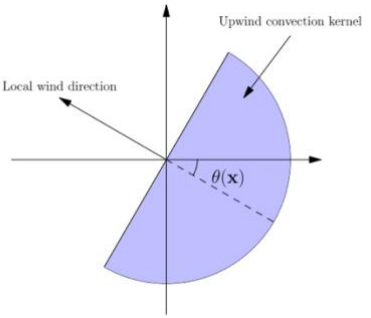


Figure 1. The upwind convection kernel used for the nonlocal convection integral operator.

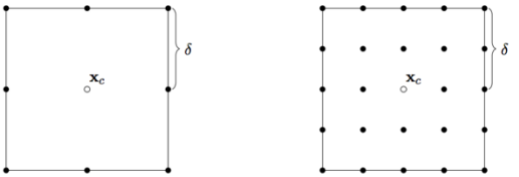


Figure 2. Discretization of the nonlocal integral in two dimensions.

Marcelina Nagales
Masters in Computational Science
Advisor: *Dennis Slice*

Three-dimensional Scanning Methodologies in Digital Reconstruction

This research project focuses on the comparison of different methodologies of three-dimensional scanning to digitally reconstruct Artifacts/Art Objects. This study used Flex Scan 3D scanning software and a rotary table in combination with an LMI HDI 109 series table top laser scanner with two cameras to take a stereoscopic image to create the mesh. The project specifically focuses on the scanning methodologies in regards to projectile points/knives from the Southeastern United States and the John House Stereograph Collection. The projectile points/knives were provided by the Florida Bureau of Archaeological Research and the John House Stereograph Collection was provided by the Department of Art History at Florida State University. Projectile points/knives are known for their influences in calculating the time periods of different archaeological sites and their digitization is key for analysis of the general morphology of projectile points/knives from different regions. The John House Stereograph Collection primarily documents Paris, Versailles, and other regions of the world. The digitization of the images of architecture and the history of the stereograph cards themselves will allow for a potential interactive museum exhibit experience and the preservation of architecture that is no longer here today. While both materials are different in origin and historical significance, they share similar morphologies that create issues in the scanning process, such as edges and small intricate details. They also have individual features that require specific methods for proper digital reconstruction. The three-dimensional scanning methodologies in digital reconstruction and the Three-Dimensional Digitization of the John House Stereograph Collection and projectile points/knives of the Southeastern United States will help pave the way for virtual preservation.

Figure 1: The left is a Citrus point and the right is the 3D reconstruction of the same Citrus point

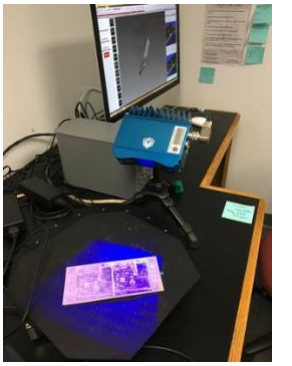
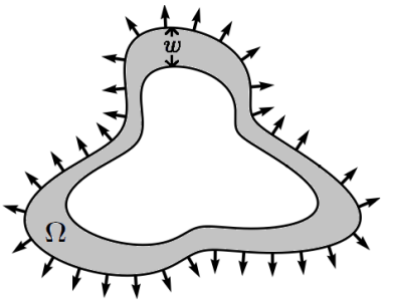


Figure 2: The LMI HDI 109 series table top laser scanner in use for digital reconstruction of a stereograph from the John House Stereograph Collection

David Robinson
Ph.D in Computational Science
Advisor: *Bryan Quaife*

Development of a Computationally Efficient Wildfire Model

Wild and prescribed fires are complex multi-physics and multi-scale systems, and accurate simulations that capture observed behavior require large amounts of computational resources. We propose a simplified model that represents the fire as a pair of closed one-dimensional interfaces in two-dimensional space. The dynamics are governed by a normal velocity that is a weighted sum of terms for the heat flux, convective sinks, and the ambient wind velocity. The weights for each contributing factor will be determined via Bayesian inference methods using both experimental data and high fidelity model runs from HIGRAD/FIRETEC. The Delayed Rejection Adaptive Metropolis (DRAM) algorithm is used on a related advection diffusion equation to infer coefficients from synthetic data. To form stable long horizon simulations, we discretize the interface with a spectrally accurate Fourier method, and a well-developed method is used to avoid any tangling of our clustering of the grid points. The goal is to capture fire line behaviors that have been observed in other models and experiments, but with a much smaller computational cost.



Top Right: Schematic of the 2D fireline model showing the interior and exterior boundary of a developing fire.

Bottom Figure: Setting up a fire box for a heat flux measurement at the Prescribed Fire Science Consortium.

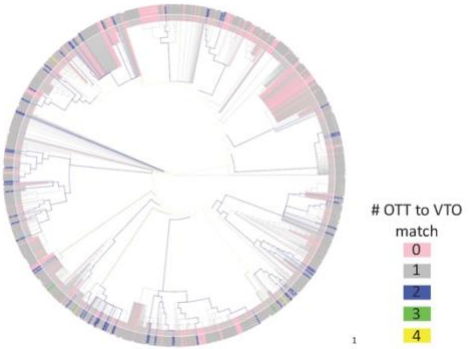
Marjan Sadeghi

Ph.D. in Computational Science
Advisor: *Peter Beerli*

Integrating Different Types of Data for the Tree of Life Project

The Open Tree of Life Project tries to identify all species on this planet and is housed at http://www.tree.opentreeoflife.org/. Many other projects collect data about species, such as Phenoscape, Taxonomic Name Resolution Service (TNRS) and NCBI. Phenoscape is a project that collects phenotypic data. NCBI is a genomic database that stores all sorts of DNA sequence information for species. TNRS is also a data set which collects plant names. Integrating all this data into one resource is difficult because these projects use different reference systems. For example, species names change over time and thus need to be carefully matched among different projects. Inclusion of fossil data seems particularly difficult because often the name is not certain. I joined a workshop and it aimed to discuss solutions to this problem. We were working on codes to assemble the Phenoscape and NCBI data into the tree of life.

We test our code with a specific species, catfish.



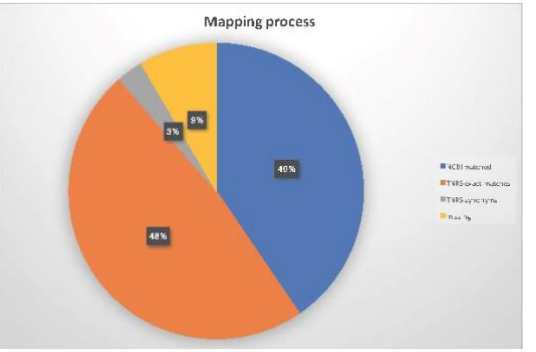


Figure 1: The figure shows the OTT (Open Tree Taxonomy) which matches VTO (Vertebrate Taxonomy Ontology) on catfish Open Tree synthetic phylogeny.
Figure 2: The figure shows the final reconciliation. Almost half the data matches with TNRS names and only a small part of the data has been missed.

Michael Schneier
Ph.D. in Computational Science
Advisor: *Max Gunzburger*

An Ensemble-Leray Proper Orthogonal Decomposition Method for the Nonstationary Navier-Stokes Equations

The definition of partial differential equation (PDE) models usually involves a set of parameters whose values may vary over a wide range. The solution of even a single set of parameter values may be quite expensive. In many cases, e.g., optimization, control, uncertainty quantification, and other settings, solutions are needed for many sets of parameter values. We consider the case of the time-dependent Navier-Stokes equations for which a recently developed ensemble-based method allows for the efficient determination of the multiple solutions corresponding to many parameter sets. To further reduce the costs of determining multiple solutions of the Navier-Stokes equations, we incorporate a proper orthogonal decomposition (POD) reduced-order model into the ensemble-based method. The stability and convergence results for the ensemble-based method are extended to the ensemble-POD approach.

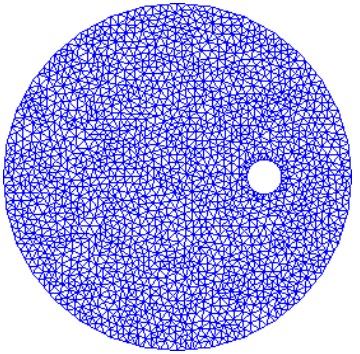


Figure 1: 16000 degrees of freedom mesh which we use in the calculations of the ensemble POD algorithm.

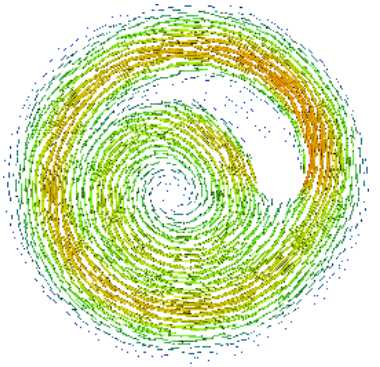


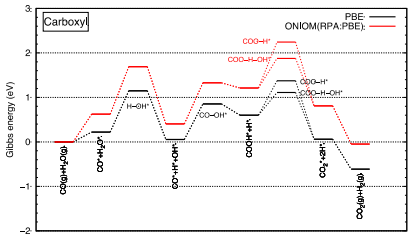
Figure 2: The velocity field of our ensemble-POD model.

Maliheh ShabanTameh
Ph.D. in Computational Science
Advisor: *Chen Huang*

Mechanism of Water Gas Shift Reaction on Cu(211): An Investigation Based on Random Phase Approximation Correlation Functional

Water-gas shift reaction (WGSR) ($CO\ (g) + H_2O\ (g) \rightleftharpoons H_2\ (g) + CO_2\ (g)$) is of great importance in industry for producing hydrogen. The catalysts are often based on copper and the catalytic process is exothermic at low temperature. Two different mechanisms were proposed in the past: redox and carboxyl mechanisms which involve complete and partial dissociation of water, respectively. To gain insight into WGSR with atomistic resolution, density functional theory (DFT) was extensively used to simulate this reaction. Most DFT studies were based on generalized gradient approximation (GGA) exchange-correlation (XC) functional, which often overestimates the adsorption energies of molecules on transition metals, therefore casting doubts on the predicted reaction energies and kinetics. In this work, we investigate WGRS reaction on copper surface by using high-level random phase approximation (RPA) XC functional. The results are compared against GGA functional to assess the accuracy of DFT-GGA for investigating WGSR.

Microkinetic modelling was performed at the pressure of 1 bar and the temperature of 473.15 K, with a syngas consisting of 7% CO, 8.5% CO_2, 22% H_2O, and 37% H_2 (balance inert). We investigated the reaction rates and surface coverages of intermediates as a function of pressure and temperature. Both PBE and RPA predict a very low coverage of carboxyl for all pressures. PBE predicts that the rate limiting step is the breaking of the O-H bond of water. In addition, RPA suggests that the decomposition of adsorbed OH is also the rate limiting step. PBE suggests that the carboxyl mechanism through the breaking of CO-OH bond is the favored path, while RPA predicts that the redox mechanism is the most favored path. By increasing the pressure, RPA shows that the turnover frequency increases, while PBE predicts the opposite trend. Our results suggest that highly accurate quantum mechanics method is required to reliably model WGSR in which competing mechanisms such as Redox and Carboxyl exist.



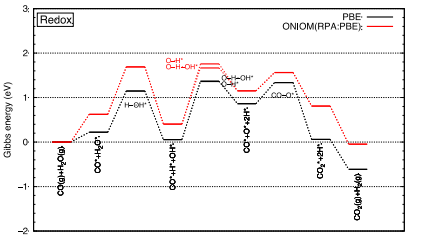


Figure 1. Gibbs free energies obtained from ONIOM(RPA:PBE) and PBE for the Carboxyl mechanism over the Cu (211) at a temperature of 473.15 K, 1 atm total pressure, 7% CO, 8.5% CO_2, 22% H_2O, and 37% H_2 (balance inert).

Figure 2. Gibbs free energies obtained from ONIOM(RPA:PBE) and PBE for the Redox mechanism over the Cu (211) at a temperature of 473.15 K, 1 atm total pressure, 7% CO, 8.5% CO_2, 22% H_2O, and 37% H_2 (balance inert).

Kyle Shaw
Ph.D. in Computational Science
Advisor: *Peter Beerli*

Understanding the Complexities of Tree Space

Finding the correct phylogenetic tree given genetic data can be very complicated and cumbersome. The number of dimensions to search increases as a double factorial of the number of samples. I studied two aspects of tree space that allow us to better understand and simplify our search space. First, I explored two priors of tree topology commonly used in Markov Chain Monte Carlo (MCMC) method including equiprobable, and random joining. The prior used determines the focus of the MCMC search; the random joining prior limits the algorithms' ability to search unbalanced trees. Second, I explored the effect of branch lengths on the likelihood. The length of a branch is fundamental to evolutionary processes. A clade with short branches implies that all samples from that clade are closely related, whereas a long branch would signify much difference between samples. Thus branch lengths have a great effect on the likelihood of a tree. Combining branch lengths and the set of tree topologies we obtain our search space. There is a potential for many smooth local maxima throughout tree space making it difficult for MCMC to find the maximum likelihood. Using a simple three taxon tree this space can be visualized with a three dimensional graph.

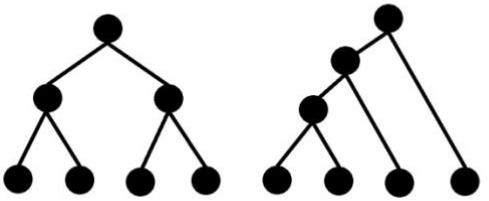


Figure 1: A balanced tree that would be more likely to show up in a random joining scheme, vs an unbalanced tree that would be more likely to show up in an equiprobable scheme.

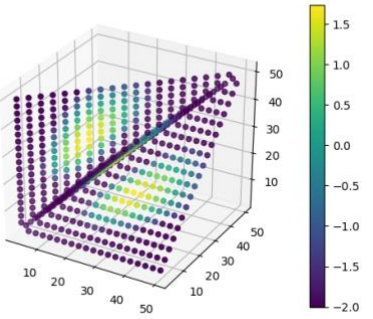


Figure 2: The tree space of a three tip tree. The three 2-dimensional planes represent the three different topologies where each point represents a different combination of branch lengths. Color is used to represent each point's likelihood.

Danial Smith
Ph.D. in Computational Science
Advisor: *Gordon Erlebacher*

Creating Educational Game Levels in a Co-creative Environment

eRebuild is a learning game initially created as a math learning game that integrates data-based assessment. As the game develops into a robust learning platform, more content becomes necessary. eRebuild will employ a number of techniques to create this content. We, as designers, will create additional levels, but the amount of content necessary for personalized learning is more than one team can create by hand. Levels will be created procedurally, but there is no guarantee these levels will target the desired math concepts. To keep the content focused on learning, we enlist math teachers as level designers. By combining our naïve procedural systems with a level editor, a mixed-initiative co-creative environment is created. Even the complete novice level designer can create a targeted learning experience with minimal effort. Each level will be categorized by the instructor after creation with a difficulty and the concepts the level focuses on. As the level library grows, our naïve level generator can learn from the instructors' designs. Ultimately, this will lead to better starting points and full autonomous level generation. A unique experience tailor-made for each student should be available give levels from all three sources: game design team, user made, and procedurally generated.

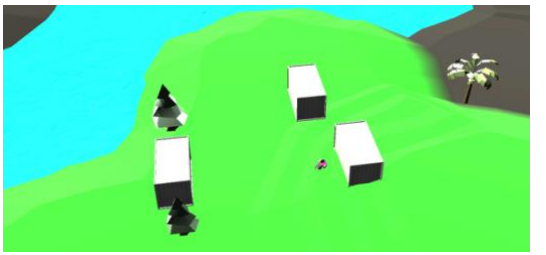


Figure 1, Left. By selecting which competencies they would like to focus on, the level editor creates a task and landscape to be edited by the instructor.

Figure 2, Above: The level created is barebones but fully functional. The instructor then modifies the play space and the narrative to fit their students.

Amirhessam Tahmassebi
Ph.D. in Computational Science
Advisor: *Anke Meyer-Baese*

Determining Disease Evolution Driver Nodes in Dementia Networks

Imaging connectomics emerged as an important field in modern neuroimaging to represent the interaction of structural and functional brain areas. Static graph networks are the mathematical structure to capture these interactions modeled by Pearson correlations between the representative area signals. Dynamical functional resting state networks seen in most fMRI experiments can not be represented by the classic correlation graph network. The changes in brain connectivity observed in many neuro-degenerative diseases in longitudinal data series suggest that more sophisticated graph networks to capture the dynamical properties of the brain networks are required. Furthermore, certain brain areas seem to act as "disease epicenters" being responsible for the spread of neuro-degenerative diseases. To mathematically describe these aspects, we propose a novel framework based on pinning controllability applied to dynamic graphs and seek to determine the changes in the "driver nodes" during the course of the disease. In contrast to other current research in pinning controllability, we aim to identify the best driver nodes describing disease evolution with respect to connectivity changes and location of the best driver nodes in functional ^{18}F-Fluorodeoxyglucose Positron Emission Tomography (^{18}FDG-PET) and structural Magnetic Resonance Imaging (MRI) connectivity graphs in healthy controls (CN), and patients with mild cognitive impairment (MCI), and Alzheimer's disease (AD). We present the theoretical framework for determining the best driver nodes in connectivity graphs and their relation to disease evolution in dementia. We revolutionize the current graph analysis in brain networks and apply the concept of dynamic graph theory in connection with pinning controllability to reveal differences in the location of "disease epicenters" that play an important role in the temporal evolution of dementia. The described research will constitute a leap in biomedical research related to novel disease prediction trajectories and precision dementia therapies.

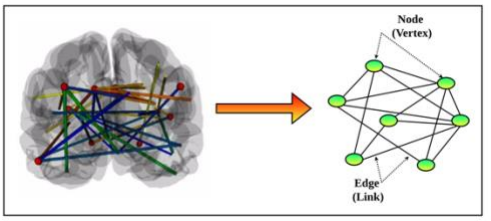


Figure 1, Above: Schematic illustration of unweighted-undirected graph of complex networks in brain. Nodes or vertices can be brain regions or voxels. Edges or links are the functional or structural connections between nodes.

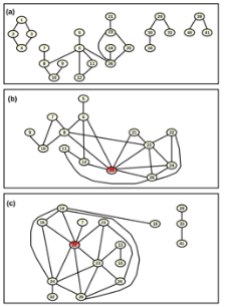


Figure 2, Right: The most influential driver nodes (shown in red) in functional connectivity graphs for (a) CN, (b) MCI and (c) AD.

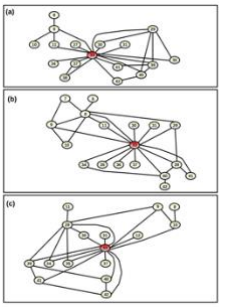


Figure 3, Far Right: The most influential driver nodes (shown in red) in structural connectivity graphs for (a) CN, (b) MCI and (c) AD.

Stephen Townsend
Ph.D. in Computational Science
Advisor: *Dennis Slice*

A Comparison of Methods for Analyzing Spatial Variation in Animal Vocalizations

Animal vocalizations are a very important and widely studied area in modern ecology. The calls and vocalizations of many different species from all across the tree of life have been studied. It is well documented that there exist patterns of spatial variation in these vocalizations. In some cases, the differences are so pronounced that it is possible that they may play a role in reproductive isolation and speciation. Here we compare two computational methods for analyzing spatial variation in animal vocalizations using frog calls as a model organism, specifically, the mating calls of the Cope's Gray Treefrog (*Hyla chrysoscelis*). Neither of these methods rely on spectrograms and both are fully computational requiring little pre-processing work on the part of the researcher. The first of these

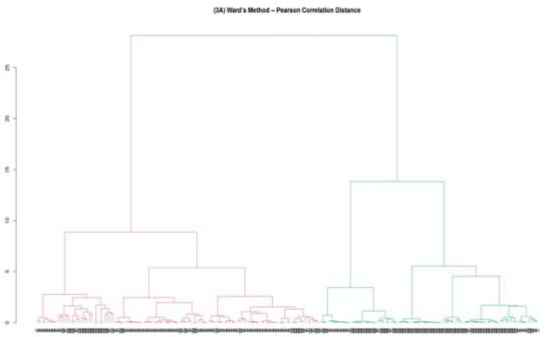
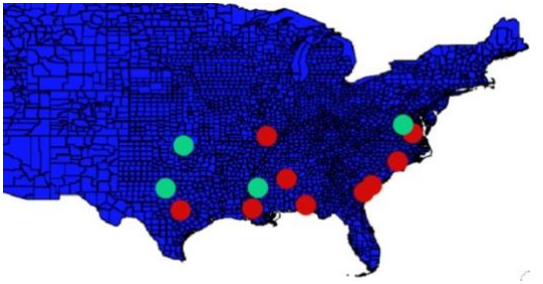


Figure 1: The figure above is a complete clustering of all 187 frog calls using Ward's Method with Pearson Product Moment Correlation as the distance measure. The green group is the Western Dialect while the red group is the Eastern Dialect.



methods is a previously developed Single-Pulse Analysis method which uses time series clustering to group animal vocalizations using Ward's Method Hierarchical Clustering. The second method uses Random Forests and Voting Classifiers to determine the best classification of vocalizations across geographic space. Random Forests and Voting Classifiers are machine learning methods widely used across science and engineering. To compare these two methods, a sample of 187 mating calls of the Cope's Gray Treefrog (*Hyla chrysoscelis*) were downloaded from the Macaulay Library. These calls were then filtered and aligned to ensure shape was the only variable. Using this data, we show that both of these methods are reliable, novel approaches to analyze spatial trends in bioacoustic data.

Figure 2: The above figure shows the results of the Random Forests and Voting Classifiers. The green group is the Western Dialect and the red group is the Eastern Dialect.

Xueming Zheng
Masters in Computational Science
Advisor: *Ming Ye*

A Computationally Efficient Method for Multi-Model Process Sensitivity Analysis

Ecological and environmental consist of multiple processes, and each process may have multiple conceptualizations for interpreting and representing the processes. This leads to conceptual model uncertainty. To address the problem of conceptual model uncertainty, Ye and the ORNL collaborator, developed a new method of multi-model process sensitivity analysis, which evaluate process sensitivity index for identifying important system processes under conceptual model uncertainty. This method integrates the concept of variance-based method with a hierarchical uncertainty quantification framework. However, a naïve implementation of the multi-model process sensitivity analysis is computationally expensive, and its computational cost increases dramatically with the number of system processes. To reduce the computational cost, a computationally efficient method has been developed. It rearranges the expression of the process sensitivity index, so that the computationally efficient quasi Monte Carlo method can be used for evaluating the process sensitivity index. The computational cost is reduced from N2 to 2N, where N (being thousands) is the number of Mont Carlo realizations. The accuracy and efficiency of the method has been evaluated using a numerical example of groundwater reactive transport modeling. The modeling results show that the number of model executions is reduced from millions to thousands. The computationally efficient method is also mathematically general and can be applicable to a wide range of ecological and environmental problems.

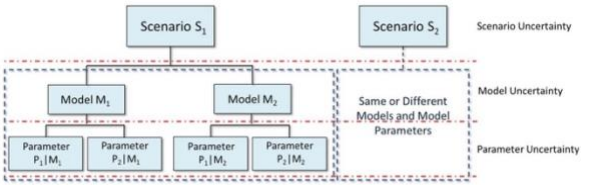


Figure 1. The three-layer hierarchical uncertainty quantification framework. It contains three layers of uncertainty factors: scenario, model, and parameter. Scenario uncertainty is aleatory and originates from the natural driving force of the model system. The scenarios may influence the model formulation and selection as each model may only be plausible for a limited range of conditions. Model uncertainty is caused by different conceptual or mathematical formulations of the targeted process, given that characterization data and process understanding is typically limited for any given complex system. Each model may have its distinct set of inputs and parameters that are also subject to uncertainty. This hierarchical framework clearly illustrates the concept of parameter uncertainty depending on both model and scenario uncertainty.

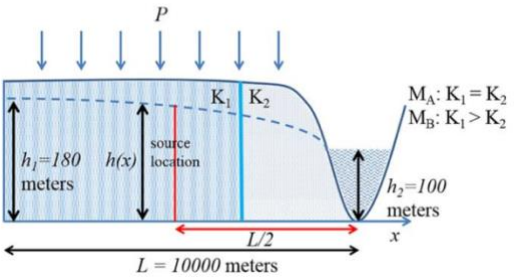


Figure 2. Diagram of the modeling domain of the synthetic study. L = 510,000 m is the domain length. Precipitation is uniform for the entire domain. Constant head (h) boundary conditions are h_1=5180 m and h_2=5100 m. The continuous contaminant source (marked in red) is located in the middle of the domain. The divide of the two zones of hydraulic conductivity at x = 57000 m is marked in blue.