SCIENTIFIC COMPUTING

Computational Xposition

2017

Welcome to Computational Xposition 2017,

a yearly event where students in the Department of Scientific Computing (DSC) showcase the results of their research in the past year. This research covers a broad spectrum of disciplines, but shares a common thread: they 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, looking at what our current students have achieved over the past several years serves as evidence 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

Front Cover: Surface scan of a small primate skull, courtesy of Benjamin Pomidor.
Back Cover: 1000 molecules of adamantine, courtesy of Benjamin Crysup.

Presenting Researchers

Research

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

Annealed pruning of neural network connections: A new regularization technique

A central goal when training a neural network is performance generalization: we want the trained neural network's accuracy on general data to be on par with the trained neural network's accuracy on the data used for network training. To facilitate generalization, neural networks are often trained with tools such as Dropout and L2-weight-regularization. We test a new approach to generalization, called Annealed Pruning, which removes connections in the neural network as the network trains. We implemented the Annealed Pruning with the Keras framework and benchmarked the performance against other regularization approaches using the MNIST and CIFAR-10 datasets. Although these other approaches (Dropout and L2-weight-regularization) demonstrate superior test-data accuracy, Annealed Pruning has the advantage of reducing the number of connections in some layers by up to 95% before performance degradation sets in. This connection reduction creates possibilities for reduced memory and computational burdens while training.
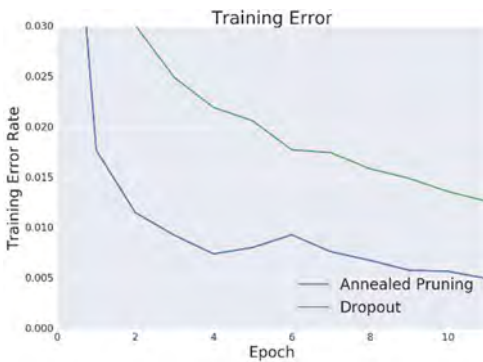

Figure 1. The training-data accuracy of two neural networks, each of which is trained with a unique regularizer identified by the legend.
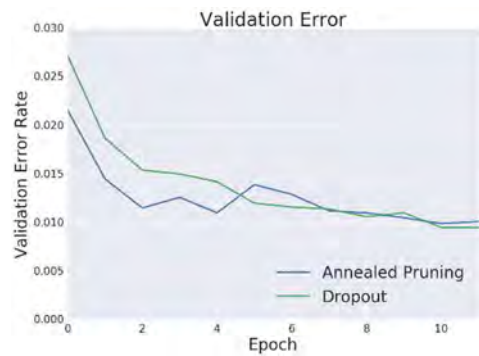

Figure 2: The validation/test-data accuracy of the trained neural networks. This graphic illustrates the ability of the networks to generalize their performance to unseen data.

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

Computational framework for evolution and nucleosynthesis studies of astrophysical objects

The chemical composition in the evolution of stellar plasma is one of the most critical components of computational models in astrophysics. We aim to construct the framework to support such computations in the context of multi-dimensional stellar evolution studies. A number of challenges arise from the fact that two different types of discrete representations of data are used in these computations. Our hydro simulations are performed on an adaptive Eulerian mesh, where we introduce passively advected Lagrangian tracer particles to continuously interpolate the thermodynamic quantities of the plasma for use in post-processing their initial isotopic abundances with a realistic set of isotopes. Thermodynamic trajectory information is output on a per-particle basis, but only once the fractional change of the density or temperature reaches a user-defined threshold, resulting in highly irregular data output. This data is buffered between checkpoint times to reduce computational cost of tracer particle output and to synchronize thermodynamic trajectory results with restart times. Following the post-processing stage,
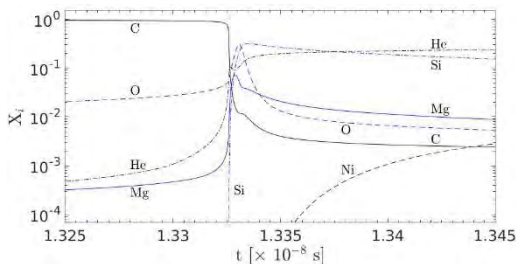
we map the more complete isotopic abundance distributions from the particles to the hydro domain using a variant of the linear Shepard's method.
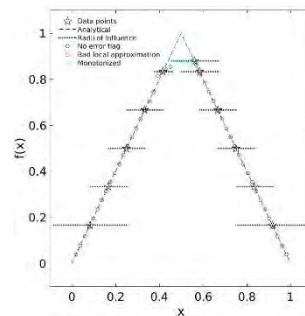
The linear Shepard's method is designed for the interpolation of only one quantity. However, we have multiple isotopic abundances and the additional requirement that these quantities must be normalized at every point in the domain. To handle this, we utilize a constrained interpolation technique involving slope-limiting and monotonization of the linear interpolant. Here we discuss the challenges involved in implementing this modified interpolation technique. Once completed, we will test this technique using the cellular detonation problem, a setup that includes multiple isotopes with highly varying abundances. Future applications of this newly-developed framework will be within explosive phenomena in astrophysics including core-collapse and thermonuclear supernovae.

Figure 1, Left: Nuclear evolution of a fluid element over time, where all nuclear species sum to 1.

Figure 2, Above: Demonstration of interpolation around an extrema, where monotonization is sometimes required.

Seth Boren
Ph.D. in Computational Science
Advisor: *Dennis Slice*

Morphometric analysis of the chimpanzee maxillary and ethmoid sinuses

The paranasal sinuses are a relatively unexamined region of the primate skull. As such, a morphometric study of the sinuses can potentially yield new information on morphological variation between individuals of the same species. The following analysis tests to see if significant differences exist between chimpanzees in the morphology of the maxillary and ethmoid sinuses. All of the sampled chimpanzee skulls were mapped for thirty landmarks present in the maxillary and ethmoid sinuses and subjected to a principal component analysis (PCA). The first three components of the PCA were then subjected to an analysis of variance (ANOVA) to see if a significant amount of variance existed between chimpanzees. No significant differences were found between chimpanzees grouped by sex and locality, but significant differences were found between chimpanzees when grouped by collection period. A greater sample of chimpanzee skulls with intact paranasal sinuses will be needed to further investigate how the observed significant differences came about. An analysis with a larger sample of species may lead to a greater understanding of variation in the overall paranasal sinus amongst great apes.
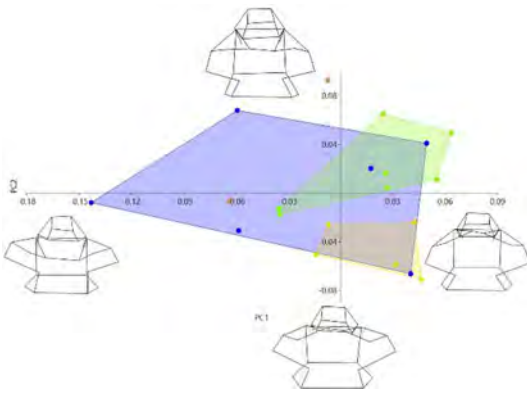

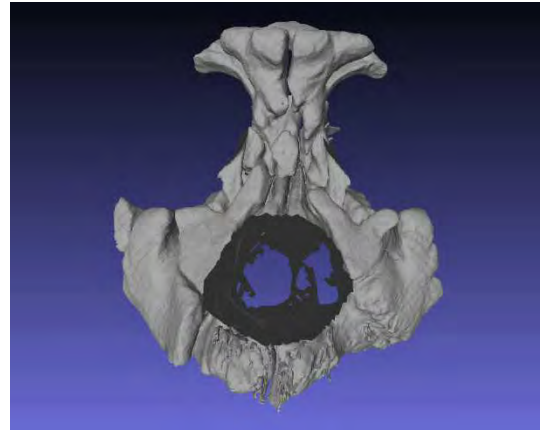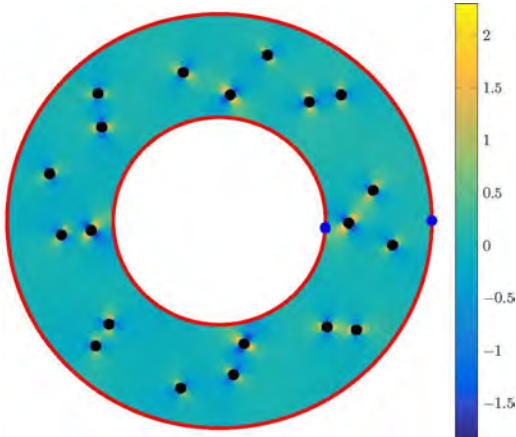Figure 1: Nuclear evolution of a fluid element over time

Figure 2, Above: Spatial variation of the average atomic mass, $\bar{A}$, for the final state of an example hydrodynamics simulation

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

Using boundary integral equations to model suspensions of rigid particles

Suspensions of particles are everywhere. Everything from mayonnaise to blood to molten plastics is a suspension. Understanding the macroscopic properties of suspensions can be aided by having a detailed understanding of the movements of the individual particles. Using boundary integral equations (BIEs) we are able to model this efficiently. Arising from classical potential theory, BIEs are a way to transform a problem over a domain into an integral around its boundary. In this way we are able to turn a 2D problem into a 1D line integral. This lets us solve the problem using far fewer unknowns. It also avoids meshing the interior of the domain, thus eliminating a potentially expensive step. Combined with fast summation techniques like the fast multipole method and iterative solvers like GMRES, BIEs have proven to be very efficient at solving linear elliptic PDEs. When applied to particle suspensions difficulties arise when needing to integrate nearly singular integrands, however we have developed near singular integration techniques to overcome this difficulty. We can use BIEs to calculate quantities of interest, including pressure, energy dissipation and stress-strain curves.
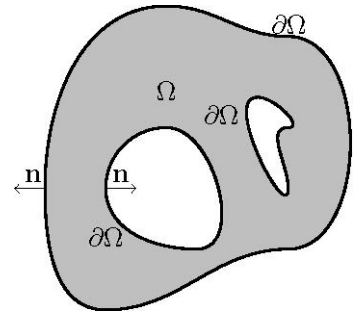

Figure 1, Above: Pressure field inside a Couette apparatus with suspended particles

Figure 2. Sample domain for bounded Stokes flow.

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

Optimal boundary condition approximation for Lagrange finite elements by polynomial extension

Curved domains pose a great problem to high order finite element methods because the accuracy of a numerical solution ultimately depends on how well the discrete domain approximates the continuous domain. As such, the majority of the higher order finite element codes utilize some sort of coordinate transformation to fit a polygonal mesh to the curved domain up to an acceptable order of accuracy. While these methods can recover optimal theoretical accuracy for the underlying approximation space, they are often difficult to implement. In our work, we present a new approach to recovering optimal accuracy for higher order finite elements with Lagrange elements for Dirichlet and Neumann problems for scalar elliptic PDE on curved domains. Instead of cumbersome coordinate transforms, we utilize a simple Taylor series to extend the underlying polynomial approximation space to approximate the given boundary condition on the curved boundary. This approximation allows us to retain our polygonal mesh, while at the same time, achieve optimal accuracy.
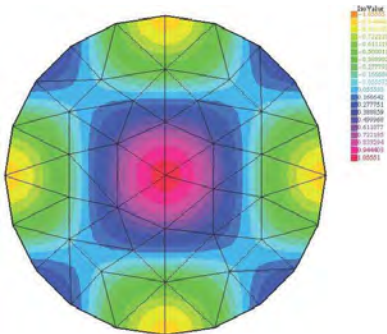

Figure: A manufactured solution computed on a circular mesh.

Michael Conry
Ph.D. in Computational Science
Advisor: *Alan Lemmon*

Simulating the effect of genetic recombination on gene-tree species-tree discordance in phylogenetic analysis

Phylogenetic studies typically attempt to take into account as many evolutionary factors as possible to maintain accuracy. In practice, however, genetic recombination is often ignored entirely leading to questions regarding the accuracy of the resulting trees. Since genetic recombination can occur in any part of the genome, closely related individuals which should have the same underlying genetic history may appear more distantly related. Herein, we describe a novel computational method to help understand the impact of genetic recombination on phylogenetic analyses by simulating populations undergoing recombination and reproduction, then tracking the exons that are exchanged between chromosomes and individuals. The process is repeated and resultant offspring's inferred phylogenies are compared to determine the rate of recombination that creates a conflicting or supported tree. The program is written in Java and in addition to the simulation has a visualization component to display the pedigrees created.
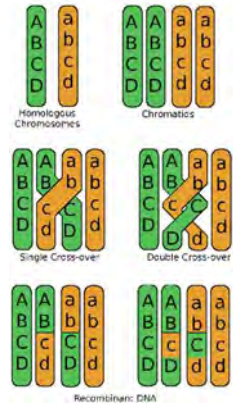

Figure 2: The above figure demonstrates crossing-over in gametes during meiosis. The homologous chromosomes split and replicate. Then, when a recombination event occurs, the chromatids cross-over and a double strand break occurs. The break is repaired which moves DNA from one chromatid to the other. This serves to increase genetic diversity.
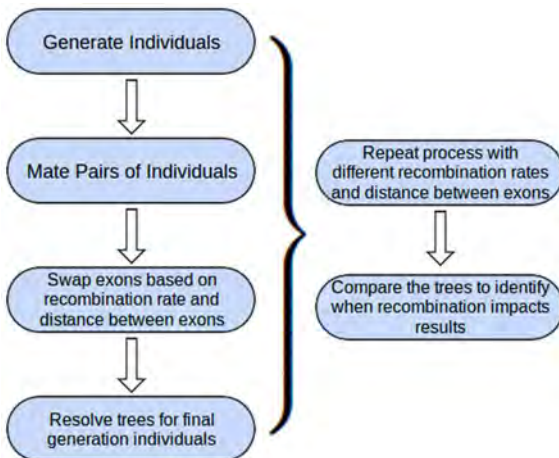
Figure 1, left: The flowchart here shows the proposed method for determining the effect different recombination rates and distances between exons have on phylogenetic trees. The process depicts simulating individuals, mating them while exchanging alleles through recombination, generating gene trees for the final generation's individuals, and repeating the process. Once this is completed for multiple runs we determine how closely the generated trees support the underlying evolutionary history.

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

Computational model for local calcium dynamics in astrocyte

Several contemporary studies show that astrocytes, a type of glial cell, are fundamental to several neural functions ranging from metabolic support to higher cognition such as recollection memory. This has resulted in the introduction of astrocytic dynamics into neural modeling. Most cellular function in the astrocyte is triggered by an increase or decrease in calcium concentration within the cytosol. Previous work considered astrocytic dynamics by representing calcium concentration as a point source or a completely spatial model in the cell. We now know, more than ever, 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, puts forward a novel approach to modeling the different levels of astrocytic calcium activity in the astro-neural system. In the model, we introduce a compartmentalized astrocyte with the purpose of understanding intra-cellular calcium dynamics. Compartmentalizing the astrocyte's
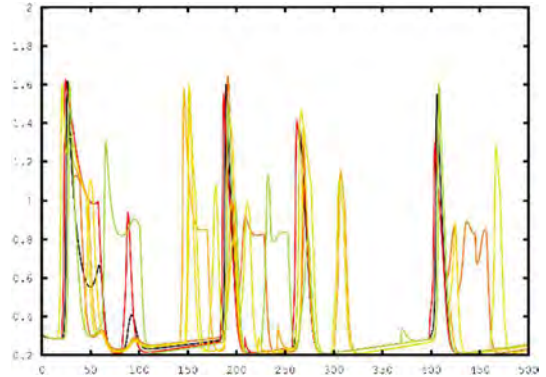
cytosol into the soma and individual branches captures the effect of the local dynamics while still holding on to the analytical power of ODE's. With this model we investigate the interaction between local and global cellular dynamics within the astrocyte in response to neural activity. This allows us to better understand the effect that astrocytes can have on both individual neurons and populations of neighboring neurons.
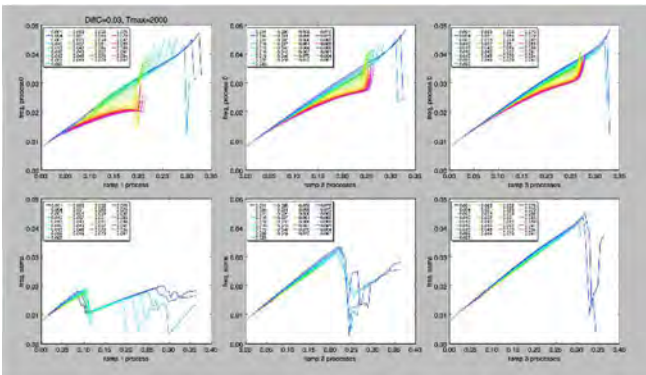
Above: Astrocytic calcium concentration in time

Left: Calcium oscillation analysis

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

New analysis techniques for new imaging techniques

The neuron doctrine has been a foothold for all neuroscientists as we ascend in our understanding of the brain. It revolutionized our knowledge of neuronal dynamics and our ability to study and diagnose diseases and disorders. Several contemporary studies show that astrocytes are more fundamental to many neural functions than previously believed. An "astrocyte doctrine," as one might call it, is sorely missing. With the advent of two-photon microscopy we are able to peer deeper into the complex inner workings of astrocytes. Studying this rich data may bring us closer to understanding their governing processes, but to do so requires the utilization of more sophisticated analysis techniques. Here we present a new method aimed at shining light on the rules governing intracellular calcium dynamics of astrocytes, which is believed to be their mechanism for communication. We employ state of the art denoising techniques, compute the flow of calcium throughout the cell using Farneback's optical flow algorithm, and measure the flux of calcium in and out of different compartments of the cell.
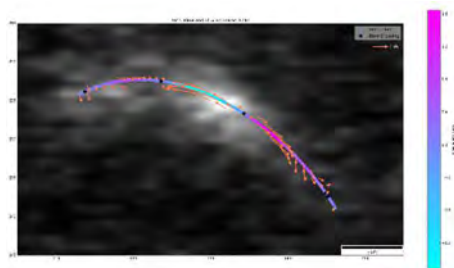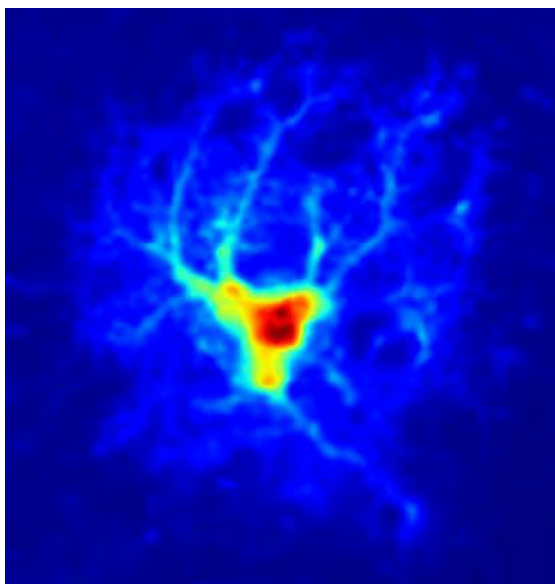
Figure 1, Left: A pixel-wise average over the entire time series allows us to see the general morphology of the cell being recorded. This is a typical astrocyte in the ferret visual cortex

Figure 2, Above: Here we see the flow of calcium along the path and the location of the hotspots detected. The vectors along the path show the flow of calcium and the color along the path is the flux at each point. Notice how the black dots are between flows of opposite directions.

Benjamin Crysup
Ph.D. in Computational Science
Advisor: *Sachin Shanbhag*

Uncertainty quantification of dynamical properties in interacting systems

Nanoparticles with a hard, inorganic core covered with long strands of polymer "hair" have potential applications as chromatography media, and coatings that produce the same structures help in the production of nanocomposites. Additionally, several other classes of materials, such as pluronic gels, take on a similar form. However, their synthesis can be challenging, so simulations would be helpful if it weren't for the fact that each individual nanoparticle can have more than 100 thousand atoms. However, if it were possible to come up with a single force that adequately represented the nanoparticle as a whole, large scale simulations would be cheaper and much more useful. There are many methods that can be used to generate a force model for the

nanoparticle given data from a (presumably smaller and shorter) atomistic simulation. These coarse graining methods include Inverse Boltzmann (an inverse modeling approach), Percus-Yevick and Hyper Netted Chain (two approaches based on the Ornstein-Zernike equation), and a direct force calculation. The aim of this work is to compare how well these coarse-graining methods preserve various quantities when applied to fuzzy nanoparticles. This is accomplished by running atomistic simulations of fuzzy nanoparticles of different core sizes and with different lengths of hair. Each method was applied, and the deviations in macroscopic quantities were noted.
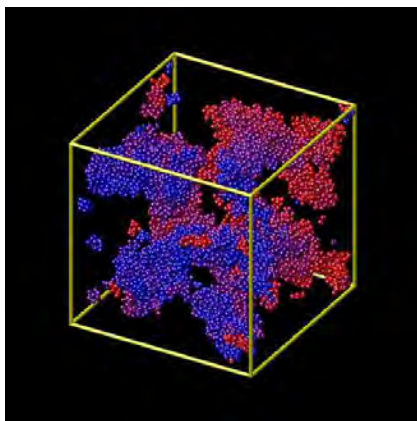

Figure 1, left: 1000 molecules of adamantane, during the burn in period of the simulation.

Figure 2, above: The error in the relative density of coarse grained particles of neopentane using various coarse-graining methods.

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

Automated parameter fitting in stellar evolution

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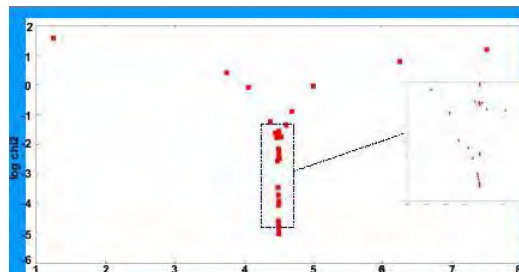
expert knowledge. Typically probing parameter space of this problem is done using a trial-and-error approach, which is inefficient, incomplete and prone to bias. Our aim is to identify the key model parameters, their most probable values, and their observationally constrained ranges in an automated way.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

Figure 1: Left

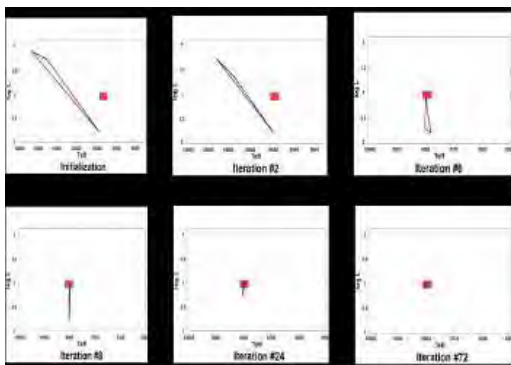The Hertzsprung- Russell diagram of a star where - during an optimization run for its parameters- the Simplex Method constructs a simplex and reflects, expands and contracts it towards a minimum.

Figure 2: Right

The cost function value versus the mass of a star. We used 1-D stellar models produced by the MESA code in order to account for the evolutionary behavior and physics of these average mass stars.

Christine Geshel
Undergraduate in Computational Science
Advisor: *Gordon Erlebacher*

A parameter study of a feed forward convolutional neural network

A convolutional neural network (CNN) was implemented to explore the impact of its various parameters on classification accuracy using training and test data. To this end, we choose the MNIST database, a collection of 60,000 28x28 images of handwritten numbers. The training data is formed from 50,000 of these images, while the testing data consists of the remaining 10,000 images.

A classical CNN consists of several pairs of convolutional and pooling layers, followed by one or more fully connected layers. The classes are determined by the last connected layer, also called the readout layer.

The structure of the CNN used in this study is as follows: two paired convolutional and pooling layers, and two fully connected layers. The first convolutional layer has 32 5x5 filters that reduce the image to a 14x14. The second convolutional layer has 64 5x5 filters that further reduce the image to a 7x7. Each of the fully connected layers have 1024 nodes. The readout layer sorts the image into one of ten possible classes.

We performed four parameter studies to increase our understanding of CNNs in a simple case. Study I varies the number of nodes in the two fully connected layers; study II varies the size the convolution filters, study III modifies the parameters of the pooling layers, and finally, test IV explores different optimization algorithms. The first three tests are performed with respect to the classification accuracy. The fourth tests compare convergence rates of the learning algorithm. Classification accuracy is not affected.

In the poster, we present the results of the four parameter studies and draw some conclusions.

Figure 2: A selection of 2s contained in the MNIST database.
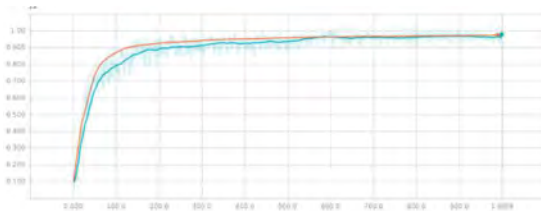

Figure 1: TensorBoard output for test and training accuracy for a 3x3 patch in parameter study II. Test accuracy is plotted in orange and training accuracy is plotted in blue. The x axis represents number of iterations, and the y axis represents the level of accuracy. .

.

Mario Harper

Ph.D. in Computational Science

Advisors: *Emmanuel Collins & Gordon Erlebacher*

Kinematic modeling of an RHex-type robot using a neural network

Motion planning for legged machines such as RHex-type robots is far less developed than motion planning for wheeled vehicles. One of the main reasons for this is the lack of kinematic and dynamic models for such platforms. Physics based models are difficult to develop for legged robots due to the difficulty of modeling the robot-terrain interaction and their overall complexity. The research presents a data driven approach in developing a kinematic model for the X-RHex Lite (XRL) platform. The methodology utilizes a feed-forward neural network to relate gait parameters to vehicle velocities.

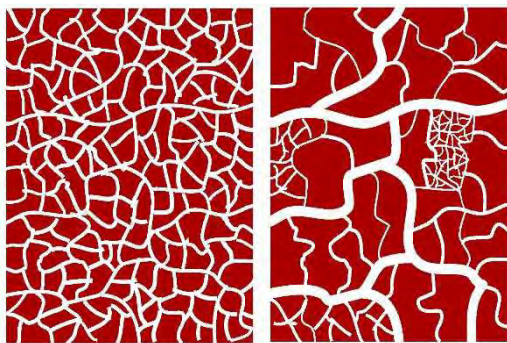Figure 1. (Left) Quadrupedal platform "Big Dog" built by Boston Dynamics

Figure 2. Hexapedal platform X-Rhex lite (XRL) built by the University of Pennsylvania

Eitan Lees

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

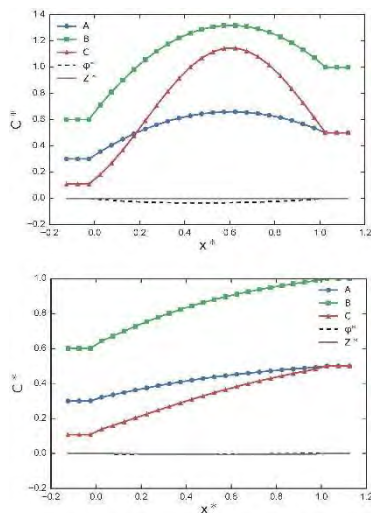## The electroneutrality constraint in nonlocal models

Heterogeneous porous media occur naturally (e.g. aquifers and oil reservoirs), and artificially (e.g. synthetic foams and membranes). Due to their heterogeneity, they exhibit variations in permeability starting from grain or pore size at the sub-micron to millimeter length scale. We develop a simple 1D nonlocal Nernst-Planck model for reaction and diffusion of multicomponent ionic solutions. We apply it to the liquid junction problem, in which two electrolytic solutions of different ionic concentrations are brought into contact via a permeable membrane. Transport of ions through the membrane induces an electric field, which is modeled using two separate nonlocal conditions: charge conservation and Gauss law.

We investigate how well they satisfy the criterion of strict electroneutrality, which stipulates that the net charge at each point in the domain is zero, by considering four different initial scenarios. Charge conservation and Gauss law yield similar results for most practical scenarios in which the initial condition satisfies strict electroneutrality. Gauss law, however, has two important advantages over charge conservation: (i) it is numerically more stable and can be applied even when the concentration of all the charged species drops to zero, and (ii) computationally, it is significantly cheaper. We discuss implications for multiphysics problems like corrosion-induced mechanical failure, where charge transport is only one of many coupled processes.

Figure 1: How the presence of cracks or long channels in the porous media lead to anomalous diffusion.

Figure 2: How charged chemical species evolved for one of our test cases.

Hongzhuan Lei
Ph.D. in Computational Science
Advisor: *Ming Ye*

Estimation of nitrogen load from septic systems to surface water bodies using ArcNLET

Excessive nitrogen loading to surface waters has caused a lot of environmental problems such as eutrophication or excess growth of algae. The ArcGIS-based Nitrogen Load Estimation Toolkit (ArcNLET) is used in this project to estimate the inorganic nitrogen load from septic systems to surface waterbodies in Sebastian, Indian River County, FL. The load estimation can be used to support sustainable decision-making and management of nitrogen pollution. Incorporating heterogeneous hydraulic parameters and considering spatial variation of locations and loads from individual septic systems, ArcNLET simulates the coupled transport of ammonium and nitrate in both the vadose zone and the saturated zone based on a simplified model of flow and nitrogen transport. The flow and transport model in ArcNLET are calibrated by using the field observations (i.e., water level and nitrogen concentration) from eight monitoring wells along the shore. Low ammonia concentration and high nitrate concentration from observations indicate that the nitrification process is largely complete under the vadose zone while the denitrification process is incomplete in the saturated zone. The final estimation of nitrogen load shows that (a) the St. Sebastian River and the Indian River Lagoon have the largest nitrogen loadings and (b) ammonium loadings account for only about 7.5% of the total nitrogen loadings which conforms to the indication of the largely complete nitrification process under the vadose zone in this area.
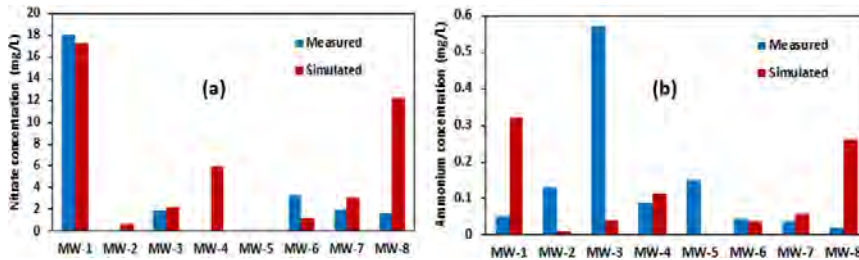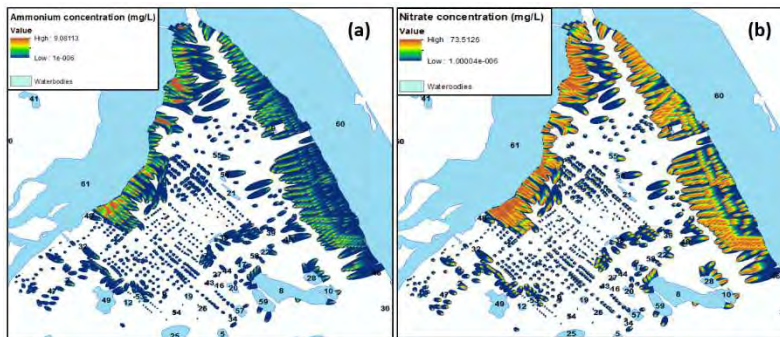

Figure 1. Measured and simulated nitrogen (a. nitrate and b. ammonium) concentration (mg/L) at eight monitoring wells.

Figure 2. Estimation: (a) plume of ammonium concentration (b) plume of nitrate concentration.

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

Using RBF-generated quadrature rules to solve nonlocal anomalous diffusion in multiple dimensions

Nonlocal models differ from traditional models by using an appropriate integral operator instead of a spatial differential operator. These nonlocal models are useful in various applications (e.g., biological systems, flow through porous media) where the observed rate of diffusion is not accurately modeled by standard diffusion thus exhibiting so-called anomalous diffusion. Nonlocal models are used to model anomalous diffusion because there are often issues calculating the spatial derivatives of a standard PDE model which contains a spatial differential operator. Solving anomalous diffusion problems is a very active research area and using a nonlocal model is a very promising approach.

In this work, we propose utilizing RBFs (Radial Basis Functions) in order to derive quadrature rules that can be used to solve anomalous nonlocal diffusion models. These quadrature rules are then used to derive a scheme for solving anomalous nonlocal diffusion problems. In particular, the method obtained from these quadrature rules extends well to higher dimensional problems. Numerical experiments are provided to show the accuracy in which this approach solves the nonlocal anomalous diffusion model and results are compared to standard finite element approximations.
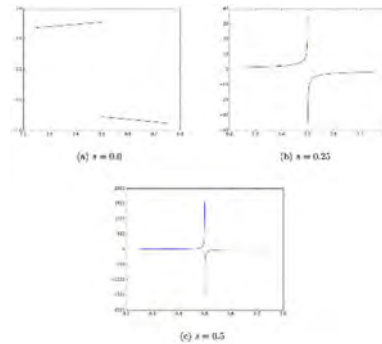

Figure 1: Discontinuous integrand corresponding to the nonlocal integral for various anomalous diffusion rates.
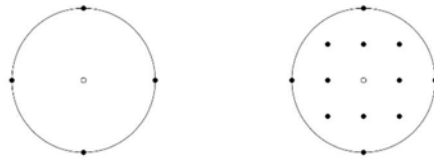

Figure 2: Quadrature points considered for the RBF-generated quadrature rule.

Alisha Mechtley
Ph.D. in Computational Science
Advisor: *Alan Lemmon*

SeedSQrrL: A utility for off-target DNA mining in non-model organisms

High throughput sequencing data are rich in information and contain many off-target sequences (reads) that are often ignored but may be biologically relevant. Seed extension, a combination of reference and de novo based assembly methods, can be used to extract the information but it is time-consuming to implement because it requires that multiple seeds (sequences from one or many

closely related species) be gathered in advance. A new tool is presented here, SeedSQrrL, that can automatically crawl the web to gather the seeds from the closest taxonomic relative for each gene and store it into a relational database. The seeds can then be used to create multiple seed extensions which are later combined into a reference and used for downstream phylogenetic analysis. The seeds can also be used to isolate segments of mitochondrial and chloroplast DNA used in species identification (Barcodes).

Figure 1: Left

The effect of enrichment efficiency and genome size on expected chloroplast coverage. The calculation assumes a chloroplast genome size of 150kb, a chloroplast copy number of 1500, and a target size of 100kb.

Figure 2: Right

A ratio of extended seed (contig) length to reference length was calculated for 6 mitochondrial genes in a sample of 18 squamates. The horizontal line indicates which seeds were extended (at or above the line), which had only partial coverage (0 < ratio < 1), and which were unable to be reconstructed (ratio = 0).

Serena Pham
M.S in Computational Science
Advisor: *Ming Ye*

Implementing a humic-ion binding model into Phreeqc

Software packages that model geochemical speciation and complexation are useful for predicting how different materials such as heavy metals and organic matter interact with the environment. However, these programs vary in what binding model is implemented to represent organic matter and its chemical behaviors. The Humic-Ion Binding Model VII is a discrete site model implemented by default in the WHAM7 software that represents binding behavior between protons, metal cations, and humic substances. We are interested in incorporating Model VII into the PHREEQC speciation program which currently does not contain a built-in model that can predict the environmental interactions of organic matter. We focus on developing this model for PHREEQC with parameters specific to mercury (Hg). These results may point to natural organic matter models and the software in which it is incorporated as another source of uncertainty in geochemical speciation modelling.
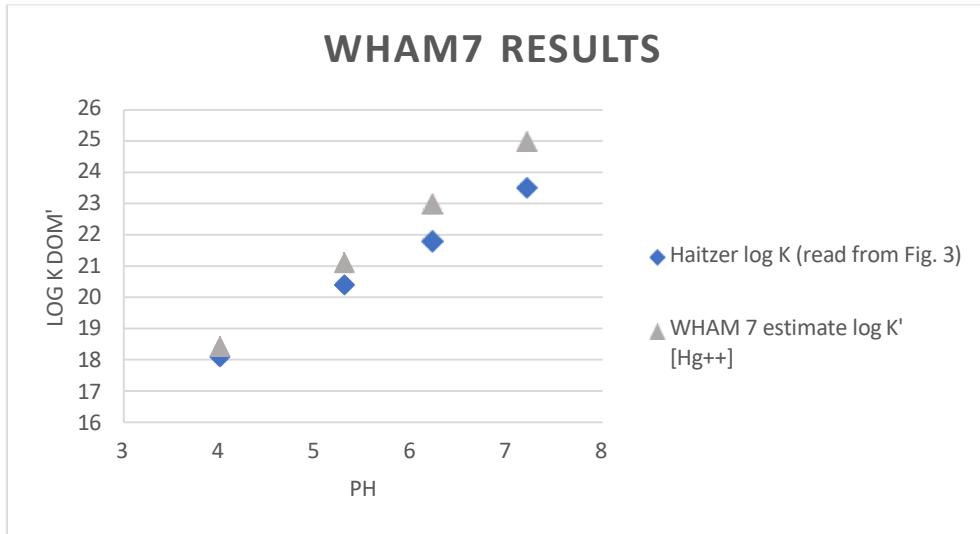
Reaction	PHREEQC log k	WHAM7 log k
$Hg^{2+} + Cl^- = HgCl^+$	7.31	7.21
$Hg^{2+} + 2\,Cl^- = HgCl_2$	14.00	13.98
$Hg^{2+} + 3\,Cl^- = HgCl_3^-$	14.925	15.06
$Hg^{2+} + 4\,Cl^- = HgCl_4^{2-}$	15.535	15.42

Top Figure: Binding constraints log K DOM' determined from experiments along with the WHAM7 estimate of log K DOM'. K DOM' = [HgDOM]/[Hg]x(DOM) where square brackets denote molar concentrations and parentheses denote mass concentration.

Bottom Figure: A sample of the database differences between PHREEQC and WHAM7.

Benjamin Pomidor
Ph.D. in Computational Science
Advisor: *Dennis Slice*

Real-time visualization and correspondence methods for geometric morphometric analysis of non-manifold surface meshes

An important aspect of current landmark-based geometric morphometric analysis is the identification of landmarks by researchers to ensure an accurate, consistent shape mapping from object to object in a sample. When moving from landmarks to surfaces, we rely upon algorithms and rules to establish this correspondence automatically. To provide insight and facilitate the further development of these automatic methods, especially as they relate to geometric morphometrics, we have constructed our own visualization package that works alongside our surface-based geometric morphometric library.

The rendering solution is written using the LWJGL OpenGL bindings for Java. It makes use of the programmable shader pipeline and instanced rendering while also taking advantage of Java's native threading to draw the large number of points and faces making up surface meshes in real-time without significantly disrupting computation, somewhat similar to particle rendering systems.

Also described here are several of the methods related to establishing surface correspondence, as this is the key reason for visualizing the actual superimposition process. The final superimposition of two surfaces, as the result of some pairwise superimposition technique like Iterative Closest Point, is simple enough to assess with a static post-computation visualization or an error metric, but determining the cause of a poor superimposition is a non-trivial process. Point and face attributes, local shape descriptors, and space partitioning work in concert to quickly determine the best point of correspondence on one surface for a given point on another surface. Understanding this system and how it compares to human-identified landmark correspondences is a significant factor in improving the robustness and accuracy of the geometric morphometric surface analysis method as a whole.
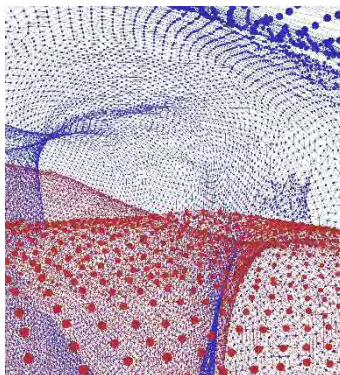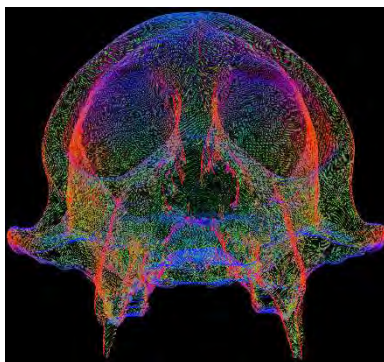

Figure 1, far left: Surface scan of a small primate skull with the normal vector mapped to RGB colors at each point.

Figure 2, left: A close-up view of two surface scans (one red, one blue) partially superimposed with both the triangulated mesh (lines) and individual vertices (circles) drawn.

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

An ensemble-proper orthogonal decomposition method for the nonstationary Navier-Stokes equations
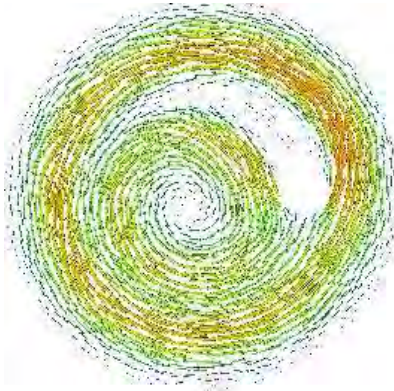
Figure 1: The velocity field of our ensemble-POD model.

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 different 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. The method uses the average of the multiple solutions at any time step to define a linear set of equations that determines the solutions at the next time step. To significantly 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 previously proven for the ensemble-based method are extended to the ensemble-POD approach. Numerical experiments are provided that illustrate the accuracy and efficiency of computations determined using the new approach.

Figure 2: 16000 degrees of freedom mesh which we use in the calculations of the ensemble POD algorithm.

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

CO and CO2 hydrogenation to methanol on stepped copper surface by random phase approximation (RPA)

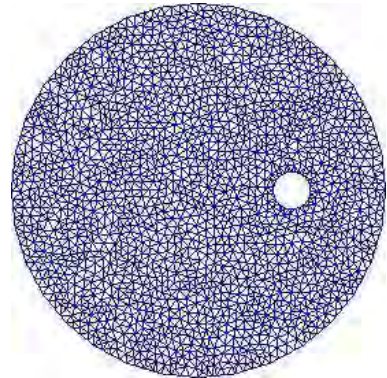Industrial methanol synthesis is about the conversion of the synthesis gas $CO/CO_2/H_2$ to methanol. The catalyst mainly consists of copper and zinc oxide. Despite extensive research on its mechanism, it is still highly debated that is what the active sites are. Resolving the active sites is the requisite for designing better catalysts. With isotope labeling techniques, experiments showed that CO_2 is the major carbon source for producing methanol, while previous density functional theory (DFT) studies suggested that the major carbon source is CO. If the DFT results were correct, we may rule out the possibility of copper as the active site and should investigate other candidate active sites, such as the sites at the interface between copper and zinc oxide. However, it is known that DFT employing density-based exchange-correlation functionals significantly overestimates the adsorption energies of molecules on transition metals, which is due to the self-interaction error (SIE). In addition, SIE underestimates the reaction barriers. In this work, we investigate the reaction energies of the CO and CO_2 hydrogenations over copper by performing the random phase approximation (RPA) calculations. We first establish the accuracy of RPA by computing the adsorption energies of selected molecules on platinum surface. RPA predictions agree well with experiments: the error is within 0.3eV. Gibbs free energy diagrams of CO and CO_2 hydrogenations over copper (211) surface are constructed using RPA (Fig. 1 and 2). The highest barriers in both pathways are determined by the last hydrogenation step and are very close, which suggests that the two pathways might have comparable rates. We are currently conducting micro-kinetic modeling to estimate the reaction rates, with which we will gain deep insight to the long-standing question of whether metallic copper can be the active site by itself.

Figure 1, left. Gibbs free energy for CO_2 hydrogenation on Cu(211) surface, computed using ONIOM(RPA:PBE) and PBE.

Figure 2, right. Gibbs free energy of CO hydrogenation on Cu(211) surface, computed using ONIOM(RPA:PBE) and PBE

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

eRebuild: increasing learning through targeted game levels

Making games has become a time consuming task, with AAA games taking thousands of people and two or three years. With this amount of man-hours spent, it is not surprising that the budgets of large games are also huge, sometimes over $100 million. With eRebuild, a game targeting math education in middle school aged children, we have neither the time or money to compete in these categories. This necessitates other solutions. In addition to the twenty levels our team has designed, we intend to address the lack of additional content on two fronts. The first is a level creation tool to allow teachers to create levels that address the subject matter of their current lessons. By reducing the scale of target audience to a class with the teacher guiding the experience, we believe the game will become more effective in its goal to increase mathematical ability. The second way we add content is a procedural content generation system that uses past gameplay as a guide to create suitable challenges for the player. By adapting to the individual player, we have again reduced the size of the target audience, and it is our hope that this will allow even greater efficiency in learning.
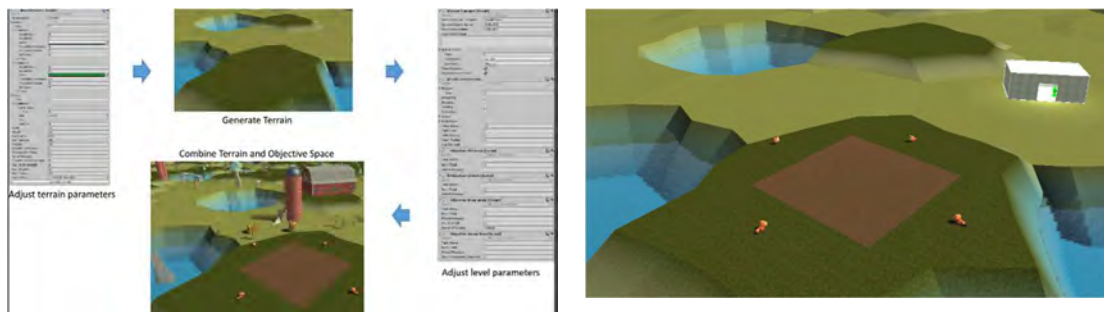

Figure 1, left. Controls and intermediate/final results of procedural level generation in eRebuild.
Figure 2, right. Combined generation of terrain and task in eRebuild.

Kenneth Sockwell
Ph.D. in Computational Science
Advisors: *Max Gunzburger & Janet Peterson*

Conservation properties and performance of exponential integrators for nonlinear conservation laws

Current climate models use explicit schemes in time over implicit schemes. This provides a reduction in computational cost but severely restricts the size of the time step that can be taken in a simulation. Recently, exponential integrators have been developed that provide explicit-type schemes, while relaxing time step restrictions. This is an attractive property for climate modeling simulations that must simulate to a time horizon of the order of millennia, while being restricted to time steps of the order of minutes to hours. This problem is further complicated by the fact that the time step size restriction is dependent on the spatial resolution as well. As the simulation grid is made finer, to resolve smaller effects, the allowable time step size becomes even smaller. Exponential integrators provide an explicit-type scheme while avoiding the time step size restriction. Although it seems clear that exponential integrators provide an advantage over explicit methods, the conservation properties of the exponential integrators are still not clear. In this work, we aim to show through analysis and computations that a family of exponential integrators possess conservation properties when applied to non-linear conservation laws and the shallow water equations.
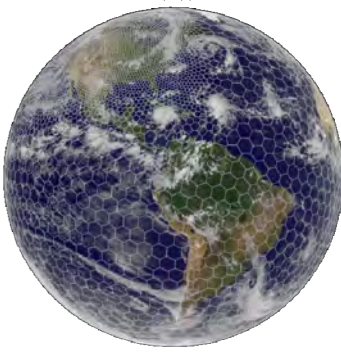

Figure 1: A highly nonuniform mesh of the Earth that resolves climatic phenomena at multiple scales. Although this mesh provides multi-resolution abilities in simulations, the smallest sized cells limit the time step of the simulation, requiring new time integration techniques to enhance simulation performance.
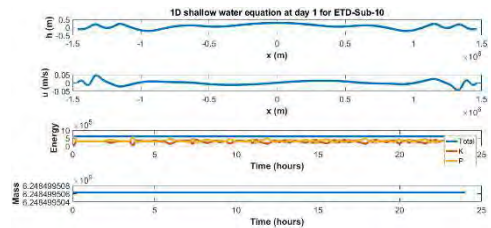

Figure 2: The simulated height h and the velocity u (top and center top respectively) of an ocean between two continental shelves after one day. These solutions were found using the shallow water equations and an exponential integrator combined with a staggered finite volume scheme. Both mass an energy are conserved using this method (bottom center and bottom respectively).

James Soda
Ph.D. in Computational Science
Advisor: *Dennis Slice*

Vector autoregressive representations of temporally-dynamic systems allow for measures of distance and statistical inference

Temporally-dynamic systems are ubiquitous in ecology and evolution. Yet, most common statistical methods represent data as static scalars or vectors. An alternative conceptualization is as stochastic processes represented as vector autoregressive (VAR) models, in which case individual samples are time series. Statistical inference is then possible. For example, a likelihood ratio test can assess whether two time series originated from processes with a single VAR representation, which we call a likelihood ratio test for a common process. So long as the system's constraints drive the processes' behaviors, this test

assesses whether two systems have identical constraints. Furthermore, a VAR representation for such systems also leads to a metric for the difference between systems. This metric is based on the process's marginal mean and on induced norms related to the model's coefficient matrices. Since it incorporates the Euclidean induced norm, we call this metric the Euclidean process distance. Here we evaluate the likelihood ratio test's effectiveness and how its results relate to the Euclidean process distance between series. To estimate Type I error rates, we simulated pairs of time series using 1,000 random VAR models under varying orders and numbers of subseries and ran the likelihood ratio test for a common process.
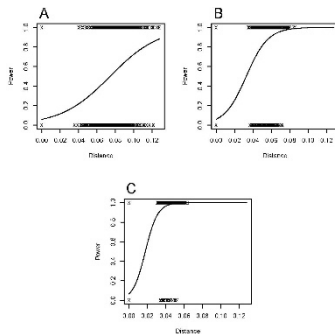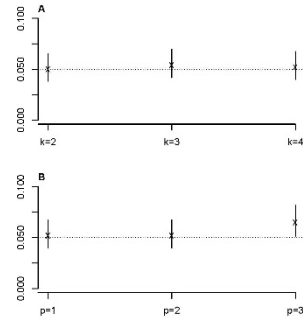
Furthermore, based on these pairs and 1,000 additional pairs of series from non-identical VAR models, we generated power curves under the same conditions using logistic regression and based on the Euclidean process distance. When α=0.05, the estimated Type I error rates were between 0.05 and 0.065, and all but one 95% confidence interval included 0.05. Estimated power curves suggest increased power with greater orders and subseries and reveal that the probability of a significant result is correlated with the process distance.

.

Figure 1, top right. Estimated Type I error rates for the likelihood ratio test for a common process under varying numbers of subseries and model orders. The X indicates a point estimate, and the bars represent a 95% confidence interval for the true rate. The dotted line indicates the desired rate of 0.05. A) Estimates for an order one model with varying numbers of subseries (k), B) Estimates under varying model orders (p) with four subseries

Figure 2, bottom left. Estimated power curves for VAR processes of various orders (p), each incorporating four subseries. Each X provides the result for one comparison with 0 indicating failure to reject the null hypothesis and 1 indicating rejection of the null hypothesis. From these points, the curves are fit via logistic regression. A) p=1, B) p=2, C) p=3

Jacob Spainhour
Undergraduate Research Opportunity Program
Mentors: *Lukas Bystricky & Michael Schneier*

Computational geometry in public policy

Every ten years, the current members of state legislatures undergo a process of redistricting, in which congressional districts are drawn to accurately reflect changing demographics. However, this process has become entangled with partisan politics, leading to divisions that intentionally leave entire populations unrepresented. While politicians are prone to manipulating the system in their favor, a computer will be objective in providing an accurate representation. This project involves the creation of software to perform the task of dividing the state of Florida into areas with regular shapes and equal populations. We use an algorithm based on the mathematical concept of centroidal Voronoi tessellations (CVTs) to accomplish this. We also show how CVTs can be applied to a seemingly unrelated problem: image compression.
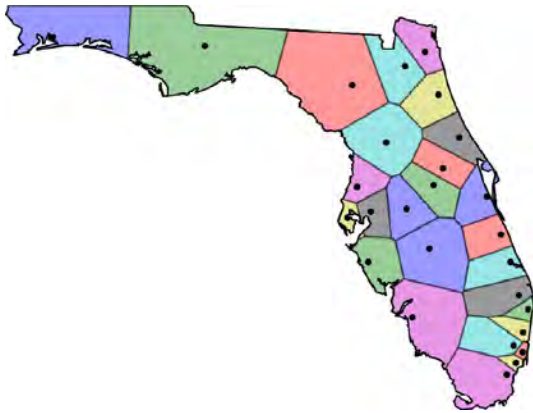

Figure 1: Map of polygonal congressional districts with equal population
.

Figure 2: CVT applied to image compression by color reduction

Detelina Stoyanova
Postdoctoral Associate
Advisor: *Dennis Slice*

Left or right pubic symphysis: Asymmetry analysis of age-at-death estimation using 3D laser scans and computational algorithms

Age-at-death estimation is crucial for building individual forensic profiles and studying mortality in past populations. For decades, anthropologists have relied on imprecise age estimation techniques based on the visual inspection of the pubic symphysis. Recently several computational methods using 3D laser symphyseal scans have been proposed as accurate, reliable and objective alternatives to current practices. The methods include two surface analysis algorithms, one ventral outline measure, and two multivariate-regression models combining each surface measure with the outline score. The five proposed models are calibrated on 3D scans from white males where the left or right pubic symphysis was randomly selected. A question remains whether the asymmetry of the two surfaces affects the age estimates. For this study both the left and right pubic symphyses from 25 white males are scanned. Both sides are used to estimate the age-at-death for each individual using the five computational models.

Additional tests are performed by selecting 25 individuals for whom the left or right symphyseal scan is included in the data for the models. For those males the age-at-death is estimated using the opposite side. The results of paired t-tests for mean differences, Wilcoxon rank sum tests for median differences and Kolmogorov-Smirnov tests for distributional differences show that there is no significant difference (p-values > 0.26) between the age estimates of the two sides for the 50 males. The Spearman and Pearson correlations are robustly positive, between 0.47 and 0.72 (p-value < 0.05), suggesting a monotonic relationship with high degree of linear dependence.
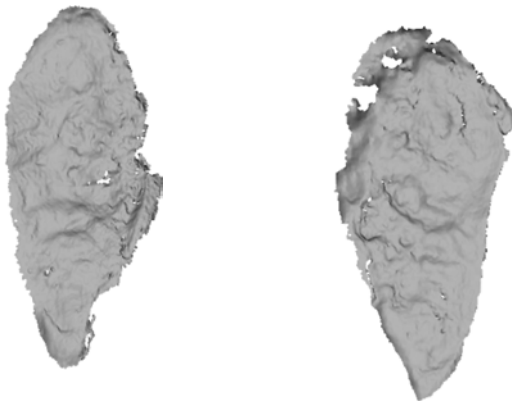
Figures at left: Example of the left and right symphyseal surfaces of the same individual.

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

An evolutionary approach for fMRI big data classification

Resting-state function magnetic resonance imaging (fMRI) images allow us to see the level of activity in a patient's brain. We consider fMRI of patients before and after they underwent a smoking cessation treatment. Two classes of patients have been studied here, one that took the drug N-acetylcysteine (NAC) and the ones that took a placebo. Our goal was to classify the relapse in nicotine-dependent patients as treatment or non-treatment based on their fMRI scans. The image slices of brain are used as the variable and as results here we deal with a big data problem with about 240,000 inputs. To handle this problem, the data had to be reduced and the first process in doing that was to create a mask to apply to all images. The mask was created by averaging the before images for all patients and selecting the top 40% of voxels from that average. This mask was then applied to all fMRI images for all patients. The average of the difference in the before treatment and after fMRI images for each patient were found and these were flattened to one dimension. Then a matrix was made by stacking these 1D arrays on top of each other and a data reduction algorithm was applied on it. Lastly, this matrix was fed into some machine learning and Genetic Programming algorithms and leave-one-out cross-validation was used to test the accuracy. Out of all the data reduction machine learning algorithms used, the best accuracy was obtained using Principal Component Analysis (PCA) along with Genetic Programming (GP) classifier. This gave an accuracy of 74%, which we consider significant enough to suggest that there is a difference in the resting-state fMRI images of a smoker that undergoes this smoking cessation treatment compared to a smoker that receives a placebo.
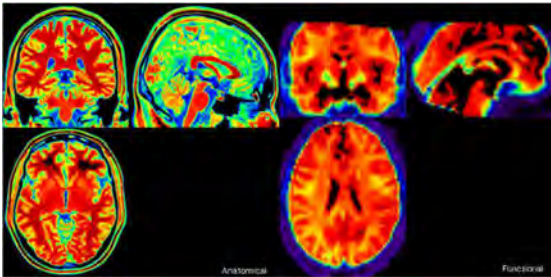

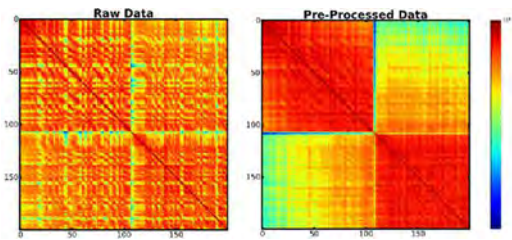Figure 1. Anatomical and functional image slices of brain subject.

Figure 2. Raw and pre-processed functional MRI data.

Stephen Townsend
M.S. in Computational Science
Advisor: *Dennis Slice*

A new method for statistical shape analysis of signals in the time-domain

Signals and Time Series appear in many applications across science and engineering, from linguistics to electronics and even into areas of business such as finance and are vital to the study and advancement of all of these subjects. One very important type of signal is sound. Sound is a very important object of study for many different fields including bioacoustics and linguistics. This project aims to produce a new method to analyze sound signals which have a pulsed and highly repetitive structure, such as speech or animal calls. Our method, unlike many common previous methods, does not rely explicitly on the representation of the signal in Fourier space, rather, we look at the signal as if it were simply a shape.
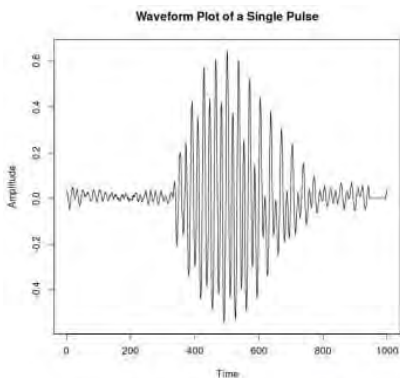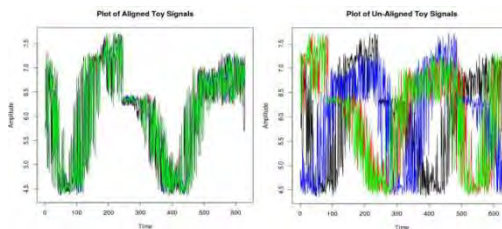

Figure 2: This figure shows the alignment algorithm with relatively complex simulated data. As the figure shows, the algorithm will align these signals accurately. Note that these signals are shifted copies of one another, however, they are noisy and therefore not perfect copies of one another.

Figure 1: This figure shows a single pulse of a frog call prior to processing.

Once digitized using a common sampling rate, each of the sample points can be seen as a type of semi-landmark. Using this representation, we then proceed to analyze the shape differences between a set of sample signals to determine relationship structures between them. This is achieved by aligning each of the signals and superimposing them so that the only remaining difference between them is waveform shape, not relative intensity or time frame. These aligned signals are then analyzed using a clustering method. Waveform shape is used because it can encapsulate many important properties of the signals. This can be applied in many contexts, ranging from understanding relationships between languages to looking at geographic variation in animal vocalizations. In this work, our method is applied to the study of geographic variations in a sample of frog calls obtained from the MacAulay Library as an example of such an application.

Bin Xu
Visiting Researcher in Computational Science
Advisor: *Ming Ye*

A multiple-reservoirs model for estimating karst aquifer parameters by studying karst spring hydrograph

Karst aquifers consist of two different but interacting systems: conduit and matrix. While the conduit system with a fast and turbulent flow distinguishes itself from the matrix system with a slow and laminar flow, the conduit and matrix systems have a strong hydraulic connection. It is always challenging to accurately characterize the karst systems with heterogeneous media. During the recession of karst systems, the recession limb of a spring hydrograph can reflect the two systems with different hydraulic parameters. To simulate the karst aquifer, we propose a multiple reservoirs model, the matrix system conceptualized as one reservoir and conduit system conceptualized as a set of serial small reservoirs. This study also introduces an approach to fit a storage-outflow model for each segment indicating a small reservoir and then average the parameter values. Matching strip method was used to analyze the matrix system. The above methods facilitate best global model selection and effective porosity estimation for the karst watershed.
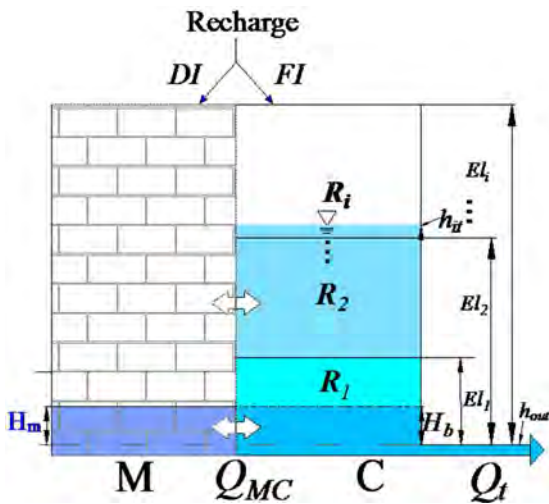

Figure 1: A multiple-reservoir model, in which the matrix system is conceptualized as one reservoir and conduit system conceptualized as a set of serial small reservoirs. *FI* fast infiltration, *DI*-delayed infiltration, Q_{MC} flow between matrix and conduit, R_i conduit reservoir i, El_i top elevation of reservoir i, Q_t outlet discharge, H_b Basic water table of conduit system, h_{out} water level of spring, *hit* the higher hydraulic head of *i*-th conduit reservoir than h_{out} at time *t*, H_m water level of matrix reservoir.
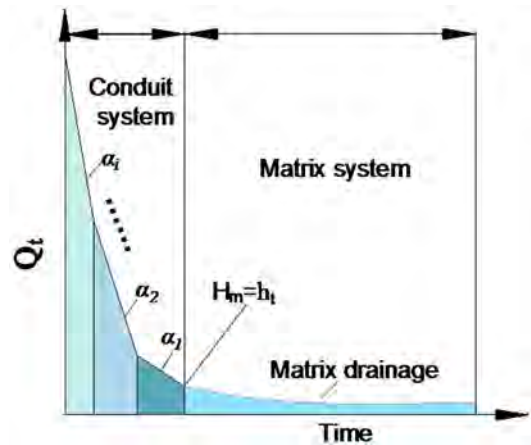
Figure 2: Different segments of a spring hydrograph recession after a storm pulse. When the water level at conduit reservoir h_t is greater than H_m, water level of matrix reservoir. Every breakpoint on the line is caused by a change in a characteristic of an underground karst reservoir.

Wenju Zhao
Ph.D. in Computational Science
Advisor: *Max Gunzburger*

Stochastic collocation method for state constrained stochastic optimal boundary control of the Navier-Stokes equations

We consider the optimal control of a system governed by the Navier-Stokes equations with a stochastic Dirichlet boundary condition. Control conditions applied only on the boundary are associated with reduced regularity, as compared to internal controls.

To simulate the system, state solutions will be approximated using the stochastic collocation finite element approach, and sparse grid techniques are applied to the boundary random field. The one shot optimality systems are derived from the Lagrange functional.

Error estimates are computed for the optimality using samples, and for the state equation using interpolated boundary conditions. Error estimates for the control equation comes via a non-conforming finite element variational crime.

A numerical simulation can then be made, using a combination of Monte Carlo and sparse grid methods, which demonstrates the efficiency of algorithms.
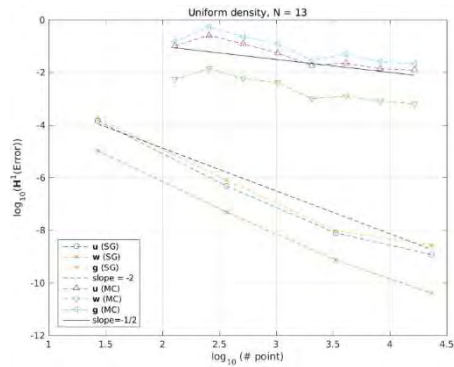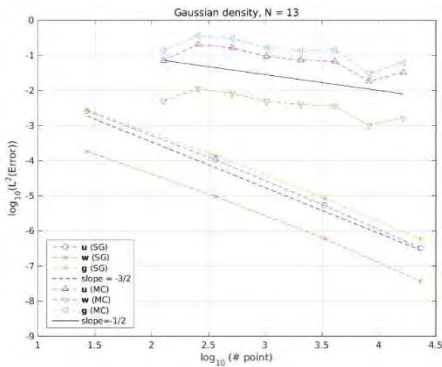
Figures: Convergence with Smolyak sparse grid methods and Monte Carlo methods.