Summer 2021

Department of Scientific Computing

CONTENTS

3

Beerli awarded funding

4

Research continues for emeritus prof

5

SC welcomes Administrative Specialist

6

Summer grad develops software

Arts & Sciences launches new grad major

SC Chair named Interim Director of Data Science

The College of Arts and Sciences recently announced a new degree program in Interdisciplinary Data Science. The new program is a collaborative, cross-departmental effort between Scientific Computing, Statistics, Mathematics, and Computer Science. Department of Scientific Computing Chair Gordon Erlebacher conceptualized, then spearheaded the effort to implement the major.

The data science major will feature a range of current and new courses – many in the student's chosen focus area – that span the four collaborating departments. The new program's emphasis will be a focus on machine learning and artificial intelligence, along with the necessary support tools, including issues such as data cleaning, feature construction, statistical analysis, database management, data privacy, regression, and a multitude of machine learning artificial intelligence techniques.

In early 2018, Erlebacher approached Sam Huckaba, Dean of the College of Arts and Sciences, with an idea: he wanted to explore the creation of a new undergraduate degree program that would explicitly address the science of data. "At that time, the other departments – statistics, computer science, and math – were already hiring in data science and were ahead of us in that regard," recounted Erlebacher. "The Medical School and the College of Communication had been discussing data science for at least the past three years. Data science had grown so much over the last decade that there are data science programs and Institutes throughout the country; FSU could not be left behind."

A year later, seeing the growth of data science on campus, the Dean called on the chairs of the four departments to discuss the idea of a Master's degree program to reflect the rapid growth of data science across campus and prioritize the need for graduate-level training in this area.

Shortly thereafter, at the Dean's direction, Erlebacher and the department chairs of computer science, mathematics, and statistics began to work towards creating a new master's degree program. The four operated as a team throughout the process, with Erlebacher serving as a catalyst, assuming the bulk of the writing and leadership duties.

From that initial meeting with the Dean, the process took months; many people participated and contributed to the project's ultimate success. Jennifer Buchanan, Associate VP for Faculty Development and Advancement, reviewed requirements, and the process of proposing a new degree program; Frank Johnson, Chair of Psychology, provided advice and counsel. The Dean and Associate Dean of the College of Communication, Larry Dennis and Paul Marty, also offered input that helped strengthen the proposal. Tim Logan provided advice on courses, reviewing existing classes, and designing discrete curricula for the data science major. Erlebacher compared program

continued from page 1

notes with Pierre Ngnepieba, Chair of Florida A&M University's Mathematics Department, and explored possible future collaborations with Richard Alo, Dean of Science and Technology at FAMU.

By mid-February 2020, the Board of Trustees had reviewed and approved the proposed program, and Erlebacher was moving forward with work toward implementation. He got back to work on the program in August, after a break necessitated by the pandemic.

"After the BOT approved the program in February, we immediately started work on the Proposal to Implement. However, in early March, COVID-19 reached Florida, priorities changed, courses went remote, and work on this proposal stopped as everyone involved in this effort worked to address the many challenges associated with responding to that crisis. Work on the Proposal to Implement resumed in earnest in early August 2020, with a renewal meeting (held remotely!) involving Dean Huckaba and the development group."

Now, two years later, the new Interdisciplinary Data Science degree is a reality. Student enrollment in the program begins immediately; the first cohort enrolls in Fall 2021.

Recently, Huckaba selected Erlebacher to lead the program as interim director, and Erlebacher has outlined what he would like to accomplish in the coming months. Among his priorities, he would like to establish a pattern of student growth over the next two years and strengthen the collaboration between the four departments that initiated the program.

"Currently, I am appointed Interim Director of the degree program, but that could change," Erlebacher said. "For the time being, I would like this program to establish a strong core curriculum, including Data Ethics, and deliver several courses currently in the development phase. I would like to establish internships with companies throughout Florida and a couple of countries in Latin America - Peru and Panama - where I have developed contacts over the past few years. Panama, in particular, awaits an influx of knowledge related to data science."

For more on the new Interdisciplinary Data Science degree program, go to https://datascience.fsu.edu.
For more on the Department of Scientific Computing, go to https://sc.fsu.edu.

Scientific Computing Chair and Interdisciplinary Data Science program Gordon Erlebacher

Beerli awarded NSF grant

Professor Peter Beerli and Assistant Professor Somayeh Mashayekhi

FSU professor and population geneticist Peter Beerli and mathematician Somayeh Mashayekhi, assistant professor at Kennesaw State University (GA), have been awarded a National Science Foundation grant that will help them study how to curb the effects of pathogens or improve the survival of endangered or commercially exploited species. Their research proposal focuses on and challenges the assumption that number of offspring in species remains constant across generations and is independent of the environment; the resulting improved population modelling will increase accuracy and decrease bias of population size estimates.

Natural populations live in heterogeneous environments, and individuals in these populations have different chances to produce offspring. In contrast, current inferences of population genomic data are coalescence-based and assume environmental homogeneity within a population. Additionally, the commonly used coalescent framework assumes a variance of relatively narrow numbers of offspring.

"These assumptions sometimes lead to interesting disconnects between theoretical expectations and observations. For example, the genetic variability of a population may suggest there are some number of individuals, but observational data show that there are many more individuals than the theory allows," Beerli said.

The grant will help Beerli and Mashayekhi study the effect of dissimilarities in offspring production of a population using a new theoretical framework that can handle heterogeneity. These new methods will be incorporated into Beerli's widely-used open-source computer software, MIGRATE. MIGRATE uses genomic data to infer population genetic parameters, such as population size, immigration rates, and population divergence times.

Beerli is enthusiastic about this work and about continuing the collaboration with Mashayekhi who was a postdoc in his group before she moved to Kennesaw State. "I hope this is a long and fruitful collaboration. We envision that our method - the fractional coalescent - may lead to many improvements of the standard population genetic theory."

Beerli and Mashayekhi's grant, Reproductive heterogeneity in the structured coalescent framework, was funded by the National Science Foundation through DBI, the Division of Biological Infrastructure. The division advances fundamental biological research to transform the future by investing in innovation and capacity-building cutting edge research infrastructure for fundamental biological science.

For more on Beerli and MIGRATE, go to peterbeerli.com.

Emeritus Prof continues active research program

I. Michael Navon at his desk.

Professor Emeritus Michael Navon has always had a very wide research circuit. With formal training in mathematics, physics and meteorology, Navon used much of his career to apply sophisticated data, statistical and mathematical models -- advanced 4-D variational data-assimilation methods, large-scale minimization, ensemble Kalman filter methods, as well as others -- to study oceans, climate, weather, and atmosphere. Scholars from all over the world continue to seek his mathematical expertise and research acumen, and despite having retired years ago, he continues as senior scholar and chief scientist for projects in the U.S. and Europe.

In recent years, Navon has become interested in Model Order Reduction (MOR), a technique for reducing the computational complexity of mathematical models in numerical simulations. It is closely related to the concept of metamodeling with applications in all areas of mathematical modelling.

"Many modern mathematical models of real-life processes are large and complex, so when they are used in numerical simulations, these factors pose challenges," Navon said. "Model Order Reduction aims to lower the computational complexity of such problems, for example, by using simulations of large-scale dynamical systems and control systems. By reducing a model's degrees of freedom, we compute an approximation to the equations of the original model. This is commonly referred to as a reduced order model.

"I started using MOR applications via POD [Proper Orthogonal Decomposition] and DMD [Dynamic Mode Decomposition – a dimensionality reduction algorithm that is data-driven, i.e., works on data measurement without requiring knowledge of the model equations] during the latter part of my research career as a means of researching these very complex real processes.

"Given a time series of data, DMD computes a set of modes each of which is associated with a fixed oscillation frequency and decay or growth rate. For linear systems in particular, these modes and frequencies are analogous to the normal modes of the system, but more generally, they are

New administrative staff at DSC

Monica Hartsfield is the department's new Administrative Specialist. In this role, Hartsfield will handle business, financial and budgetary process for the department, approve financial transactions, assist professors with pre- and post-award grant matters, and supervise several positions. She applied for the position because she wanted to work with and be a part of higher education, and she's glad for the opportunity to again work at FSU.

Originally from Fayetteville, NC, Hartsfield grew up next to an army base, Fort Bragg, where her dad began to work in the early 1980's. After graduating high school, she spent three years in Auburn, AL, following in her father's footsteps to Auburn University where in 2003, she completed the Bachelor of Arts degree in history.

Following undergrad, Hartsfield came to Tallahassee and Florida State to pursue graduate work. She finished her master of arts degree in 2005, and has held positions at Tallahassee Community College, FSU, Geek Speak, and the Apalachee Center where she was an administrative assistant and accountant. Since arriving at Scientific Computing in April, Hartsfield has taken steps to familiarize herself with the department and university systems by taking training and certification courses.

Hartsfield enjoys movies, and is looking forward to returning to movie theatres – she especially misses the aspect of sound and total immersion theatres provide. In the coming years, Hartsfield plans to introduce her three-year old son, Wyatt, to birds of prey through the Southeastern Raptor Center, a rehabilitation, rescue and release center within the Auburn University College of Veterinary Medicine. She is married to Alex, a sixth generation Tallahassean.

Monica Hartsfield, Admin Specialist

Hartsfield likes to travel, and spends time pour painting when she can. She has two cats, including a 7 month old orange tabby.

For more on the department, go to: https://www.sc.fsu.edu/.
You may contact Hartsfield at m.hartsfield@fsu.edu.

approximations of the modes and eigenvalues of the composition operator [also called the Koopman operator].

"We say that DMD is data-driven." By using the inherent suppleness of mathematics, scholars can maintain and adaptably shift among multiple representations of numbers and problem-solving strategies. "Mathematics is the unique tool that allows elegant problem solving – it's one of the many benefits of math's built-in flexibility."

A scholar of world renown, Navon continually demonstrates research is most effective when those with expert knowledge in different areas collaborate on a project of overlapping interest. The overlap allows for common ground, while the respective areas of expertise cover a greater "surface area" of possible knowledge brought to bear on a specific question.

"The benefits of scientific collaboration are numerous," said Navon.

In the past few years, he added to his still growing catalog of research topics, having most recently undertaken subjects in machine learning and epidemiology. With those additions, one of his most recent publications adds to the body of knowledge on the world's single most urgent global crisis: COVID-19.
"I was inspired to undertake epidemiology after reading a 2015 seminal article written by Proctor and Eckhoff and by the work of scientists at the University of

see Navon, page 7

Sharkey spearheads software development

Eric Sharkey, MS

For the past year, in addition to rigorous studies as an SC grad student, Eric Sharkey has been working as a machine learning engineering intern at NewSci Labs, a machine learning/artificial intelligence think tank located in Tallahassee. The goal of NewSci is to use and develop state-of-the-art AI technologies to improve the public, health, and education sectors; the company develops novel AI-powered products and makes them available to industry.

Recently, the primary project Sharkey has worked on since interning at NewSci - Call Simulator -has been launched. "Call Simulator is an AI powered training program that we've developed over the past few years that can simulate a caller in an emergency situation," said Sharkey. "We use Natural Language Understanding techniques to comprehend the operator, and Natural Language Generation techniques to create a realistic response for] our AI caller. Emergency call center operators need to be trained to respond appropriately to a large variety of situations. We plan to use this software to improve the training quality of emergency call center operators by exposing them to a wide range of scenarios for them to practice."

Nathan Crock, SC doctoral grad and NewSci Labs Director, is enthusiastic about the work Sharkey has done to launch Call Simulator as well. "Our Call Simulator is a quintessential example of an AI technology created in our lab consistent with our "AI for good" motto. It is now deployed around the world, improving 911 emergency telecommunicator training programs, and helping prepare them to save more lives. Eric Sharkey and the skills he acquired at the Department of Scientific Computing were invaluable in the creation of this technology," Crock said.

NewSci Labs' Call Simulator software vastly improves on current instruction protocols in training emergency call/911 operators. "Our software replaces human trainers with an AI system. Trainees can run simulations anywhere at anytime, and can practice any simulation

as much as they need," said Sharkey. "We automatically score the operators to determine how they performed and how they can improve. In the future, we're planning to expand to a number of different areas that use scripted calls, for example telehealth."

NewSci has been working on the simulator since 2018, building a number of interdependent components needed for launching the software. The first core component was the AI system used to identify and respond to scripted questions. "Over the years, we built different iterations to create more realistic responses. Moving forward we are planning to create new versions of our simulator that use more complex Natural Language Processing techniques to handle more complicated calls and create more robust simulations."

Another component of the software that benefitted from Sharkey's expertise is the system that streams audio from the user's microphone to the simulator, generates the caller's speech, and returns the speech to the user all in real time. "We recently integrated with ProQA Discovery, a browser based tool that operators can use to run simulations while practicing entering call information." Sharkey graduated with his MS in computational science in the summer 2021 class, and plans to continue his work on the project. "At NewSci Labs, I work with machine learning researchers and developers on interesting and challenging problems. It's been a great experience to work alongside the talented members of our team at NewSci, and I've learned and grown so much from them," Sharkey said. "I've accepted a full-time position as a machine learning engineer at NewSci, and I'm going to be the technical lead on developing the Call Simulator software."

Crock is glad to have Sharkey and his advanced knowledge on staff. "With the skillset Eric acquired in the Department of Scientific Computing's graduate program, he can approach various problems in the machine learning discipline," Crock said. "He is already helping us develop the next-generation simulator by creating faster and more efficient paraphrasing, text-style transfer, and efficient question-answering models."

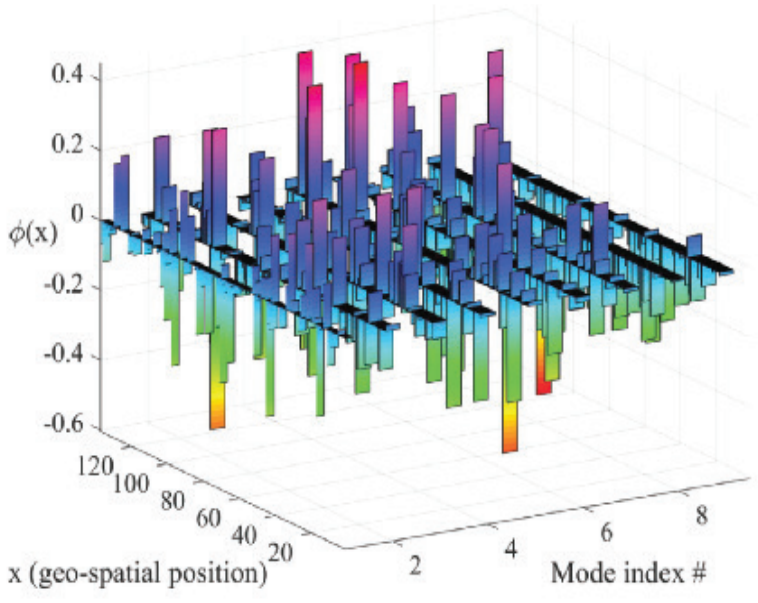
For more, go to sc.fsu.edu & https://newsci.ai/.

Navon, continued from page 5

Washington in Seattle. These studies of infectious disease data used Dynamic Mode Decomposition, and I wanted to try using that model. DMD can analyze snapshots of infectious disease data from experiments, numerical simulations, or historical records to extract coherent spatiotemporal patterns.

"Data collected from separate spatial locations can be formed into vectors representing state snapshots that evolve in time. Prof Bistrian developed an ingenious economical way to implement DMD- the Randomized DMD.

"The dynamic modes discovered by



a.

The leading DMD modes used to generate the reduced order model.

Department of Scientific Computing
400 Dirac Science Library
P. O. Box 3064120
Tallahassee, FL 32306-4120
www.sc.fsu.edu

First Class
US Postage
PAID
Tallahassee FL
Permit 55

The department's mission is to be the focal point of science and computation at Florida State University. Gordon Erlebacher is Chair of the Department of Scientific Computing. He can be reached at 850.644.7024. Newsletters are issued three times each year. Subscriptions and single copies are available by calling 850.644.0196. This publication is available in an alternative format on request.

DMD allow for a number of interesting and relevant epidemiological interpretations of large-scale dynamic patterns of infectious disease spread. DMD can automatically identify the frequency content of time series. In addition to identifying oscillations, the dynamic modes associated with the related eigenvalues can indicate the relative phase of oscillation by examining the angle of each element in the dynamic mode, as shown in both flu and measles data examples. The phase information alone can be useful for allocating vaccine resources for the year, sending surveillance teams to monitor the disease, and time interventions to leverage natural disease dynamics."

Working with Professors Diana A. Bistrian and Gabriel Dimitriu, two Romanian university affiliates, Navon requested and received data from the World Health Organization (WHO). The data consisted of the number of COVID-19 infections registered daily between February 28 – May 18, 2020. This is an 80-day period that includes data from twelve countries, including the North American continent and Europe.

Using the adaptive randomized dynamic mode decomposition algorithm (ARDMD) the researchers identified the optimal rank for the mathematical model of the infection spread. They then used a small part of the raw data to initiate the predictive model and the remaining data to validate the prediction results.

This method is especially efficient in the case of large amounts of snapshot data. The method is data-driven in that it does not require knowledge of the underlying governing equations, only snapshots of state and actuation data from historical, experimental, or black-box simulations.

"We demonstrate the method on high-dimensional dynamical systems, including a model with relevance to the analysis of infectious disease data with mass vaccination (actuation)."

For more on the Department of Scientific Computing, go to sc.fsu.edu.

sc.fsu.edu