Newsjournal of the Society for Industrial and Applied Mathematics sinews.siam.org

siam news

Volume 54/ Issue 5
June 2021

Fighting a Pandemic with Medical Imaging and Machine Learning: Lessons Learned from COVID-19

By Michael Roberts, Derek Driggs, Ian Selby, Evis Sala, and Carola-Bibiane Schönlieb

Many researchers in applied mathematics have recognized that machine learning (ML) models can help clinicians during the ongoing COVID-19 pandemic. Models that utilize chest radiographs and computed tomography (CT) scans to detect patients with severe cases of COVID-19 have received considerable attention over the last year. However, pervasive systematic errors in existing research make it dangerous for doctors to consider these models for clinical use.

Here we identify a few of these complications and discuss avenues that researchers can explore to close the gap between model development and model deployment. In particular, we emphasise the need to (i) source high-quality data, wherein the biases are understood and appreciated; (ii) incorporate data from multiple sources (as a clinician would when making a decision); and (iii) employ a multidisciplinary team for model development.

The work, effort, and dedication of the mathematical imaging and ML communities during this pandemic has been inspirational and clearly shows ML's potential for clinical decision support. It also demonstrates the possible pitfalls of ML during a global emergency.

By focusing on image-based diagnosis and prognosis for COVID-19, we make several observations about the quick and reliable development of ML-based clinical support tools. Our consequential discussion stems from a recent systematic review [12] and editorial piece [4].

Imaging and Machine Learning for the COVID-19 Pandemic

Chest imaging is a useful tool for the initial triage of patients with COVID-19 at hospitals with first-line utilization of chest X-rays (CXRs) and CT imaging. Although European and American radiological societies initially discouraged the use of CT and chest radiographs for COVID-19 diagnosis in early 2020 [1, 13], this position softened as the high false negative rate of existing tests became apparent and the pandemic began to strain the resources that are required to test patients quickly [14]. China has employed imaging exams as the primary initial diagnostic tool since the outbreak's onset [13]. In addition, several studies indicate that the extent of opacification in the lungs of COVID-19 patients is a significant prognostic marker of mortality [2]. Figure 1 displays common presentations of COVID-19 in CT scans and chest radiographs.

The COVID-19 pandemic is the first of the ML era, and pattern recognition algorithms have the potential to aid clinicians in the diagnosis and prognostication of

See **Medical Imaging** *on page 3*

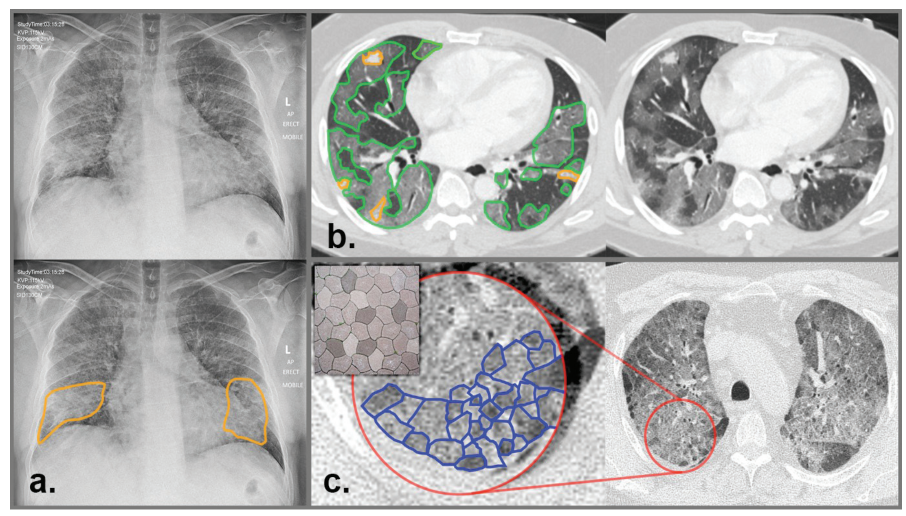

Figure 1. *Annotated examples of COVID-19 scans.* **1a.** *Chest X-ray (CXR) with ground-glass opacification (GGO) in both lungs and consolidation (outlined in orange).* **1b.** *A computed tomography (CT) scan that shows GGO (green) and consolidation (orange).* **1c.** *A CT scan that indicates severe COVID-19 with a crazy-paving pattern. Images courtesy of [8] and inset courtesy of [5].*

Nonprofit Org
U.S. Postage
PAID
Permit No 360
Bellmawr, NJ

siam. SOCIETY for INDUSTRIAL and APPLIED MATHEMATICS
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA

Balancing Homeostasis and Health

By Matthew R. Francis

Human beings are not bicycles. However, mechanistic metaphors for the human body abound. For instance, we compare athletes to finely-tuned machines and look for equations that are derived from mechanics to describe biological processes — even when the relationship is no better than an analogy.

However, the concept of homeostasis clearly exemplifies the breakdown of mechanistic models when one applies them to the human body. Homeostasis is the process by which an organism maintains a stable output regardless of input (within reasonable limits). The most familiar example is human body temperature, which stays within a remarkably small range of values regardless of whether one is sitting in a cold room or walking outside on a hot day.

"In a bicycle, you know what each part is for," Michael Reed, a mathematician at Duke University, said. "We are not machines with fixed parts; we are a large pile of cooperating cells. The question is, how does this pile of cooperating cells accomplish various tasks?"

More specifically, how can researchers mathematically describe the function of these processes? Homeostasis is the conceptual opposite of bifurcation, wherein a small variation in initial conditions results in a massive change to the system's evolution. But it is also unlike dynamical equilibrium, in which system perturbations often lead to oscillations.

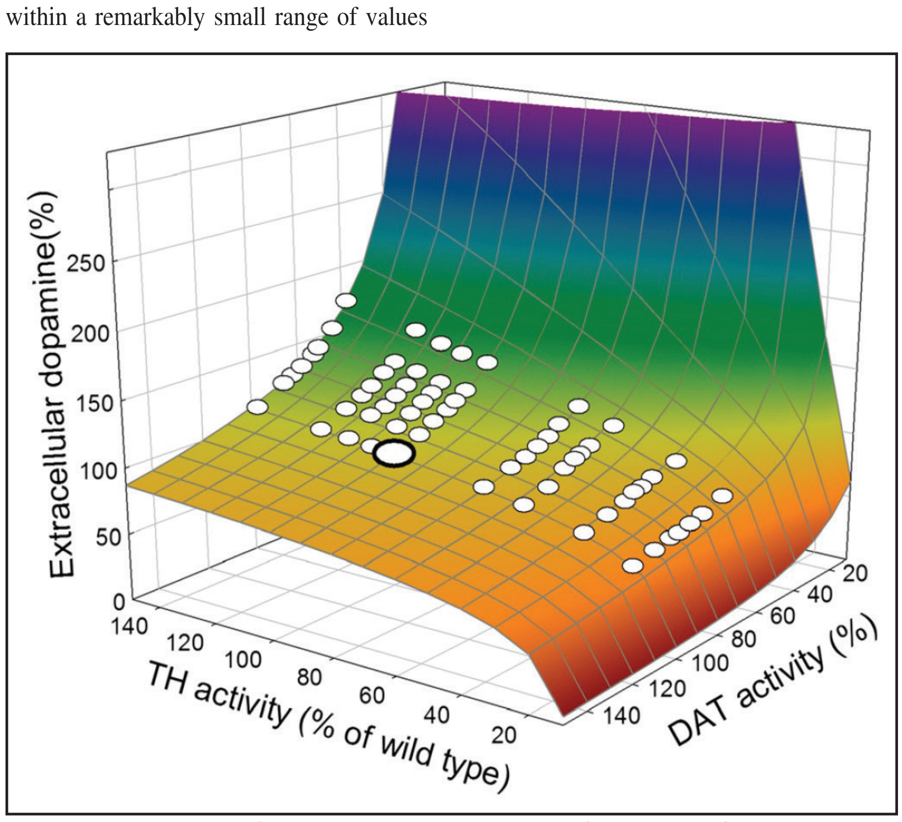

Figure 1. *Homeostasis of dopamine under a wide range of production of the enzyme tyrosine hydroxylase (TH) and the dopamine reuptake transporter (DAT), which are to varying degrees controlled by genes. The stated values are all relative to an average genotype, which the authors of the study call the "wild type." Figure courtesy of [1] and reproduced with permission under the Creative Commons Attribution 4.0 International License: https://creativecommons.org/licenses/by/4.0.*

All living things rely on homeostatic mechanisms, and homeostasis is an essential component of a wide range of biological and biomedical phenomena, including cancer growth, physiological responses to drugs, and hormone therapies. Mathematical modeling of these systems—in tandem with animal experiments—promises new ways to treat diseases and imbalances, both by identifying healthy homeostatic schemes and disrupting homeostasis in pathogens or tumors.

"Not only are these very complicated biological mechanisms, but overlaid on top of the biochemistry and physics are various control mechanisms that adjust for this kind of variation," Reed said. During his presentation at the 2021 American Association for the Advancement of Science (AAAS) Annual Meeting,[1] which took place virtually this February, Reed noted that the amount of a particular enzyme that is produced in the liver can vary as much as 25 percent without affecting liver function. Such preservation requires a fine level of control that results from a complex nonlinear relationship between input, regulation, and output functionality.

"Homeostasis is a biological concept, [so] you have to make it into a mathematical concept," Martin Golubitsky, a mathematician at Ohio State University who also spoke at the AAAS session, said. "It's a really difficult mathematical problem analytically; what does it mean to be 'approximately constant'? You know it when you see it to some extent, but you don't know how to search for it very easily."

Homeostasis Giveth and Homeostasis Taketh Away

Human beings are *like* bicycles, at least in a dynamical sense. The forward motion that stems from pedaling helps to maintain balance and keep the bicycle from falling over. Similarly, equilibrium for a living thing is

See **Homeostasis and Health** *on page 4*

[1] https://www.aaas.org/events/2021-aaas-annual-meeting

siam news
Volume 54/ Issue 5/ June 2021

SIAM Partakes in the 2021 Virtual National Math Festival

By Wesley Hamilton

Every two years, the Mathematical Sciences Research Institute—in cooperation with the Institute for Advanced Study and the National Museum of Mathematics—organizes the National Math Festival[1] (NMF) to bring together mathematicians, math educators, and the broader community. The NMF typically takes place in Washington, D.C., but it occurred virtually this year due to the ongoing COVID-19 pandemic. The festival ran from Friday, April 16 through Sunday, April 18 and was hosted on the Hopin event platform (see Figure 1).

This year's event included five film screenings and associated panels, seven special math talks (with subjects ranging from black holes and spacetime to "Changing the 'Face' of Mathematics"[2]), an "Imagine Math Class" video competition[3] for youths aged 13-18, and a plethora of interactive sessions and booths. The festival began on Friday night with several screenings and a Mathical book reading,[4] but most festival activities took place over the weekend. All of the interactive sessions and booths were open to attendees on Saturday and Sunday.

Steven Strogatz delivered a main stage talk titled "Infinite Powers: The Story of Calculus."[5] During his presentation, which was meant for a general audience, Strogatz discussed the history of calculus and how it has shaped modern science. Austin Ferguson, a mathematics graduate student at the University of North Carolina at Chapel Hill, attended and enjoyed the talk. "It was great listening to Dr. Strogatz talk about the basics of calculus," he said. "Hearing a professor who I respect so much, whose textbook helped steer me in the direction I've gone, explain something so clearly and with so much passion was wonderful."

[1] https://www.nationalmathfestival.org
[2] https://www.nationalmathfestival.org/event/changing-face-mathematics
[3] https://www.nationalmathfestival.org/join/imagine-math-class
[4] https://www.nationalmathfestival.org/event/mathical-author-reading-brittney-morris-slay
[5] https://www.nationalmathfestival.org/event/infinite-powers-story-calculus

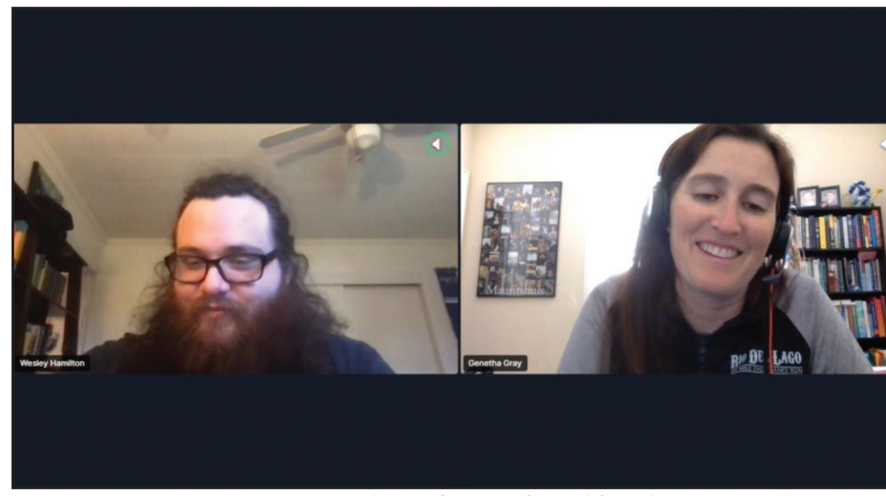

Figure 2. *Author Wesley Hamilton (left) and Genetha Gray of Salesforce prepare for a "Meet a Mathematician" session at the virtual National Math Festival (NMF) in April 2021.*

Another main stage event featured the founders of Mathematically Gifted and Black:[6] Erica Graham (Bryn Mawr College), Raegan Higgins (Texas Tech University), Candice Price (Smith College), and Shelby Wilson (Johns Hopkins University Applied Physics Laboratory). Their session, which explored the genesis and impact of their website, was followed by a question-and-answer (Q&A) period with the audience. The four mathematicians discussed the importance of representation in mathematics, addressed the ways in which media portrayal of mathematicians has shifted in the past few decades, and commented on the origins of their website. "Typically [in the media] we see white men as mathematicians," Graham said. "If you try to think about some sort of counterexample to that, there are very rarely additional representations of mathematicians who look like the four of us."

This presentation highlighted a second NMF goal: increasing representation and humanizing mathematics in a way that the standard math curriculum often overlooks. A selection of film screenings and associated Q&As also furthered this objective:

– 2016's *Hidden Figures*, about the African American female mathematicians who worked for NASA during the Space Race

– *Secrets of the Surface*, the 2020 documentary about Maryam Mirzakhani — the first female and Iranian recipient of the Fields Medal

– *The Man Who Knew Infinity*, the 2015 biopic of Indian mathematician Srinivasa Ramanujan.

Many of the special events were recorded and are available on the NMF website.

The SIAM Education Committee[7] hosted a number of sessions and a booth at the NMF to encourage festival attendees to interact with applied mathematicians and learn about careers in applied math. We decided to organize "Meet a Mathematician" sessions, which were inspired by a similar series of interviews by the Girls Talk Math summer camp[8] in 2020. We invited seven mathematicians from diverse backgrounds and fields to speak about their lives, mathematical experiences, and use of math in their careers. The speakers were as follows:

– Torina Lewis of the American Mathematical Society (AMS) detailed the AMS's efforts to organize programs that engage, support, advance, and uplift the entire mathematical community

– Kerisha Burke of Phillips 66 spoke about working in the energy industry as a midstream analyst

– Mario Banuelos of California State University, Fresno explained how he uses math and computer programming to study biology, computer vision, and health

– Tim Chartier of Davidson College discussed his work with sports analytics

– Aaron Luttman of Pacific Northwest National Laboratory talked about his nuclear security research

– Genetha Gray of Salesforce described how she incorporates mathematical models in human resource departments (see Figure 2)

– Sara Del Valle of Los Alamos National Laboratory discussed epidemiology and her responsibilities at the lab.

These sessions were well attended—most had roughly 30 unique attendees—and generated engaging dialogue. Moreover, we compiled audience questions into a central document that the SIAM Education Committee will use when preparing for future outreach endeavors and festival appearances. SIAM volunteers helped the sessions run smoothly by monitoring the chat for questions, managing virtual attendees, and otherwise engaging in conversations.

Ferguson volunteered for two "Meet a Mathematician" sessions and reflected on the value of this type of outreach. "As a kid, I never actually saw what being a

See **National Math Festival** *on page 5*

[6] https://mathematicallygiftedandblack.com

[7] https://www.siam.org/about-siam/committees/education-committee
[8] https://girlstalkmath.com

Figure 1. *Attendees at the virtual National Math Festival (NMF), which took place in April 2021, entered via the Hopin Lobby.*

siam news

ISSN 1557–9573. Copyright 2021, all rights reserved, by the Society for Industrial and Applied Mathematics, SIAM, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688; (215) 382-9800; siam@siam.org. To be published 10 times in 2021: January/February, March, April, May, June, July/August, September, October, November, and December. The material published herein is not endorsed by SIAM, nor is it intended to reflect SIAM's opinion. The editors reserve the right to select and edit all material submitted for publication.

Advertisers: For display advertising rates and information, contact Kristin O'Neill at marketing@siam.org.

One-year subscription (nonmembers): Electronic-only subscription is free. $73.00 subscription rate worldwide for print copies. SIAM members and subscribers should allow eight weeks for an address change to be effected. Change of address notice should include old and new addresses with zip codes. Please request address change only if it will last six months or more.

Medical Imaging

Continued from page 1

COVID-19 via chest imaging [11]. Indeed, our systematic review [12]—which examines the entire literature from January 1, 2020 to October 3, 2020—identifies 320 published and preprint manuscripts that develop ML models using chest CT or radiographs for COVID-19 diagnosis or prognostication. Unfortunately, most papers contained systematic issues pertaining to image sourcing, quality, and documentation that introduce bias in developed models, ultimately making them unlikely to perform well in practice [4].

We now highlight some of the methodological issues in these studies and provide recommendations for the creation of ML models that incorporate imaging features and are suitable for clinical use. Fundamentally, we endorse the acquisition of high-quality data (preferably temporal) and detailed associated metadata.

Sourcing Issues

CXRs and CT scans for COVID-19 patients are commonly available in public repositories that do not independently verify the accuracy of the ground-truth labels of COVID-19 diagnosis. Researchers also frequently use a large pediatric pneumonia cohort [9] as the control "non-COVID-19" group during model development, which inadvertently trains a model to distinguish adults (COVID-19) from young children (non-COVID-19); the imaging differences in Figure 2 demonstrate this point. Furthermore, we discovered a prevalence of "Frankenstein" datasets, which are compiled from existing datasets and released under a new name. Scientists who train models with such Frankenstein datasets are unknowingly testing their models on *overlapping data* and thus producing optimistic performance metrics. Finally, many manuscripts utilize data from image repositories that consist of COVID-19 images from publications, preprints, or social media posts without acknowledging these images' potential bias towards more interesting or unusual cases.

Quality Issues

Image quality can also drastically affect a model's performance and robustness. In our review, we found that researchers rarely discuss image pre-processing steps. Image resizing is a common pre-processing step for deep learning models, but the effects of initial image resolution, input resolution, and aspect ratio adjustments on a model's performance are unclear. We also do not know whether an image that one extracts from a publication—a "picture of a picture"—contains the same level of useful information as the original [10].

Documentation Issues

Although researchers should fully document the source and metadata for the images that contribute to model development, such documentation is lacking in most of the literature. For example, one must know whether radiographs were taken with mobile scanners and understand the distribution of scanner manufacturers, CT reconstruction kernels, and CT slice thicknesses. Without this information, scientists cannot comprehend biases in the dataset.

Multiple Data Streams

In the absence of a polymerase chain reaction test result or a suspected false positive/negative, one can make a clinical diagnosis with information from multiple data sources. A ML model for use in clinical trials must endeavour to emulate this ability.

In late February 2020, the Diamond Princess cruise ship had the largest cluster of positive COVID-19 cases outside of China. A study of 104 of these COVID-19-positive patients found that 73 percent (76 out of 104) were asymptomatic. However, 54 percent (41 out of 76) of these asymptomatic individuals displayed lung opacities on their CT scans. The converse was also true, as roughly 21.5 percent (six out of 28) of symptomatic patients had normal CT findings [6]. Imaging features alone are clearly not sufficient for accurate diagnosis.

Multidisciplinary Approach

Clinicians, ML experts, imaging specialists, mathematicians, and statisticians should all partake in the development process of a trustworthy model for clinical use. Clinical insight regarding model usability and data quality is invaluable. For instance, clinicians know that unstable patients may be primarily imaged from the front (i.e., anteroposterior) or whilst lying on their backs (supine) if they are critically unwell. Annotations that indicate this specification are commonly burnt into CXR images, and state-of-the-art models that classify lung pathologies on the widely used CheXpert dataset [7] frequently employ these annotations to inform the models (see Figure 3). Although such insights are not obvious to a non-clinician, they should heavily influence the development of ML algorithms to avoid irrelevant links between radiographs and outcomes.

Eye on the Prize

Unfortunately, most models have no viable path towards regulation and clinical use. The many hastily developed, poor-quality models in some manuscripts risk polluting the entire literature, obscuring high-quality models, and alienating clinicians who are eager to embrace ML methods. Maintaining a clear path towards the clinical adoption of algorithms throughout their development process—and working with relevant industrial partners and healthcare authorities—is crucial for ensuring model suitability for clinical implementation.

Acknowledgments: The authors thank their colleagues and partners in the AIX-COVNET collaboration[1] for their contributions and input to this article, which is the result of many months of discussions and teamwork that began in March 2020.

[1] https://covid19ai.maths.cam.ac.uk

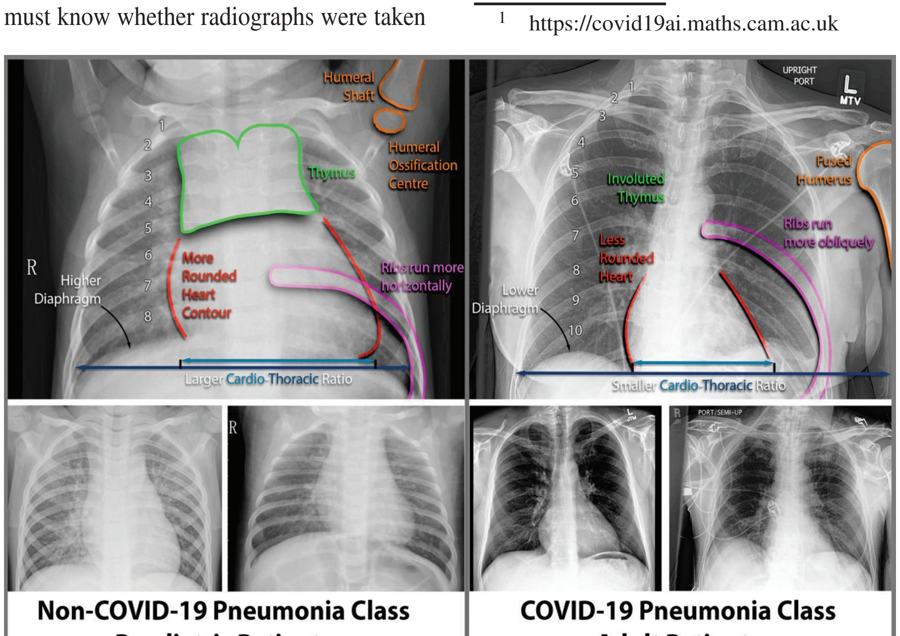

Figure 2. *Annotated chest radiograph indicating the differences between pediatric and adult patients. Pediatric scans courtesy of [9] and adult scans courtesy of [3].*

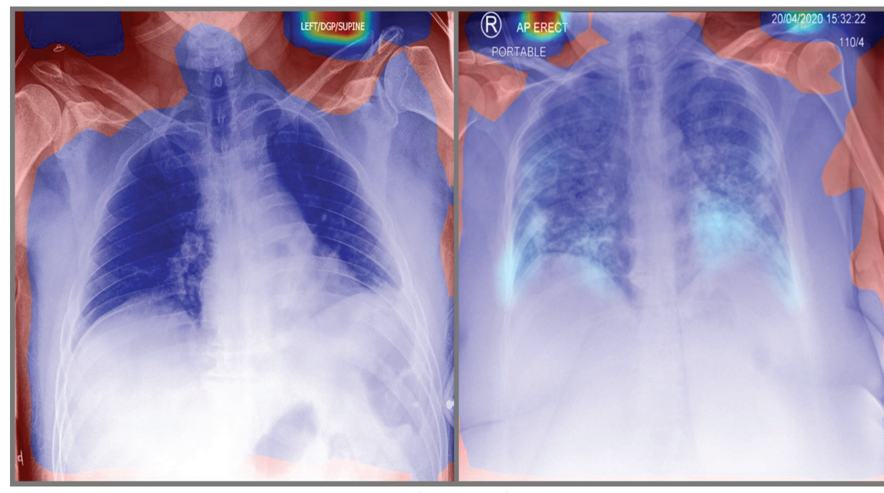

Figure 3. *Activation map depicting regions of interest for a neural network that detects lung abnormalities. Red areas indicate discriminative regions that the model uses to make its predictions. Original scans courtesy of the National Collaborating Centre for Infectious Diseases dataset [8].*

References

[1] American College of Radiology. (2020, March 22). ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. Retrieved from https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection.

[2] Cleverley J., Piper, J., & Jones, M.M. (2020). The role of chest radiography in confirming covid-19 pneumonia. *Brit. Med. J.*, *370*, m2426.

[3] Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., & Ghassemi, M. (2020). COVID-19 image data collection: Prospective predictions are the future. Preprint, *arXiv:2006.11988*.

[4] Driggs, D., Selby, I., Roberts, M., Gkrania-Klotsas, E., Rudd, J.H.F., Yang, G., …, Schönlieb, C.-B. (2021). Machine learning for COVID-19 diagnosis and prognostication: Lessons for amplifying the signal whilst reducing the noise. *Radiol. Art. Int.*, e210011.

[5] Harvey, D.M. (2005, June 25). Example of wallpaper group type p3. Computer-enhanced photograph of a street pavement in Zakopane, Poland. Retrieved from https://commons.wikimedia.org/wiki/File:Wallpaper_group-p3-1.jpg.

[6] Inui, S., Fujikawa, A., Jitsu, M., Kunishima, N., Watanabe, S., Suzuki, Y., …, Uwabe, Y. (2020). Chest CT findings in cases from the cruise ship Diamond Princess with coronavirus disease (COVID-19). *Radiol. Cardio. Imag.*, *2*(2).

[7] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., …, Ng, A.Y. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the thirty-third AAAI conference on artificial intelligence* (pp. 590-597). Honolulu, HI: Association for the Advancement of Artificial Intelligence.

[8] Jacob, J., Alexander, D., Baillie, J.K., Berka, R., Bertolli, O., Blackwood, J., …, Joshi, I. (2020). Using imaging to combat a pandemic: Rationale for developing the UK national COVID-19 chest imaging database. *Euro. Respir. J.*, *56*(2).

[9] Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., …, Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*(5), 1122-1131.

[10] Lensink, K., Parker, W., & Haber, E. (2020, July 13). Deep learning for COVID-19 diagnosis. *SIAM News*, *53*(6), p. 1.

[11] Rajpurkar, P., Joshi, A., Pareek, A., Chen, P., Kiani, A., Irvin, J., …, Lungren, M.P. (2020). CheXpedition: Investigating generalization challenges for translation of chest X-ray algorithms to the clinical setting. Preprint, a*rXiv:2002.11379*.

[12] Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., …, Schönlieb, C.-B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Int.*, *3*, 199-217.

[13] The Royal College of Radiologists. (2020, March 12). The role of CT in patients suspected with COVID-19 infection. Retrieved from https://www.rcr.ac.uk/college/coronavirus-COVID-19-what-rcr-doing/rcr-position-role-ct-patients-suspected-COVID-19.

[14] Yang, W., Sirajuddin, A., Zhang, X., Liu, G., Teng, Z., Zhao, S., & Lu, M. (2020). The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). *Eur. Radiol.*, *30*(9), 1-9.

Michael Roberts is a Senior Research Associate in Applied Mathematics at the University of Cambridge and a Senior Postdoctoral Fellow in Oncology R&D at AstraZeneca. Derek Driggs is a Ph.D. student in applied mathematics at the University of Cambridge. Ian Selby is a Clinical Research Associate at the University of Cambridge and an honorary radiology registrar at Cambridge University Hospitals NHS Foundation Trust. Evis Sala is a professor of oncological imaging at the University of Cambridge and an honorary consultant radiologist at Addenbrooke's Hospital. Carola-Bibiane Schönlieb is a professor of applied mathematics at the University of Cambridge.

SIAM Members Elected to the U.S. National Academy of Sciences

In late April, the National Academy of Sciences (NAS) announced the election of 120 new members—59 of whom are women, the most ever elected in one year—and 30 international members, bringing the total number of active members to 2,461 and the total number of international members to 511. Inductees are honored for their distinguished and continuing achievements in original research. SIAM offers its sincere congratulations to all inductees, including those SIAM members who were elected this year:

- **Peter Constantin,** *director of the Program in Applied and Computational Mathematics and John Von Neumann Professor, Department of Mathematics, Princeton University*
- **Glenn H. Fredrickson,** *Mitsubishi Chemical Chair in Functional Materials, Department of Chemical Engineering, University of California, Santa Barbara*
- **Kenneth Lange,** *Rosenfeld Professor of Computational Genetics, Department of Computational Medicine and Departments of Human Genetics and Statistics, University of California, Los Angeles*
- **Randall J. LeVeque,** *professor emeritus, Department of Applied Mathematics, University of Washington, Seattle*
- **Linda Petzold,** *distinguished professor, Department of Computer Science, University of California, Santa Barbara*

A full list of the newly elected members of the NAS is available online.[1]

[1] http://www.nasonline.org/news-and-multimedia/news/2021-nas-election.html

Homeostasis and Health

Continued from page 1

death; homeostasis upholds the balances that keep organisms alive and healthy.

Reed's research involves the creation of mathematical models that analyze the body's regulation of dopamine — a chemical that helps transmit signals between cells, particularly in the nervous system. Low dopamine levels are a key factor in Parkinson's disease, while hyperactive dopamine receptors may be associated with schizophrenia.[2] Reed and his collaborators found that their model displayed an impressive stability in dopamine levels across a wide range of genetic variations that produce different levels of regulatory enzymes (see Figure 1, on page 1). This stability serves as a strong demonstration of homeostasis in the dopamine cycle [1].

In some cases, immune systems can hijack homeostatic processes to fight pathogens. Reinhard Laubenbacher, director of the Laboratory for Systems Medicine at the University of Florida, is particularly interested in the way in which cancer disrupts iron metabolism and the immune system's ability to stop fungal infections. During his AAAS talk, Laubenbacher described how certain fungi parasitize iron from lung cells. Under ordinary circumstances, however, the immune system disrupts homeostasis in the invader cells' iron metabolism to starve them of this vital element and ultimately kill them.

This mechanism for interrupting homeostasis provides researchers with possible alternative treatment options beyond antifungal medication. Like antibiotics, many antifungals are becoming less effective as fungi evolve. "We work very closely with immunologists," Laubenbacher said. "They might say, 'An immunocompromised patient doesn't have a certain type of white blood cell, so how can we make up for it? What if we inject some more of this kind of substance, would that make a difference?' In the [computer] model, we can do that and see if it does make a difference."

Such results inform laboratory experiments that can lead to new treatments, particularly when other therapies are ineffective. While administering drugs to an already weakened patient might be harmful, interventions that restore homeostasis after its disruption could be possible.

Networks and Nodes

Human beings are like networks. Despite having individual roles, cells work together to ensure the function of organs and the entire body. The output of a single essential biochemical might involve many cell types across multiple organs, but sometimes one can abstract the process when studying homeostasis. A three-node network that consists of a single input ι, output o, and regulatory node ρ serves as a simple example. Researchers can model networks of these three nodes to produce three types of homeostasis, one of which corresponds conceptually to the dopamine regulatory system that Reed and his colleagues described.

Consider a network wherein a given input I leads to output $x_o(I)$. Homeostasis occurs when $x_o'(I_0)=0$ for some input value I_0 (the derivative is with respect to I). One can write the three-node network generically as a system of three differential equations:

$$\dot{x}_\iota = f_\iota(x_\iota, x_\rho, x_o, I)$$
$$\dot{x}_\rho = f_\rho(x_\iota, x_\rho, x_o)$$
$$\dot{x}_o = f_o(x_\iota, x_\rho, x_o).$$

The choice of functions f_k describes specific biochemical networks. For instance, for a "feedforward" excitation in which the biochemical substrate activates an enzyme that removes a product,

$$\dot{x}_\iota = f_\iota(x_\iota, I) = I - g_1(x_\iota) - g_4(x_\iota)$$
$$\dot{x}_\rho = f_\rho(x_\iota, x_\rho) = g_1(x_\iota) - g_2(x_\rho) - g_5(x_\rho)$$
$$\dot{x}_o = f_o(x_\iota, x_\rho, x_o) = g_2(x_\rho) - h(x_\iota) g_3(x_o).$$

Particular choices of the functions $\{g_n, h\}$ lead to homeostatic behavior. Other homeostatic networks involve different combinations of variables and functions [2].

Three-node networks are mathematically tractable. However, they are not descriptive of many realistic systems, which could have as many as 50 nodes. "How many four-node networks are there? 199!" Golubitsky said. "How many different homeostasis configurations are there? 20. That's huge!"

However, some of these seemingly complicated systems might reduce through symmetries or redundancies. Golubitsky and his collaborators are investigating this possibility by drawing on graph theory and catastrophe theory for guidance.

So Very Simple, Only a Child Can Do It

Human beings are not actually networks. One cannot reduce Laubenbacher's models for homeostasis in iron metabolism to the type of differential equations that Golubitsky and Reed use — at least not yet. Laubenbacher also distinguishes between complicated and complex systems; though complex systems may be conceptually simple, complicated homeostatic systems often require greater levels of detail.

"[Our model] is a multiscale model," Laubenbacher said. "It has intracellular networks, tissue-level phenomena, [and] a whole-body component. It's made up of molecule diffusion, partial differential equations, and cells that are moving around. The model is really a hybrid of mechanistic and phenomenological modeling and has altogether maybe 150 variables."

Laubenbacher then paraphrased Ludwig Wittgenstein's aphorism, *"Whereof one cannot speak, thereof one must be silent,"* and added his own interpretation. "Mathematics provides a language for you to formulate the properties of the systems that you encounter in the life sciences," he said. "The goal is to be really translational. We would actually like to say that if you treat these patients in this particular way, it's going to make a difference."

Homeostasis can simultaneously be similar and dissimilar to both bicycles and networks. Researchers might need to utilize novel mathematics to grasp these contradictions, simply because life is not mechanical. As with the early years of nonlinear dynamics and chaos theory, much current work in this field involves categorizing and searching for global patterns that suggest the underlying order. In the end, the mathematics of homeostasis may lead to a new and deeper understanding — and ultimately save lives.

References

[1] Nijhout, H.F., Best, J.A., & Reed, M.C. (2015). Using mathematical models to understand metabolism, genes, and disease. *BMC Biol.*, *13*, 79.

[2] Golubitsky, M., Stewart, I., Antoneli, F., Huang, Z., & Wang, Y. (2020). Input-output networks, singularity theory, and homeostasis. In O. Junge, O. Schütze, G. Froyland, S. Ober-Blöbaum, & K. Padberg-Gehle (Eds.), *Advances in dynamics, optimization, and computation*. Cham, Switzerland: Springer.

Matthew R. Francis is a physicist, science writer, public speaker, educator, and frequent wearer of jaunty hats. His website is BowlerHatScience.org.

[2] https://www.sciencedirect.com/topics/neuroscience/dopamine-hypothesis-of-schizophrenia

Mathematical Models of Traffic Flow

By Helge Holden and Nils Henrik Risebro

Preliminary mathematical models of traffic flow date back to at least the early 1950s [10]. Two distinct classes of models prevail: follow-the-leader (FtL) models and traffic hydrodynamics. The former tracks individual vehicles, while the sufficiently dense traffic in the latter justifies a continuum approach wherein vehicle density is the fundamental quantity. Given the prevalence of traffic in modern society, the development of unprecedented computer power and progressive tracking devices, and continued advances in mathematical research, mathematical traffic modeling has become increasingly important in recent years.

Here we describe a novel mathematical model that allows for the analysis of multilane traffic [9]. But first we start with the basics. Consider dense unidirectional traffic on a single lane. At the most fundamental level, your velocity is determined by your distance from the vehicle just ahead — the closer you are, the slower you drive. If z_i is the position of the ith vehicle on a single-lane road, we can model this simple point by

$$\frac{d}{dt}z_i = v\left(\frac{\ell}{z_{i+1}-z_i}\right). \qquad (1)$$

Here, z_{i+1} is the position of the vehicle directly in front of you, ℓ is the vehicle length, and v is a decreasing velocity function. This calculation amounts to the FtL model and generates a system of ordinary differential equations with a size that is equal to the number of vehicles. However, society is getting too accustomed to (prohibitively) dense traffic, for which a "particle" description becomes inadequate. It is thus natural to wonder whether we can take advantage of available mathematical technology to study the continuum limit of particle models in the context of traffic dynamics. It turns out that we can.

Define $\rho_i = \ell/(z_{i+1}-z_i)$; a straightforward calculation then yields

$$\frac{d}{dt}\left(\frac{1}{\rho_i}\right) - \frac{1}{\ell}\Big(v(\rho_{i+1}) - v(\rho_i)\Big) = 0. \qquad (2)$$

Now we can let $\ell \to 0$ and $\#(\text{vehicles}) \to \infty$ with $z_i = i\ell$ fixed to obtain $\rho_i(t) \to \rho(t,z)$, where ρ satisfies

$$\frac{\partial}{\partial t}\left(\frac{1}{\rho}\right) - \frac{\partial}{\partial z}v(\rho) = 0, \qquad (3)$$

since (2) is a first-order (in ℓ) semi-discrete scheme for (3). In this case, the continuum limit follows from a result in finite difference approximations of nonlinear partial differential equations. Equation (3) is an example of a first-order hyperbolic conservation law. Solutions of the Cauchy problem for this equation develop singularities in finite time that are independent of the initial data's smoothness. We therefore must develop the machinery of weak solutions and entropy conditions to single out unique solutions [6]. Because (3) is not in the standard form, we introduce the classical transformation—well-known from fluid dynamics—from Lagrangian to Eulerian variables. This transformation yields

$$\rho_t + (\rho v(\rho))_x = 0,$$

which is the celebrated Lighthill-Whitham-Richards (LWR) model for traffic flow. In the simplest case, we assume that the velocity depends only on density. If we scale the maximum density to unity and assume a linear dependence in the velocity—that is, $v = 1-\rho$—we obtain the equivalent of the inviscid Burgers' equation

$$\rho_t + (\rho(1-\rho))_x = 0.$$

We have thus connected the two most fundamental traffic models by establishing the convergence of a numerical scheme. By examining more general velocity functions and allowing these functions to depend on time and position, we see that the "hydrodynamic" approach to traffic on a single-lane road is a rich source of interesting mathematical problems — even in this very simple case.

The aforementioned reasoning is formal and assumes the differentiability of all quantities, but one can rigorously establish that the limit ρ exists and is an entropy solution [3, 7, 8] (see Figure 1).

Two Lanes

We model two lanes of traffic as two individual roads, where vehicles move according to the FtL model (1) and are allowed to change lanes. Our basic assumption is that the likelihood of a driver changing lanes is zero if doing so would lead to a decrease in speed, and is otherwise proportional to the potential gain in speed. This simple idea is a bit complicated to describe mathematically.

Let $\{z_i\}$ and $\{y_i\}$ denote the vehicle positions in the two lanes z and y respectively. We assume that the drivers continuously monitor the prospective speeds (and

See **Traffic Flow** *on page 7*

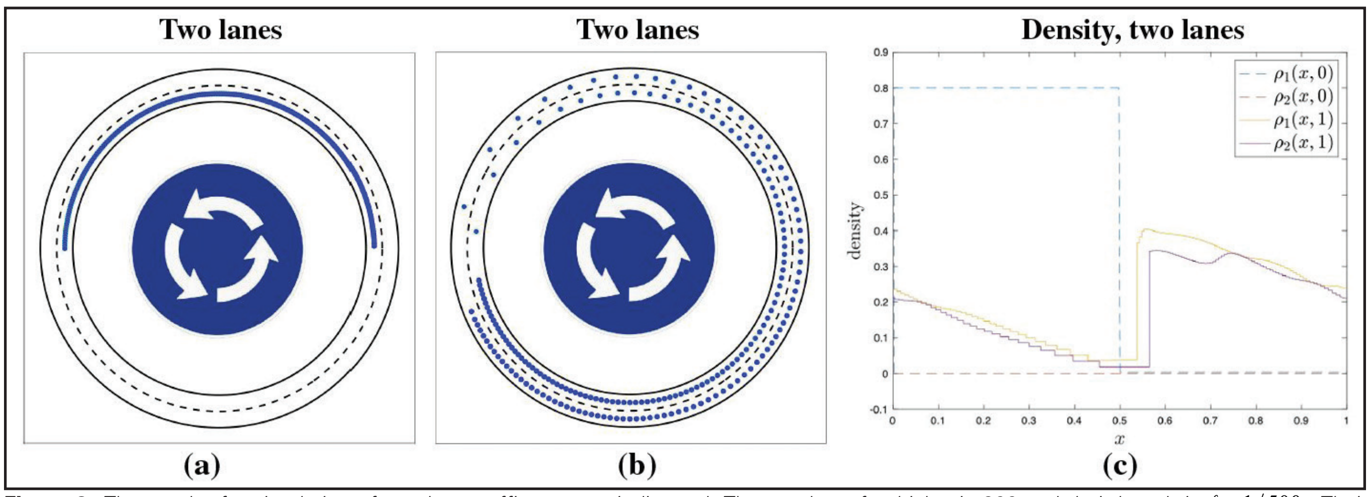

Figure 2. *The result of a simulation of two-lane traffic on a periodic road. The number of vehicles is 200 and their length is $\ell = 1/500$. Their initial locations are such that on half of the road, $x \in [0, 1/2)$ and density $\rho = 0.8$.* **2a.** *Initially, the vehicles are equally spaced on half of lane 1; the other lane (lane 2) is empty.* **2b.** *As time passes, some vehicles switch to the outer lane. The tendency for vehicles to change lanes is determined randomly, and probability is proportional to the gain in velocity that a vehicle obtains by changing lanes. The velocity function $v(\rho) = 1-\rho$ is the same in both lanes. Vehicles begin moving to lane 2 and traffic becomes denser there. The distance between vehicles in lane 1 starts to increase, which makes the vehicles' velocities increase. The image depicts vehicle positions at $t = 1$.* **2c.** *The density for the two lanes at $t = 1$ again displays an N-wave, as in Figure 1. See the online version of this article for an associated animation. Figure courtesy of the authors.*

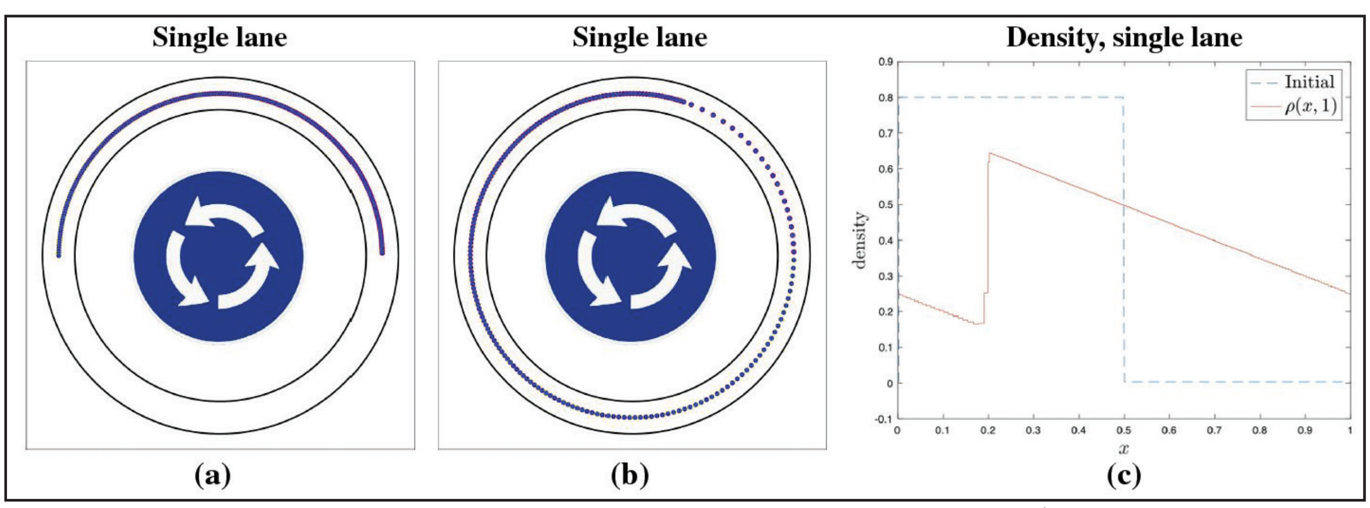

Figure 1. *A simulation of traffic on a periodic road. The number of vehicles is 200, their length is $\ell = 1/500$, and the velocity function is $v(\rho) = 1-\rho$.* **1a.** *Initially, the vehicles are equally spaced on half of the road so that the density $\ell/(x_{i+1}-x_i) = 0.8$.* **1b.** *As time begins to pass, the first vehicle instantly approaches the maximum velocity. Vehicles at the tail move more slowly and are eventually approached by those that have completed one loop (at $t = 1$).* **1c.** *Vehicle density approaches the familiar N-wave from conservation laws, wherein a shock is sandwiched between two rarefaction waves. This graph is virtually indistinguishable from the solution of the corresponding Lighthill-Whitham-Richards (LWR) model. See the online version of this article for an associated animation. Figure courtesy of the authors.*

National Math Festival

Continued from page 2

mathematician entails and always had this mental image of a man at a chalkboard in a university office," he said. "But the two mathematicians I worked with—Genetha Gray and Aaron Luttman—have led such interesting careers that have taken them places I didn't know you could go. I think it's great to expose kids to the varied ways a mathematician can work so they can better connect themselves to the math that they learn." The SIAM Education Committee expresses its deep gratitude to all participating speakers and thanks them for volunteering their time to engage with a curious and motivated group of festival attendees.

In addition to these sessions, SIAM also sponsored a booth that was staffed by both undergraduate and graduate students in applied mathematics. The booth aimed to humanize mathematics by providing attendees with the opportunity to ask students about their experiences with pursuing applied math degrees. Many festival-goers that dropped by were parents or educators who were interested in engaging their students with real-world math, although undergraduates, retired educators, and practicing mathematicians also attended. Approximately eight to 10 people—including several SIAM volunteers and one Hopin staff member who offered technical and logistical support—were at the SIAM booth at any given time throughout the weekend. In addition, the booth provided an opportunity for SIAM to share the following resources that highlight the stories of mathematicians from underrepresented groups:

– Mathematically Gifted and Black
– Latinx and Hispanics in the Mathematical Sciences (Lathisms)[9]
– The Society for Advancement of Chicanos/Hispanics & Native Americans in Science (SACNAS)[10]
– Indigenous Mathematicians[11]
– Meet a Mathematician,[12] which is an independent collection of short interviews.

The SIAM volunteers represented student chapters from nine institutions across the U.S.: Colorado School of Mines (Ethan Lewellin and Brendan Stewart); Virginia Tech (Jovan Zigic); Worcester Polytechnic Institute (Kayla Fabry); Indiana University of Pennsylvania (Aziegbemi Okisamen); University of Delaware (Nicholas Russell); Western Kentucky University (Ahmet Kaan Aydin); Arizona State University (Gabriela Navas-Zuloaga); California State University, Fresno (Mario Banuelos); and the University of North Carolina at Chapel Hill (K. Medlin, Franklin Rea, Dylan Bruney, Andrew Ford, Austin Ferguson, Kate Daftari, Gargi Dixit, Mary Mac Cole, and Nick Tapp-Hughes). SIAM's presence at the NMF would not have been possible without the time and hard work of all of these volunteers, and the SIAM Education Committee appreciates every one of them.

As this was the first-ever virtual iteration of the NMF, many presenters and volunteers were unsure of what to expect. By default, volunteers could see the number of attendees who were watching a booth or session at any given time, but only a few volunteers actually turned on their cameras. Moreover, the use of Hopin required that all participants learn a new virtual meeting environment. Nevertheless, the "Meet a Mathematician" sessions and the SIAM booth experienced a healthy level of engagement, and attendees actively asked questions and interacted with the speakers and volunteers. Linder Global Events, the NMF's logistical organizer, offered plenty of training sessions for Hopin, and all of the presenters and volunteers were fully prepared at the start of each event.

The NMF continues to play an important role in exposing the broader community to the joy of mathematics. "SIAM puts boots on the ground at the NMF because we want to see the joy, beauty, and practice of mathematics shared with as many children and families as possible," SIAM's executive director Suzanne Weekes said. From film screenings and a mathematical rap workshop to career sessions with applied mathematicians, the 2021 festival had something for everyone.

[9] https://www.lathisms.org
[10] https://www.sacnas.org
[11] http://indigenousmathematicians.com
[12] https://www.meetamathematician.com

Wesley Hamilton recently graduated with a Ph.D. in mathematics from the University of North Carolina at Chapel Hill and will start his position as a Wylie Assistant Professor at the University of Utah this summer. He also serves on the SIAM Education Committee.

The SIAM Industry Committee: What We Do and How You Can Engage

By John Abbott, Sharon Arroyo, Kevin Bongiovanni, Amr El-Bakry, and Lalitha Venkataramanan

Where can you learn about the latest mathematical advances that inspire the development of new algorithms to reduce business costs, improve product designs, and increase safety? How can you meet colleagues who will introduce you to novel avenues of research? Or perhaps you want to influence the next generation of industrial mathematicians by exposing students to industry and government careers that stimulate their curiosity and creativity while simultaneously providing value to companies and laboratories. The SIAM Industry Committee[1] is here to enable all of these goals.

The SIAM Industry Committee was formed in 2010 and has 11 members who serve three-year terms, with the possibility of up to three terms of total service. In 2020, we added 15 ad-hoc members to enhance committee diversity and extend our scope. The committee pursues the following three focus areas:

– Increasing industry member participation within SIAM

– Improving SIAM services for industry members

– Providing education support related to careers in industry.

Each committee member focuses on one or more of these areas and attends monthly meetings that allow them to share ideas and collectively chart committee progress.

In 2018, the Industry Committee sent a survey to nearly 5,000 SIAM members who work in non-academic settings (e.g., industry/business, government laboratories, nonprofit organizations, self-employed, and other). We received more than 600 replies that expressed respondents' needs from SIAM. The subsequent survey highlights helped drive recent committee activity:

– The majority of respondents indicated a desired focus in data science, machine learning, and computer science.

– Members rated SIAM journals and *SIAM Review* as the most valuable SIAM products, followed closely by conferences and workshops.

– Survey responses indicated high levels of enthusiasm for resources that identify candidates for job openings, new books in specific areas, and newly formed joint conferences with other societies.

– Young members expressed an interest in mentoring, guest speakers, training, and job listings.

Given these findings and other objectives, the Industry Committee maintains several routine initiatives. For example, every year at the SIAM Annual Meeting, a member of the Industry Committee hosts a career panel of four to six individuals from a variety of industries and laboratories with a mix of experience levels. Participants provide overviews of their backgrounds and answer questions about their career trajectories. Each panel has an associated theme; the theme at the 2020 SIAM Annual Meeting (AN20) was entrepreneurship.[2] A well-attended virtual mixer after the session encouraged further interaction with the panelists. Inspired by the success of these panels, the Industry Committee hosted another virtual event[3] about industrial and entrepreneurial careers in October 2020 that attracted over 200 participants. We are planning additional panels to serve student needs in 2021.

Here we outline several other recent accomplishments:

– The Industry Committee developed and hosted a tutorial session at AN20 on emerging research areas, in coordination with the SIAM Education Committee.

– Student chapters in Massachusetts, Washington, Oregon, and Texas connected with local industry mathematicians, allowing chapter members to pose questions about industry and business.

– Members of the Industry Committee continue to participate in SIAM's Visiting Lecture Program,[4] which promotes topics that are of interest to developing professional mathematicians.

– The committee updated information on SIAM's website that concerns industry, with a focus on industry careers.

– The Industry Committee consistently provides input to the Committee on Programs and Conferences[5] and the Career Opportunities Committee,[6] and shares industry perspectives during committee discussions.

In addition, the committee closely partners with the BIG Math Network,[7] with a particular focus on the Tondeur Initiatives.[8] In 2018, Philippe and Claire-Lise Tondeur made donations to SIAM, the American Mathematical Society, and the Mathematical Association of America to support U.S.-based mathematicians who are interested in business, industry, and government (BIG) careers. SIAM is proud to work with the BIG Math Network and its fellow mathematics societies to implement the Tondeurs' vision and facilitate career pathways for students in the mathematical sciences.

In October 2020, the Industry Committee hosted a nine-hour workshop over two days with more than 30 participants from the committee, SIAM staff, SIAM leadership, academia, and industry. The workshop aimed to develop essential elements of a three- to five-year strategic plan for non-academic membership, accounting for the 2018 survey and other recent activity.

See **Industry Committee** *on page 7*

[1] https://www.siam.org/about-siam/committees/industry-committee

[2] https://sinews.siam.org/Details-Page/an20-panel-offers-guidance-to-future-applied-mathematics-entrepreneurs

[3] https://sinews.siam.org/Details-Page/siam-panel-addresses-applied-mathematics-in-industry-and-entrepreneurship

[4] https://www.siam.org/students-education/programs-initiatives/siam-visiting-lecturer-program

[5] https://www.siam.org/about-siam/committees/committee-on-programs-and-conferences

[6] https://www.siam.org/about-siam/committees/career-opportunities-committee

[7] https://www.siam.org/students-education/programs-initiatives/big-math-network

[8] https://www.siam.org/students-education/programs-initiatives/tondeur-initiatives

Traffic Flow

Continued from page 5

thereby positions) of their vehicles if they were to move to the other lane. As such, $\tilde{z}_i$ signifies the dynamics of vehicle z_i if it were in the other lane. The probability that the vehicle changes lanes within a small time interval is then given by

$$\phi\left(\tilde{z}_i(t+\Delta t) - z_i(t+\Delta t)\right),$$

where ϕ is an increasing smooth function with $\phi(s)=0$ for $s \le 0$ and $\phi(\infty)=1$. In this model, drivers behave rather selfishly and do not consider the consequences of their lane changing for other drivers (in particular, for vehicle y_{j-1}).

If we again take the formal limit of reducing the time interval and $\ell \to 0$,

$$\frac{\ell}{z_{i+1} - z_i} \to \rho_1(x,t),$$
$$\frac{\ell}{y_{j+1} - y_j} \to \rho_2(x,t),$$

just as in the single lane case. Here, ρ_1 is the density of vehicles in lane z and ρ_2 is the density of vehicles in lane y. The lane-changing model leads to a flux from lane z to lane y. This flux allows for different velocity functions in the two lanes:

$$S\left(\rho_1,\rho_2\right) = K\left[\left(v_2(\rho_2) - v_1(\rho_1)\right)^{+}\rho_1 - \left(v_2(\rho_2) - v_1(\rho_1)\right)^{-}\rho_2\right],$$

where $a^{\pm} = (|a| \pm a)/2$ and K is a constant. Therefore, conservation of vehicles reads as

$$\begin{aligned}\frac{\partial \rho_1}{\partial t} + \frac{\partial}{\partial x}\left(\rho_1 v_1(\rho_1)\right) &= -S(\rho_1,\rho_2)\\ \frac{\partial \rho_2}{\partial t} + \frac{\partial}{\partial x}\left(\rho_2 v_2(\rho_2)\right) &= S(\rho_1,\rho_2).\end{aligned} \tag{4}$$

This is a weakly coupled system of conservation laws. Its special structure allows for the sharp estimate of the difference between two solutions [9], stating that the sum over all lanes of the L^1 norm of the difference between two solutions does not exceed the initial difference, again measured in the L^1 norm. In contrast to the single lane case, the scaling limits that lead to (4) are not rigorously established (see Figure 2, on page 5).

Figure 3 compares the numerical solution of (4) to the same initial data and reveals some similarities to the data in Figure 2.

We can expand this analysis to arbitrarily many lanes. The density ρ_i of vehicles in lane i with velocity function v_i satisfies

$$\partial_t \rho_i + \partial_x\left(\rho_i v_i(\rho_i)\right) = S_{i-1}(\rho_{i-1},\rho_i) - S_i(\rho_i,\rho_{i+1}), \qquad i = 1,\ldots,N, \tag{5}$$

with $S_0 = S_N = 0$.

It is tempting to mathematically scale the lane "width" to allow for infinitely many lanes. Even if the connection to traffic flow is absent, doing so gives rise to an interesting non-heterogeneous diffusion model with Neumann boundary conditions [1, 9]. One can also extend the LWR model to a network of roads [4, 5].

Modeling traffic flow provides a treasure chest for interesting mathematical problems. Much of the work assumes that the velocity is a decreasing function of the density, but the experimental data in Figure 4 indicates that this is not always the case.

It is also wise to recall the advice of Sherlock Holmes, courtesy of Arthur Conan Doyle: "It is a capital mistake to theorize before one has data," he says. "Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

Acknowledgments: This work is supported in part by the Waves and Nonlinear Phenomena (WaNP) grant from the Research Council of Norway.

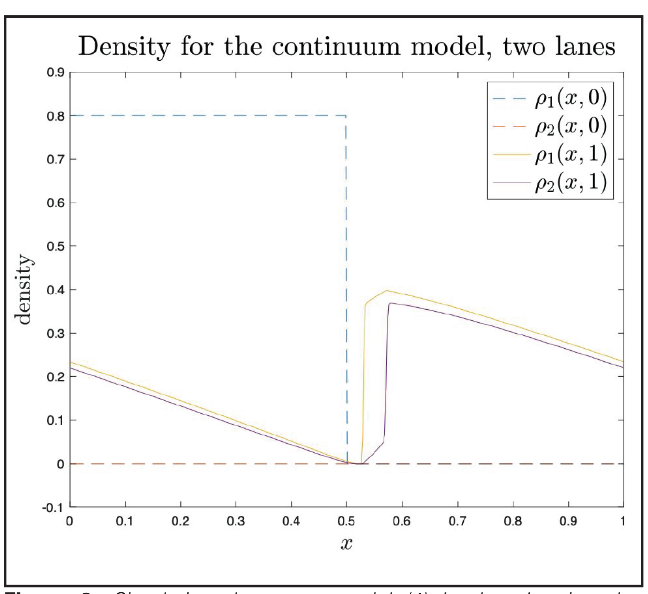

Figure 3. *Simulation that uses model (4) in the situation that corresponds to that in Figure 2. Here we utilize $\rho_1(x,0)=0.8$ for $x \le 1/2$—and zero otherwise—and $\rho_1(x,0)=0$. Furthermore, $v_1 = v_2 = 1-\rho$ and $K=1$. Figure courtesy of the authors.*

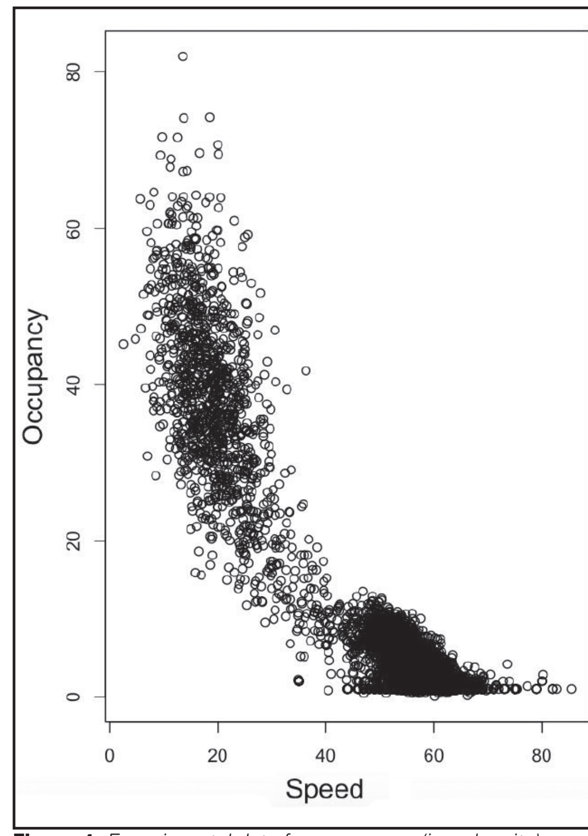

Figure 4. *Experimental data for occupancy (i.e., density) versus speed. Figure courtesy of [2].*

References

[1] Coclite, G.M., Holden, H., & Risebro, N.H. (2021). Singular diffusion with Neumann boundary conditions. *Nonlin.*, *34*(3), 1633.

[2] Delle Monache, M.L., Chi, K., Chen, Y., Goatin, P., Han, K., Qiu, J.-M., & Piccoli, B. (2021). A three-phase fundamental diagram from three-dimensional traffic data. *Axioms*, *10*(1), 17.

[3] Di Francesco, M., & Rosini, M.D. (2015). Rigorous derivation of nonlinear scalar conservation laws from follow-the-leader type models via many particle limit. *Arch. Ration. Mech. Anal.*, *217*(3), 831-871.

[4] Garavello, M., Han, K., & Piccoli, B. (2016). *Models for vehicular traffic on networks*. In *AIMS series on applied mathematics* (Vol. 9). Springfield, MO: American Institute of Mathematical Sciences.

[5] Holden, H., & Risebro, N.H. (1995). A mathematical model of traffic flow on a network of unidirectional roads. *SIAM J. Math. Anal.*, *26*(4), 999-1017.

[6] Holden, H., & Risebro, N.H. (2015). *Front tracking for hyperbolic conservation laws* (2nd ed.). In *Applied mathematical sciences* (Vol. 152). New York, NY: Springer.

[7] Holden, H., & Risebro, N.H. (2018). The continuum limit of follow-the-leader models — a short proof. *Discrete Contin. Dyn. Syst.*, *38*(2), 715-722.

[8] Holden, H., & Risebro, N.H. (2018). Follow-the-leader models can be viewed as a numerical approximation to the Lighthill-Whitham-Richards model for traffic flow. *Netw. Heterog. Media*, *13*(3), 409-421.

[9] Holden, H., & Risebro, N.H. (2019). Models for dense multilane vehicular traffic. *SIAM J. Math. Anal.*, *51*(5), 3694-3713.

[10] Knoop, V.L (2018). *Introduction to traffic flow theory* (2nd ed.). The Netherlands: Delft University of Technology.

[11] Lighthill, M.J., & Whitham, G.B. (1955). On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proc. Roy. Soc. A*, *229*(1178), 317-345.

[12] Richards, P.I. (1956). Shockwaves on the highway. *Oper. Res.*, *4*(1), 42-51.

Helge Holden is a professor of mathematics at the Norwegian University of Science and Technology. Nils Henrik Risebro is a professor of mathematics at the University of Oslo.

Industry Committee

Continued from page 6

As a result of the workshop's outcomes, SIAM and the Industry Committee are currently working on the following projects:

– Teaming up with the SIAM Major Awards Committee[9] to develop a SIAM Prize for Industry, which will recognize significant accomplishments in the use of mathematics to provide value in industry and business settings

– Planning to advertise SIAM's benefits to industry members so they utilize SIAM resources more fully

– Developing a community lecture for future Annual Meetings to offer insights into exciting industrial careers for the broader community.

In the future, SIAM and the Industry Committee plan to (i) increase industry participation in other SIAM committees to ensure the inclusion of diverse views, and (ii) expand grassroots efforts to connect industry members with local student chapters.

We will also explore a number of other longer-term initiatives. Such initiatives include expanding SIAM's reach and programs for data scientists with bachelor's and master's degrees, creating an Industry Activity Community, reviewing the Fellows Program[10] and mid-career support, developing an enhanced marketing strategy, and evaluating the corporate benefits model.

The SIAM Industry Committee remains active and maintains a strong vision for servicing our community. We welcome your insights and feedback — together we will meet these aforementioned challenges and strengthen SIAM's connection with industry. Please submit your name to the SIAM Leadership Suggestion Form[11] if you are interested in getting involved.

John Abbott is a development fellow in the Modeling, Software, and Analytics group at Corning Incorporated. Sharon Arroyo is a Technical Fellow at The Boeing Company and the Vice President for Industry at SIAM. Kevin Bongiovanni is a systems engineer at Raytheon Technologies. Amr El-Bakry is a senior Computational & Data Science Advisor at ExxonMobil Upstream Integrated Solutions Inc. Lalitha Venkataramanan is a data science advisor at Schlumberger. All of the authors are members of the SIAM Industry Committee.

[9] https://www.siam.org/about-siam/committees/major-awards-committee-mac

[10] https://www.siam.org/prizes-recognition/fellows-program

[11] https://www.siam.org/forms/leadership-suggestions

Is There an Artificial Intelligence in the House?

By Matthew R. Francis

Medical care routinely involves life-or-death decisions, the allocation of expensive or rare resources, and ongoing management of real people's health. Mistakes can be costly or even deadly, and healthcare professionals—as human beings themselves—are prone to the same biases and bigotries as the general population.

For this reason, medical centers in many countries are beginning to incorporate artificial intelligence (AI) into their practices. After all, computers in the abstract are not subject to the same foibles as humanity. In practice, however, medical AI perpetuates many of the same biases that are present in the system, particularly in terms of disparities in diagnosis and treatment (see Figure 1, on page 9).

"Everyone knows that biased data can lead to biased output," Ravi Parikh, an oncologist at the University of Pennsylvania, said. "The issue in healthcare is that the decision points are such high stakes. When you talk about AI, you're talking about how to deploy resources that could reduce morbidity, keep patients out of the hospital, and save someone's life. That's why bias in healthcare AI is arguably one of the most important and consequential aspects of AI."

While AI as a subfield of computer science has been around for over 60 years, it has only made progress in usable algorithms over the last decade. Such usability comes in the form of machine learning, which extrapolates patterns in training data to solve various problems. This training data is one spot where problems can arise.

"Healthcare has been late to the party with AI and machine learning," Parikh said. "We've had to learn from a lot of lessons that have come up with other industries. [It] isn't that AI is inherently biased; even traditional predictive tools and other things can be biased as well. The problem is that the data used in healthcare is subject to biases that certain clinicians, health systems, or payers might perpetuate, including coverage decisions that may disproportionately affect certain minority groups."

Parikh spoke about improving medical AI at the 2021 American Association for the Advancement of Science (AAAS) Annual Meeting,[1] which took place virtually in February. He identified three specific forms of statistical bias (in addition to normal measurement error): undersampling, labeling problems, and heterogeneity of effects. Undersampling can occur because white people tend to have better access to healthcare, while systemic problems lead to the inclusion of a disproportionately small percentage of people of color in the data. As a result, the widely used Framingham Risk Score—which estimates the 10-year cardiovascular risk of an individual—is nearly 20 percent lower for Black patients than white patients with the same set of clinical characteristics.

Access to care is also a source of the labeling problem. For instance, an algorithm might conclude that a Black patient no longer needs care when they actually discontinue treatment because various reasons—such as a lack of transportation, job issues, or family obligations—prevent them from getting to the medical center. Finally, heterogeneity of effects includes high rates of false negatives for Black patients because training data can miss factors that occur more frequently in minority groups.

Although these biases are statistical, they have obvious connections to larger systemic problems that include transportation disparities, insurance issues (particularly for those without job-related benefits), a lack of hospitals in rural areas and minority neighborhoods, and so forth. AI did not create these problems, but it certainly should

See **Artificial Intelligence** *on page 9*

[1] https://www.aaas.org/events/2021-aaas-annual-meeting

Exploring COVID-19's Impact on Undergraduate and Graduate Education

By Kathleen Kavanagh, Robyn Hannigan, and Joe Skufca

The COVID-19 pandemic's impact on K-12 and higher education—including both students and their educators—should not be underestimated. But can we truly estimate it at all? The fall semester is quickly approaching, and institutions must grapple with their expectations of incoming students and determine how to best support them after more than a year of nontraditional learning. Some recent high school graduates have not been in a classroom since March 2020 and will suddenly find themselves in a class of 120 college students. Despite herculean efforts by high school teachers, students might not have mastered trigonometry or algebra and thus will not be able to apply those fundamentals to related rates problems. The same is true for returning college students, many of whom learned their *entire calculus sequence over Zoom*. Will they be able to utilize these concepts in an in-person thermodynamics course?

In addition to the potential knowledge gap, one must also consider the pandemic's mental and emotional toll. It is foolish to think that educators can teach the same way they did before March 2020, at least in the immediate future. So, what information do we have and how can we use it to plan for the coming semester?

A Data Anecdote for Undergraduate Coursework

Examining the questions of "outcomes evidence" in any detail requires a bit more pause. For context, Clarkson University shifted to fully remote learning after spring break in 2020. The fall 2020 (F20) semester was a mix of on-campus, hybrid, and completely online courses. With grade distributions only ranging from spring 2020 (S20) to F20, there is not much data with which to resolve the longitudinal impacts. However, we can still begin to scour the data we do have for anecdotal evidence.

To account for as many variables as possible, we examined a specific two-course sequence that crossed disciplines: Differential Equations (MA232) and Engineering Science (ESXXX), where MA232 is a prerequisite for ESXXX. We compared the cohort from the S19-F19 semesters (both courses taught in person) with the equivalent cohort from S20-F20 to attempt to understand the differences.

The unexpected mid-semester shift to online coursework impacted MA232 students in the second cohort (S20-F20). Though students were back on campus the following semester, social distancing limited class capacity and ESXXX was taught as a fully online course.

Figure 1 shows pre-COVID and post-COVID distributions for the aforementioned two courses, where the sample population is students who took this specific two-course sequence. We note a minor shift to the right (representing improved grades) for the MA232 comparison and a much stronger shift toward higher grades in ESXXX. Based on this observation, one could conclude that MA232 students were sufficiently well prepared for the follow-up course — despite the shift to online learning in the middle of the semester. However, we acknowledge that many confounding possibilities lurk behind this simple interpretation. *Did the online shift actually help students? Are teachers simply grading more leniently? Did educators simplify the courses to the core material, which provides the minimal requirements but perhaps does not challenge the students to think? How did assessments change?* As educators, we have many questions to ponder over the summer, especially after we receive additional results from the most recent semester.

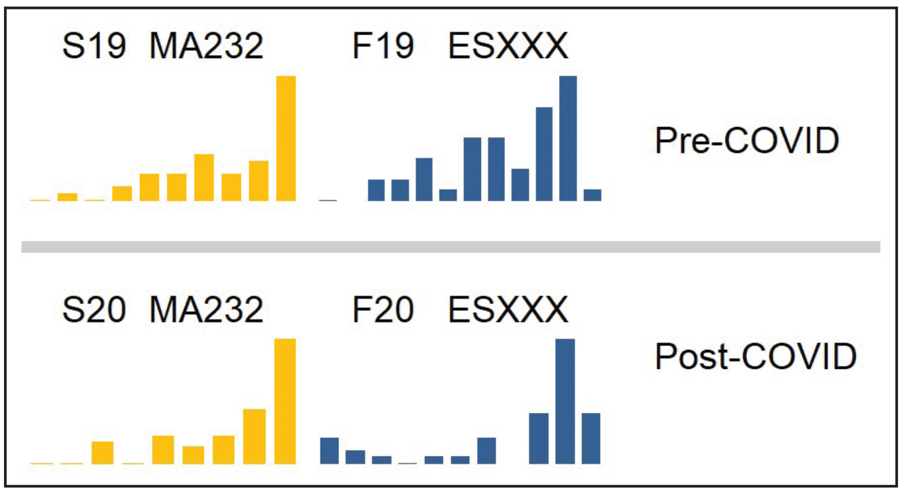

Figure 1. *Histogram that represents the grade distributions for Differential Equations (MA232) and Engineering Science (ESXXX) in "pre-COVID" (spring 2019/fall 2019) and "post-COVID" (spring 2020/fall 2020) settings. Note the apparent shift to the right in distributions for both courses in the "post-COVID" setting. Figure courtesy of Joe Skufca.*

Mental and Emotional Impacts on Graduate Students

The COVID-19 pandemic disrupted every aspect of the graduate student pathway to a Ph.D.; research, learning, and even teaching responsibilities. As graduate students, teaching assistants are a bit more prepared for online participation because they already possess independent study skills. However, student-student interactions and an overall sense of community are both important in graduate school, and social distancing has disrupted that mechanism.

Like professors, graduate students who teach had to move their pedagogy online — but perhaps with less support, especially in regards to technology. Many of them have families and were thus taking classes and teaching online courses while also homeschooling or caring for small children. Some institutions made adjustments for faculty members who were under similar levels of stress, like delaying the tenure clock. What accommodations could and should we provide for Ph.D. students? At Clarkson, the Department of Mathematics delayed all qualifying exam requirements but did not formally adjust any other expectations in terms of degree completion (though we are monitoring the situation and remaining flexible).

The social impacts of COVID-19 for students have been significant. Grad students typically operate as a cohort, and much of the information comes from that collective group. Without close communication, things can quickly derail. In a situation that we assume is not unique to Clarkson, one of our first-year graduate students appeared to be a bit withdrawn (which is harder to identify with remote interactions). One of his instructors dug a little deeper and noted that the student's cohort was small; given social distancing, it was essentially nonexistent. The student thought he needed 30 credits for the year (the stated maximum, not a minimum), believed that there was no way he could keep up, and was afraid that he would lose his assistantship. Without other students to commiserate with, he felt like he was not meeting the standard; in reality, he was doing fine — struggling in the same way everyone else was struggling. His advisor had noticed that the student seemed more stressed and had therefore reduced some research expectations, which the student then interpreted as a lack of faith in his ability to contribute to the research group. These misunderstandings would likely not have arisen if students were working face to face, with ready access to advice from more senior peers and the usual easy communication between advisors, faculty, and grad students.

As we slowly begin to emerge on the other side of COVID-19, we will need to take a personal and institutional inventory to understand any necessary adjustments. As faculty, advisors, mentors, and academic leaders, this is everyone's first time navigating the aftershock of a pandemic. SIAM provides many opportunities to network, collaborate, share, and plan as a community, and we encourage readers to organize a minisymposium on the topic of education, consider writing an article for SIAM News Online, and continue to engage in discussions with fellow SIAM members. Moving forward, we hope that applied mathematics education will be at the forefront of discussions so we can overcome the challenges of the last year. We urge members to dive into their own data and leverage the SIAM community so we can all learn from each other and determine the best path forward.

We invite readers to share their own experiences with virtual teaching or learning, offer suggestions for tips and techniques, and brainstorm with one another. Comment on the online version of this article to get the discussion started!

Kathleen Kavanagh is a professor of mathematics at Clarkson University and the Vice President for Education at SIAM. Robyn Hannigan is the provost at Clarkson. Her academic work includes publications about the opportunities of diversity in advancing student achievement and learning outcomes assessment. Joe Skufca is a professor and Chair of Mathematics at Clarkson who conducts research in dynamical systems, data science, and applied modeling. He is an active member of the SIAM Activity Group on Applied Mathematics Education with a special interest in ethics in mathematics.

Artificial Intelligence

Continued from page 8

not perpetuate them either. Such avoidance requires active intervention.

Trust and Explainability

"When we're talking about machines and algorithms, one goal should be understanding biases," Elham Tabassi, a researcher at the U.S. National Institute of Standards and Technology (NIST), said. "We need to understand what biases [are present]; have tools to quantify them; and [have] practices, standards, and documents about how to manage and mitigate them."

Because it defines standards for both government and industry, NIST is particularly concerned with the complications of AI applications. During her AAAS presentation, Tabassi addressed the need for trustworthy AI — with the unspoken implication that many current machine learning implementations are not trustworthy enough.

"Trust and risk are two sides of the same coin," she said. "What things should we be worried about? Discrimination is one, bias is one, accuracy is one. Can we build zero-bias algorithms? Maybe one day. The same thing [is true] for machines as for humans: we are not going to reach zero bias. The expectation is understanding, identifying, and managing bias."

These issues are present in many current applications—from self-driving cars to facial recognition—thus adding a sense of urgency to the need for standards and accountability. "The issue with healthcare AI is that it uniquely has the potential to mask bias and make it seem like you're generating an accurate prediction," Parikh said. "A lot of the output from AI and machine learning is a black box. We don't understand the variables that go into a prediction, [which] exists as a complex association of nonlinear relationships."

This concept—called *explainability* in computer science terms—is separate from bias, though it can have similar effects in terms of trust and risk assessment. To complicate matters, explainability varies widely from application to application. "What developers expect from an explainable AI is very different than the financial sector, based on the legal requirements needed from explainability of the algorithms," Tabassi said. "We're bringing enough knowledge and understanding about the risks…to develop a risk management framework as a tool for everybody to make the right decisions."

Do No Harm

Healthcare naturally has separate ethical standards from other fields that utilize AI, including the classic "do no harm" mantra. This sentiment may limit deployment until researchers can solve some of the problems that are related to bias and trust. Nevertheless, both Parikh and Tabassi are hopeful. "It's going to be rare even five or 10 years from now that an AI device replaces a healthcare worker, because the decisions are so high stakes," Parikh said. "AI hasn't gotten to a performance level yet where we're talking about replacing humans, and ultimately I don't think it's safe to replace clinicians. It can be reassuring that there is a human check on a potentially biased algorithm."

The reverse—a machine check on potentially biased humans—is also true in ideal circumstances. Parikh believes that assistance could come from a simple tool with widespread use in clinical settings: checklists, which help surgeons and other workers keep track of every step during a complicated procedure. "At various agencies, there are preliminary checklists around appropriate reporting for potential bias in an algorithm for publication," he said. "But we need checklists when it comes to potentially vetting an algorithm for clinical implementation as well. That type of checklist could really help clinicians filter through a lot of the noise and difficult-to-understand concepts of what it takes to declare an algorithm biased."

One such check is software that requires doctors to perform end-of-life conversations with patients, which they often neglect to do for people of color. Simple, automated prompts of this sort that remind practitioners to avoid bias might make major differences in quality of care.

AI cannot solve systemic problems in the healthcare sector on its own. However, researchers and clinicians are becoming more aware of how and where bias creeps in, rather than assuming that computers are free of such issues. Establishing standards will help achieve trust and identify the origins of biases. Even though computers cannot care the way that humans do, they can help us fix our own shortcomings in the medical field.

Figure 1. *As healthcare providers begin using artificial intelligence (AI) to help guide medical decisions, it is important that computerized medicine does not perpetuate systemic inequalities in diagnosis or treatment methods. Public domain image.*

Matthew R. Francis is a physicist, science writer, public speaker, educator, and frequent wearer of jaunty hats. His website is BowlerHatScience.org.

Errata

The May issue of *SIAM News* erroneously featured an outdated advertisement from two years ago for the Call for Nominations of the 2020 William Benter Prize in Applied Mathematics. Please disregard it and refer to the correct ad for the 2022 Prize to the right.

Strolling Through Jacobi Fields

MATHEMATICAL CURIOSITIES

By Mark Levi

What is a Jacobi Field?

A Jacobi vector field governs the separation of two nearby geodesics, to the leading order; the Jacobi equation is the linearization of the geodesic equation around a geodesic. In mechanical terms (and for embedded surfaces in $\mathbb{R}^3$), the distance s between two point masses that are sliding abreast on a surface (see Figure 1) in the absence of gravity and friction satisfies *to the leading order* the Jacobi equation

$$\ddot{s} + K(r)v^2 s = 0; \qquad (1)$$

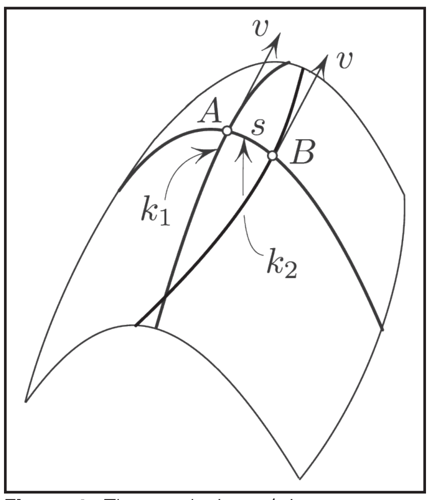

Figure 1. *The geodesic at A is tangent to a principal line of curvature of normal curvature k_1 (not drawn).*

here, $r = r(t)$ is the position of one of the masses, $v = |\dot{r}|$ is the speed, and K is the Gaussian curvature. Most books on differential geometry derive (1), but the derivations require some background — as well as some space and some time. Instead, I would like to give a back-of-the-envelope derivation of (1) *in a special case* using little more than the high school formula $F = mv^2/R$ for the centripetal force.

The Setup

Consider a unit point mass A with a velocity that points in the direction of a principal curvature at the instant in question; an identical particle B is near to and abreast of A (i.e., the arc AB is perpendicular to the velocity of A). Both A and B have the same constant speed v, and B's direction of motion is close to that of A by assumption (see Figure 1). As mentioned, there is no gravity or friction.

A Heuristic Derivation of (1)

An observer who is sliding with the reference frame of A feels the centrifugal g-force due to the curvature of A's path:

$$F = k_1 v^2. \qquad (2)$$

This inertial force—which also acts on B from A's point of view—has the tangential component $F_{\text{restoring}} = F\sin\theta$ that pulls B towards A (see Figure 2). But $\theta = k_2 s + o(s)$ by the definition of curvature (see Figure 3). Therefore,

$$F_{\text{restoring}} = Fk_2 s + o(s) \overset{(2)}{=} k_1 k_2 v^2 s + o(s).$$

And since $k_1 k_2 = K$ is the Gaussian curvature, this explains (1) — but only for the special case when the velocity points in a principal direction of curvature.

How to Explain (1) Heuristically for an Arbitrary Direction?

I would like to leave this question as a fun problem and may address it in the next installment. It turns out that the special case that I describe here misses an interesting aspect, one that is also hidden in the formal machinery of standard derivation. For example, what if the geodesic is a straight line on a ruled surface? Considering this question yields a mechanical interpretation of the Hessian determinant that has not occurred to me before.

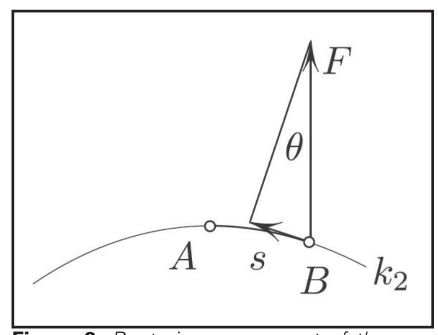

Figure 2. *Restoring component of the centrifugal force. Here, s is the arc length.*

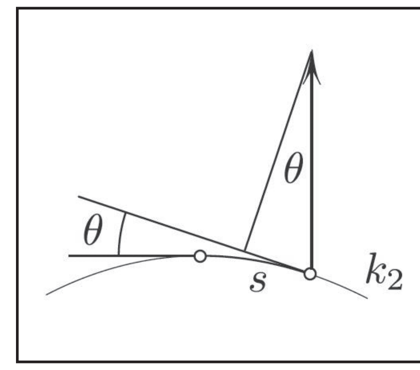

Figure 3. *Angle between the normals equals the angle between the tangents; the latter $\approx k_2 s$.*

The figures in this article were provided by the author.

Mark Levi (levi@math.psu.edu) is a professor of mathematics at the Pennsylvania State University.

An Exploration of Inliers

By Ronald K. Pearson

FROM THE SIAM BOOKSHELF

The following is a short reflection from the author of Mining Imperfect Data: With Examples in R and Python (Second Edition),[1] which was published by SIAM in 2020. This updated edition is more complete and contains techniques and applications that were not available in 2005 when the first edition—titled Mining Imperfect Data: Dealing with Contamination and Incomplete Records—initially published.

The second edition of *Mining Imperfect Data* focuses on the definitions, consequences, detection, and treatment of 10 forms of "imperfection" that commonly occur in real datasets. These imperfections include well-known data anomalies like outliers and missing observations, as well as more obscure issues like *inliers*. We can define inliers as "data values that are consistent with the distribution of the bulk of the data, but are in error" [3]. The following example [1] from *Mining Imperfect Data* illustrates the phenomenon of *disguised missing data* [4] as one possible source of inliers: "Recently, a colleague rented a car in the U.S. Since he was Dutch, his post code did not fit into the fields of the computer program. The car hire representative suggested that she use the zip code of the rental office instead."

Another possible source of inliers is extreme *coarsening*, such as when one uses the first day of a month or year as a surrogate for "exact date unknown."

Although the aforesaid definition of inliers unfortunately provides no basis for their detection, inliers in numerical data typically represent values that repeat unusually frequently. By adopting this condition as a working definition, researchers can detect inliers by tabulating the frequency of each distinct value and searching for those with atypically high frequencies. This method converts the inlier detection problem into an outlier detection problem, for which a variety of solutions exist.

The best known of these solutions is likely the "three-σ edit rule," which declares points that lie more than three standard deviations from the mean to be outliers. Despite its popularity, this approach often performs badly in practice, leading to the development of alternatives like the *Hampel identifier* or the *boxplot rule* (discussed in chapter two of *Mining Imperfect Data*, "Dealing with Univariate Outliers"). Ironically, these alternative approaches fail catastrophically in the presence of numerical data with more than 50 percent *ties* (i.e., repeated values). Since there is zero probability of ties under the continuously distributed random variable model that researchers commonly use to characterize numerical data, one would expect to see few of them in inlier-free data. This means that the majority of distinct value *frequencies* that are computed via the aforementioned inlier detection strategy should be 1. As a consequence, the three-σ edit rule probably represents the best approach for identifying unusually large frequencies in inlier detection, despite its limitations.

The Australian vehicle insurance dataset [2] provides a real-world inlier example, which is also available on the companion website[2] (see Figure 1). This dataset contains 67,856 records, with 11 fields that provide losses, claim counts, and vehicle and driver characteristics. The loss variable `claimcst0` exhibits 3,257 distinct values that range from 0 to 55,922.13; 3,182 of the values appear only once, while the most frequent value appears 63,232 times. The five most frequently occurring values and their frequencies (in parentheses) are as follows: 0 (63,232), 200 (695), 353.76999998 (219), 389.94999981 (94), and 390 (35).

Applying the three-σ edit rule to this count sequence yields a mean frequency of 20.83, a standard deviation of 1108.02, and an upper outlier detection threshold of 3344.90. The only frequency that exceeds this threshold is that for the value 0. This extremely high frequency of 0s is characteristic of variables like insurance loss data—where claims are relatively rare—or daily rainfall amounts, which are 0 for most days except in extremely wet regions.

The fact that 0 is the only value that is identified by the proposed inlier detection procedure illustrates the weakness of the three-σ edit rule. An extension of this approach that sometimes greatly improves its performance is *inward detection*. For this extension, we detect outlying counts as before, then remove these records and reapply the outlier detection procedure. By adopting this approach, we exclude the very large 0 count and recompute the mean frequency as 1.42 and the standard deviation as 12.96, thus giving an upper outlier detection limit of 40.29. The second pass of this inward detection procedure therefore identifies the second through fourth most frequent values as candidate inliers (200, 353.76999998, and 389.94999981). Of these results, the first value (200) is the most interesting because it is such a round number. Upon further investigation, it turns out to be the smallest loss value.

A more detailed discussion of this example is available in section 6.1.2 of *Mining Imperfect Data:* "Inward Detection of Inliers."

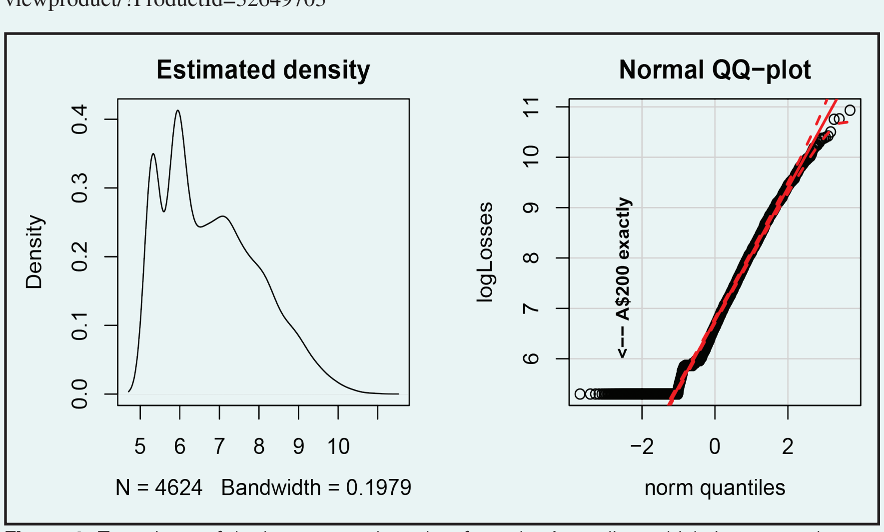

Figure 1. *Two views of the log nonzero loss data from the Australian vehicle insurance dataset.*

References

[1] Adriaans, P., & Zantinge, D. (1996). *Data mining*. Reading, MA: Addison-Wesley.

[2] de Jong, P., & Heller, G.Z. (2008). *Generalized linear models for insurance data*. Cambridge, U.K.: Cambridge University Press.

[3] DesJardins, D. (2001). Paper 169-26: Outliers, inliers, and just plain liars—new graphical EDA+ (EDA Plus) techniques for understanding data. In *Proceedings of the SAS User's Group International Conference (SUG126)*. Cary, NC: SAS Institute.

[4] Pearson, R.K. (2006). The problem of disguised missing data. *SIGKDD Explor.*, *8*(1), 83-92.

Enjoy this passage? Visit the SIAM Bookstore[3] to learn more about Mining Imperfect Data[4] and browse other available SIAM titles.

Ronald K. Pearson is a senior data scientist at GeoVera Holdings, a U.S.-based property insurance company. He has held positions in both academia and industry and has been actively involved in research and applications in several data-related fields, including industrial process control and monitoring, signal processing, bioinformatics, drug safety data analysis, property-casualty insurance, and software development.

[1] https://my.siam.org/Store/Product/viewproduct/?ProductId=32649703

[2] http://www.acst.mq.edu.au/GLMsforInsuranceData

[3] https://my.siam.org/Store

[4] https://my.siam.org/Store/Product/viewproduct/?ProductId=32649703

New Jersey High School Team Wins Top Prize for Models that Optimize High-Speed Internet Connectivity

2021 MathWorks Math Modeling Challenge Confronts the Digital Divide

By Lina Sorg

In today's increasingly digital world, many people take high-speed internet for granted. Individuals with stable internet connections often do not think twice about uploading or accessing social media content, streaming their favorite shows, or video chatting with friends and family. Yet the benefits of high-speed internet extend well beyond entertainment purposes. When many parts of the world shifted to predominately virtual schooling and employment at the onset of the COVID-19 pandemic last year, internet connectivity became more important than ever before.

Those who lack reliable access to high-speed internet are at a significant disadvantage for many daily tasks. It is far more difficult for them to attend online classes and complete assignments, work from home, utilize healthcare portals, participate in civic duties, consume news and information, and so forth. These limitations are especially salient for people in rural and low-income communities, which are often disproportionately affected by connectivity issues. And despite the numerous ways to access the internet—cable, satellites, fiber-optic lines, mobile broadband, etc.—there is no clear path to optimal connectivity; bandwidth requirements depend on region, household type, and usage frequency.

The topic of internet connectivity lent itself well to this year's MathWorks Math Modeling (M3) Challenge,[1] an annual mathematics competition that is a program of SIAM with MathWorks as its title sponsor. Now in its 16th year, M3 Challenge offers 11th and 12th graders in the U.S. and sixth form students in England and Wales the chance to compete for more than $125,000 in scholarship funds. The contest allows participating teams just 14 consecutive hours to tackle a complex, real-world problem with mathematical modeling and generate a comprehensive report that explains their solutions. All solution papers then undergo three rounds of blind judging by 150 applied mathematicians. This year, the 10 finalist teams virtually presented their solutions to a live panel of judges in late April. An online awards ceremony—in lieu of the live ceremony that is traditionally hosted by Jane Street, a quantitative trading firm in New York City—followed these presentations. Recordings of both the presentations[2] and the awards ceremony[3] are available online.

This year's problem tasked students with combatting the "digital divide": the gap between those who benefit from sufficient internet access and those who do not. "The COVID-19 pandemic really brought to light issues with internet connectivity in the U.S. and the U.K. that haven't been as clear in the past," Chris Musco of New York University said. "There are many families with insufficient access to the internet who were unable to get their students into online schooling, for example, when communities went into lockdown."

Musco—who is a 2008 M3 Challenge winning alumnus, co-author of this year's Challenge problem, and director of judging for the Technical Computing Award—elaborated on the relevance of internet connectivity as an appropriate subject for mathematical modeling and computational thinking. "We're at a really interesting time for the internet because technology is changing rapidly," he said. "The past couple of years, we've seen 5G technology roll out that is able to provide broadband, connected, wire-internet-like speeds over the air without any need for a wireless connection. This changing technology is clearly changing the conversation around this topic."

The three-part 2021 Challenge problem[4] first asked students to build a mathematical model to estimate the cost per unit of bandwidth per Megabits per second (Mbps) over the next 10 years for consumers in the U.S. and U.K. Next, they had to create a model to predict a typical household's need for internet over the course of a year, apply that model to three sample households with varying levels of internet usage, and determine the minimum amount of required bandwidth to cover their total internet needs both 90 and 99 percent of the time. Finally, teams developed an optimal plan for distributing cellular nodes and demonstrated the flexibility of their models in three hypothetical regions.

The champion team from Livingston High School in Livingston, N.J., trained their initial model on the relationships between population density, cost of living, average download speed, and average price per Mbps in the 48 mainland U.S. states. The students found that fixed infrastructure costs are higher in regions with low *population densities* because the costs are divided between fewer people; the opposite is true in high-density areas. They also noted that internet is more expensive in areas with a higher *cost of living,* and used *average download speed* as a measure of the level of infrastructure development that correlates with internet prices. "After collecting data for these three factors, we found that traditional regression techniques would be ineffective and time consuming due to the various skews and nonlinearities in the data," Leo Stepanewk of Livingston High School said. "We thus utilized random forest regression, a robust machine learning algorithm that is capable of handling and learning complex relationships on its own."

Once they achieved a root mean square error of 0.660, Stepanewk and his teammates applied the random forest regression to the entire U.S. and U.K., adjusted the population densities and costs of living based on the expected percent change in 10 years, and employed a regressed exponential function to calculate future average download speeds. Ultimately, their model predicted a decrease of $0.23 per Mbps in the U.S. and a decrease of $0.57 per Mbps in the U.K. over the next decade.

Next, the team calculated the bandwidth demands for a given household over the course of a year. "The main factors that we considered were age and occupation status, since these variables had the greatest impact on internet usage patterns," team member Edward Wang said. "The internet usage patterns for each individual consisted of a probability that they would perform a certain internet task at any given time and a range of bandwidth values for that task. Using these patterns, we created Monte Carlo simulations for each household and simulated the bandwidth demand for 1,000 trial weeks."

These simulations yielded the minimum required bandwidth of predicted demand for three example scenarios. The students determined that 14.5 and 15.5 Mbps is sufficient to respectively satisfy 90 and 99 percent of predicted demand for a couple in their early 30s with a three-year-old; 20.8 and 21.8 Mbps is adequate for a retired woman in her 70s who cares for two school-aged grandchildren twice a week; and 20.6 and 21.9 Mbps meets the demands of three former M3 Challenge participants who are sharing an off-campus apartment while completing their undergraduate degrees and working part time.

Finally, the Livingston team developed a model to optimally distribute 4G and 5G cellular nodes in arbitrary regions. To calculate the locations for node placement, the high schoolers utilized population density to compute a region's center and concluded that placing nodes at the center of mass would allow them to reach the most users.

"After learning about the intricacies and expenses of 5G, we decided that it would be worthwhile to create a model that only pinpointed areas where 5G would be beneficial," Sidhant Srivastava of Livingston High School said. "This is where demographics played a role. We created an equation to calculate the minimum average household income for a region to be eligible for 5G. We also calculated the minimum population density for that region to maximize the consumers for a 5G node." The team deduced that to qualify for a 5G network, a region should have a population density of at least 777 people per square mile and an annual household income of at least $103,689.32. By combining these two elements, the students created a criteria-based model that optimally identified regions for 5G networks and generated an efficient distribution plan for cellular nodes.

The Livingston team—which included Aditya Desai and Charles Yu in addition to Srivastava, Stepanewk, and Wang—took home $22,500 in scholarship funds for their top-notch solution. As they celebrated their impressive earnings, the students considered the ways in which this experience will impact their forays into higher education and future career trajectories. "Partaking in M3 Challenge has further piqued my interest in the intersection of mathematics, business, and technology," Desai said. "Having experienced the interrelated forces that play a role in the disciplines, I will look to consciously make these connections and apply my problem-solving approach from this problem down the road."

This is precisely the goal of M3 Challenge, which seeks to expose talented students to the complex facets of mathematical modeling that are not common in standard high school curricula. It introduces participants to relevant and timely topics and presents problems in an unfamiliar way, forcing them to think critically about real issues; quantify and organize data; and represent, analyze,

See **Internet Connectivity** *on page 12*

The Livingston High School team from Livingston, NJ, took home the top prize of $22,500 for their mathematical models of internet connectivity in the 2021 MathWorks Math Modeling (M3) Challenge. Top row, left to right: Aditya Desai, Sidhant Srivastava, and Leo Stepanewk. Bottom row, left to right: Edward Wang, Charles Yu, and coach Cheryl Coursen.

1 https://m3challenge.siam.org

2 https://m3challenge.siam.org/archives/2021/videos

3 https://go.siam.org/BIKZ70

4 https://m3challenge.siam.org/archives/2021/problem

Professional Opportunities and Announcements

Send copy for classified advertisements and announcements to ***marketing@siam.org****. For rates, deadlines, and ad specifications, visit* ***www.siam.org/advertising****.*

*Students (and others) in search of information about careers in the mathematical sciences can click on "Careers" at the SIAM website (****www.siam.org****) or proceed directly to* ***www.siam.org/careers****.*

National Institute of Standards and Technology

National Research Council

Postdoctoral Research Associates

The Applied and Computational Mathematics Division (ACMD) of the National Institute of Standards and Technology (NIST) invites applications for two-year National Research Council (NRC) postdoctoral research positions at NIST laboratories in Gaithersburg, Maryland and Boulder, Colorado. NIST is a Federal government research laboratory specializing in measurement science. ACMD consists of some 46 full-time professional staff, along with part-time faculty appointees and guest researchers. Staff members engage in collaborative research with scientists throughout NIST, providing expertise in applied mathematics, mathematical modeling, and computational science and engineering.

Research areas of interest include complex systems and networks, computational materials science, computational fluid dynamics, computational electromagnetics, computational biology, orthogonal polynomials and special functions, applied optimization and simulation, combinatorial software testing, data mining and visualization, parallel and distributed algorithms, quantum information science, and uncertainty quantification in scientific computing.

Candidates and their research proposals are evaluated in a competitive process managed by the NRC Associateship Programs. The current stipend is $72,750 per year; there is also a $5,500 travel and equipment allowance. For further details, see **https://www.nist.gov/itl/math/postdoctoral-opportunities**. Application deadlines are **August 1** and **February 1**. Appointments commence within one year of selection. For questions, contact Tim Burns at **burns@nist.gov**.

NIST is an equal opportunity employer. The NRC Associateship Program at NIST is restricted to U.S. citizens.

Getting to Know the Mexico Section of SIAM

By Gerardo Hernández-Dueñas, Jorge X. Velasco-Hernández, Irma García-Calvillo, and Daniel Olmos

The Mexico Section of SIAM (MexSIAM),[1] which was initially formed in 2000 and revived in 2018, serves as an international forum for applied mathematics in Mexico. It provides a broad and inclusive channel for the exchange of information and ideas between professionals in the mathematical sciences to promote research in mathematics and further its application in a wide variety of settings. The response of the mathematical sciences community in Mexico—which includes individuals in scientific programs that incorporate methods and techniques from engineering, mathematics, statistics, and computer science—has been outstanding. This enthusiasm reflects the growing importance and vitality of applied math in both academia and industry.

A significant number of applied mathematical scientists in Mexico work in the petroleum, mining, and energy sectors (e.g., Pemex, the Federal Electricity Commission, and the Secretariat of Energy); financial institutions; the aeronautical industry; environment and ecology fields (e.g., the National Commission for the Preservation of Biodiversity); medicine and healthcare; technological design; and engineering. Students also represent an important part of the applied mathematics community, and exposure to mathematical sciences applications early in one's career is crucial.

As with most countries, the COVID-19 pandemic has put an increased burden on Mexico's educational system. Economic inequity and the concomitant lack of technological infrastructure in many parts of the country has made it difficult for schools and institutions to offer remote online classes and other educational activities in certain communities, cities, and regions. The communication channel and support that MexSIAM provides—although circumscribed to the mathematical sciences—is therefore exceedingly important during this time. Our students have responded enthusiastically to MexSIAM activities, thus inspiring the creation of the first SIAM student chapter in the country. The student chapter at the National Autonomous University of Mexico (UNAM) will begin to host events later this year, with Raul Esquivel Sirvent and Gerardo Hernández-Dueñas as its faculty advisors.

Mexico is home to several internationally recognized academic research centers and university departments that advance and teach the mathematical sciences. Addressing the country's most pressing issues—including national innovation and technological development, sustainable natural resource use, and public health—requires the multi- and interdisciplinary engagement of mathematics. Mexico is very heterogeneous, and social inequity remains a major unsolved problem. Scientific infrastructure is predominantly located in major urban centers, with a significant presence in Mexico City. MexSIAM is fully committed to the promotion of science, technology, engineering, and mathematics (STEM) at the undergraduate and graduate levels across the various states of the Mexican federation — with a particular focus on women and underrepresented groups/sectors that are primarily located in states with high poverty indexes and low average per capita annual incomes. To strengthen this commitment, section members routinely organize summer schools that are open to students from all over the country. Members have also been involved in the development of a new applied mathematics program that will soon launch at UNAM Campus Juriquilla.

Mexico and the U.S. share a long border, an extensive history, mutual regional and bilateral interests, and an important tradition of academic exchange in many areas of STEM. The mathematical sciences have always served as a bridge between the academic establishments of both countries, and many applied mathematicians in Mexico have been SIAM members for many years. As such, MexSIAM is an important milestone in the history of our academic interrelation. We expect that MexSIAM will increase its membership in the coming years and continue to attract mathematically trained professionals who work in industry and the national productive sector. These connections will reinforce national development of the applied mathematical sciences.

The first Annual Meeting of the Mexico Section of SIAM[2] took place in December 2019 at the Ensenada Center for Scientific Research and Higher Education (CICESE) in Ensenada, Mexico. 82 registered participants delivered 64 talks and 12 poster presentations. An employee of Samsung Research Tijuana also hosted a workshop. Participants came from more than 20 universities in Mexico and the U.S. (see Figure 1).

In August 2020, UNAM and MexSIAM joined forces and organized a *Seminar Series on COVID-19* that centered on mathematical modeling approaches to control, mitigate, forecast, and understand evolutionary trends of the epidemic (see Figure 2). The emergency response generated by the worldwide spread of SARS-CoV-2 necessitates the involvement of all individuals to mitigate, contain, and control COVID-19. The seminar series focused on strategies to achieve these goals and explored ways in which applied mathematics and mathematical models can contribute. These virtual seminars—which were organized by Isaac Pérez-Castillo (Universidad Autónoma Metropolitana – Iztapalapa) and Jorge X. Velasco-Hernández (UNAM)—took place biweekly and featured speakers from Mexico, Spain, the U.S., and Canada.

This year, MexSIAM is partnering with the Mexican Mathematical Society, Mexican Society of Scientific Computing and Applications, Mexican Society for Operations Research, and Mexican Association of Statistics for a diverse set of academic activities that focus on the dissemination of applied mathematics research in Mexico. These societies are organizing a virtual applied mathematics colloquium this semester to promote interactions and the exchange of ideas between their members during the ongoing pandemic.

The 2021 Annual Meeting of the Mexico section of SIAM[3] will take place virtually from June 21 to 23. Students and academic professional members of the section who work in multiple areas of the mathematical sciences and related fields will deliver contributed talks, organize minisymposia, and present posters in their respective research sectors. This year, some of the meeting's main topics include geoscience; mathematics of planet earth; climate, ocean, and atmospheric modeling; mathematical and theoretical epidemiology (particularly in relation to the COVID-19 pandemic); mathematical and theoretical big data in public health; and ecological perspectives. There is particular interest in research that involves the use, development, and analysis of mathematical models as tools to study the dynamics of natural phenomena and their impacts on human activities.

The invited plenary speakers at the upcoming Annual Meeting are distinguished researchers who appropriately represent each of the areas that are highlighted at the meeting. They include Edgar Knobloch (University of California, Berkeley), Pablo A. Marquet (Pontifical Catholic University of Chile), Francisco J. Ocampo Torres (CICESE), Beatrice Rivière (Rice University), Mauricio Santillana (Harvard University), and Pauline van den Driessche (University of Victoria).

MexSIAM looks forward to continued growth and will remain actively involved in the larger SIAM community to promote the applied mathematical sciences.

Gerardo Hernández-Dueñas[4] is an applied mathematician who is interested in applications in the area of fluid dynamics. He is president of the Mexico Section of SIAM (MexSIAM). Jorge X. Velasco-Hernández is a biologist, mathematician, educator, enthusiastic amateur pianist, and a SIAM Fellow. He is vice president of MexSIAM. Irma García-Calvillo is a mathematician, reader, educator, and promoter of applied mathematics who enjoys traveling. She is the secretary of MexSIAM. Daniel Olmos is a biomathematician, lecturer, and active promoter of applied mathematics at every level who is always looking for all kinds of equilibrium. He is the treasurer of MexSIAM.

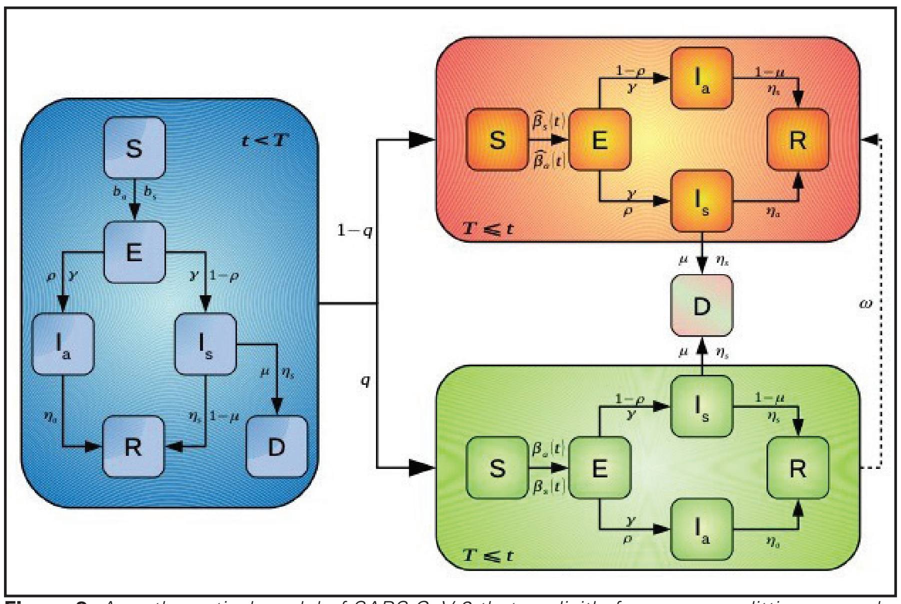

Figure 2. *A mathematical model of SARS-CoV-2 that explicitly focuses on splitting a population that is undergoing a selective lockdown. Each population on the right side of the diagram is characterized by a different time-dependent effective contact rate. Figure courtesy of Manuel Adrian Acuña-Zegarra.*

Figure 1. *Attendees of the 2019 MexSIAM Annual Meeting in Ensenada, Mexico. Photo courtesy of the local organizers at the Ensenada Center for Scientific Research and Higher Education.*

[1] https://mexsiam.org

[2] https://mexsiam.org/noticias/f/mexsiam-annual-meeting-2019-took-place-at-the-ensenada-center-for

[3] https://mexsiam.org/reunion-anual-2021

[4] https://paginas.matem.unam.mx/gerardo

Internet Connectivity

Continued from page 11

and predict trends in real-world situations. "The open-endedness of M3 Challenge, along with being able to interpret the question in a variety of ways, is so much different from the math education in school," Yu said. "In school, there's always a right answer and it's pretty easy to know if you're right or wrong. But for M3 Challenge, we had to pitch a lot of ideas and weigh their strengths and weaknesses before settling on one that made the most sense."

The Livingston team was coached by Cheryl Coursen, a mathematics instructor at the high school who teaches AP Calculus BC and Multivariable Calculus. She tries to routinely incorporate modeling into her courses and praised M3 Challenge for familiarizing students with mathematical applications. "There are no right or wrong answers in life, as it is not a neat and tidy problem situation," Coursen said. "Every choice has consequences, good and bad. Students need to see and witness this to understand and be able to face life head on, not just in the job force but in their everyday lives."

Livingston High School's paper is available online,[5] as is their final presentation.[6]

*Do you have an idea for a real-world problem that would lend itself well to mathematical modeling? A topic that is not well understood or would make a difference to society, the environment, or general quality of life? The M3 Challenge Problem Development Committee is always looking for problem ideas for future competitions and will work with authors to shape their suggestions and locate relevant data. Submit problem drafts or even just rough ideas online,[7] or send them to **M3challenge@siam.org**!*

Lina Sorg is the managing editor of SIAM News.

[5] https://m3challenge.siam.org/sites/default/files/uploads/CHAMPION_14817.pdf

[6] https://go.siam.org/02RRWo

[7] https://m3challenge.siam.org/challenge/suggest-problems