

COMMUNICATIONS

CACM.ACM.ORG OF THE **ACM** 09/2019 VOL.62 NO.09

Computational Sustainability



Computing for a Better World and a Sustainable Future

Augmented Reality Gets Real

Alloy: A Language and Tool for Exploring Software Designs

Sustaining Open Collaboration in Universities

CARLA GOMES
Cornell University
THOMAS DIETTERICH
Oregon State University
CHRISTOPHER BARRETT
Cornell University
JON CONRAD
Cornell University
BISTRA DILKINA
University of Southern California
STEFANO ERMON
Stanford University
FEI FANG
Carnegie Mellon University
ANDREW FARNSWORTH
Cornell University
ALAN FERN
Oregon State University
XIAOLI FERN
Oregon State University
DANIEL FINK
Cornell University
DOUGLAS FISHER
Vanderbilt University
ALEXANDER FLECKER
Cornell University
DANIEL FREUND
Massachusetts Institute of Technology
ANGELA FULLER
U.S. Geological Survey
JOHN GREGOIRE
California Institute of Technology
JOHN HOPCROFT
Cornell University
STEVE KELLING
Cornell University
ZICO KOLTER
Carnegie Mellon University
WARREN POWELL
Princeton University
NICOLE SINTOV
The Ohio State University
JOHN SELKER
Oregon State University
BART SELMAN
Cornell University
DANIEL SHELDON
University of Massachusetts Amherst
DAVID SHMOYS
Cornell University
MILIND TAMBE
Harvard University
WENG-KEEN WONG
Oregon State University
CHRISTOPHER WOOD
Cornell University
XIAOJIAN WU
Microsoft AI Research
YEXIANG XUE
Purdue University
AMULYA YADAV
Pennsylvania State University
ABDUL-AZIZ YAKUBU
Howard University
MARY LOU ZEEMAN
Bowdoin College

Computer and information scientists join forces with other fields to help solve societal and environmental challenges facing humanity, in pursuit of a sustainable future.

Computational Sustainability: Computing for a Better World and a Sustainable Future

THESE ARE EXCITING times for computational sciences with the digital revolution permeating a variety of areas and radically transforming business, science, and our daily lives. The Internet and the World Wide Web, GPS, satellite communications, remote sensing, and smartphones are dramatically accelerating the pace of discovery, engendering globally connected networks of people and devices. The rise of practically relevant artificial intelligence (AI) is also playing an increasing part in this revolution, fostering e-commerce, social networks, personalized medicine, IBM Watson and AlphaGo, self-driving cars, and other groundbreaking transformations.

Unfortunately, humanity is also facing tremendous challenges. Nearly a billion people still live below the international poverty line and human activities and climate change are threatening our planet and the livelihood of current and future generations. Moreover, the impact of computing and information



technology has been uneven, mainly benefiting profitable sectors, with fewer societal and environmental benefits, further exacerbating inequalities and the destruction of our planet.

Our vision is that computer scientists can and should play a key role in helping address societal and environmental challenges in pursuit of a sustainable future, while also advancing computer science as a discipline.

For over a decade, we have been deeply engaged in computational research to address societal and environmental challenges, while nurturing the new field of *Computational Sustainability*. Computational sustainability aims to identify, formalize, and provide solutions to computational problems concerning the balancing of environmental, economic, and societal needs for a sustainable future.¹⁸ Sustainability problems offer challenges but also opportunities for the advancement of the state of the art of computing and information science. While in recent years increasingly more computer

and information scientists have engaged in research efforts focused on social good and sustainability,^{12,14,16,24,29,31,35} such computational expertise is far from the critical mass required to address the formidable societal and sustainability challenges that we face today. We hope our work in computational sustainability will inspire more computational scientists to pursue initiatives of broad societal impact.

Toward a Sustainable Future

In 1987, *Our Common Future*, a United Nations report by the World Commission on Environment and Development,^a raised serious concerns about the state of our planet, the livelihood of current and future generations, and introduced the groundbreaking notion of “sustainable development.”

Sustainable development is develop-

^a United Nations. *Our Common Future*. Retrieved Aug. 25, 2018; <http://www.un-documents.net/our-common-future.pdf>

ment that meets the needs of the present without compromising the ability of future generations to meet their needs.

» key insights

- **Computer science enriches sustainability.** Computer scientists can and should make important contributions to help address key societal and environmental challenges facing humanity, in pursuit of a sustainable future. The new field of computational sustainability brings these efforts together.
- **Sustainability enriches computer science.** In turn, working on sustainability problems, which involve uncertainty, machine learning, optimization, remote sensing, and decision making, enriches computer science by generating compelling new computational challenge problems.
- **Sustainability concerns human well-being and the protection of the planet.** A large group of computer science researchers, collaborating with an even larger group of domain specialists from social, environmental, and natural sciences, can drive computational sustainability in ways that would not be possible in a smaller or less interdisciplinary setting.

The sustainable development goals (SDGs)^b identify areas of critical importance for human well-being and the protection of the planet and seek to integrate and balance the economic, social, and environment dimensions for sustainable development (see Figure 1).^c

Computational Research in Sustainability

We illustrate some of our computational sustainability research, which has focused on three general sustainability themes: Balancing environmental and socioeconomic needs; biodiversity and conservation; and, renewable and sustainable energy and materials. The research is also centered on three broad computational themes: optimization, dynamical models, and simulation; data and machine learning; and, multi-agent systems, crowdsourcing, and citizen science (noted in Figure 2). This section is organized in terms of our three sustainability themes, highlighting crosscutting computational themes, as depicted in the subway lines of Figure 3.

Balancing environmental and socioeconomic needs. The elimination of

poverty is one of the most challenging sustainable development goals. Globally, over 800 million people live below the international poverty line of \$1.90 per person per day.^d Rapid population growth, ecosystem conversion, and new threats due to conflicts and climate change are further pushing several regions into chronic poverty.

The lack of reliable data is a major obstacle to the implementation of policies concerning poverty, food security, and disaster relief. In particular, policies to eradicate poverty require the ability to identify who the poor are and where they live. Poverty mapping can be very challenging, especially in the case of the developing countries, which typically suffer from large deficiencies in terms of data quantity, quality, and analysis capabilities. For example, some countries have not collected census data in decades.^e To mitigate this challenge, Ermon and collaborators are introducing novel approaches for obtaining large-scale spatial and temporal socioeconomic indicators from publicly available satellite and remote sensing data (Figure

4). The approaches take advantage of advances in machine learning and are quite effective for estimating a variety of socio-economic indicators of poverty, even comparable to the predictive performance of expensive survey data collected in the field, and are currently being used by the World Bank.²⁰

In the arid regions of sub-Saharan Africa, one of the world’s poorest regions, migratory pastoralists manage and herd livestock as their primary occupation. During dry seasons they must migrate from their home villages to remote pastures and water points. Barrett and collaborators are developing models for studying well-being dynamics and poverty traps associated with migratory herders and other populations.⁵ The herders’ preferences are also key in the design of policies for sustainable development. Unfortunately, such preferences are often unknown to policymakers and must be inferred from data. Ermon et al.¹¹ developed generative models based on (inverse) reinforcement learning and dynamic discrete choice models, to infer the spatiotemporal preferences of migratory pastoralists, which provide key information to policymakers concerning a variety of decisions, in particular, the locations for adding new watering points for the herders.

Access to insurance is critical since uninsured losses can lead to a vicious cycle of poverty.^{5,8} Unfortunately, agricultural and disaster insurance are either unavailable or prohibitively expensive in many developing countries, due to the lack of weather data and other services. To mitigate this problem, the Trans-Africa Hydro-Meteorological Observatory (TAHMO) project is designing and deploying a network of 20,000 low-cost weather stations throughout sub-Saharan Africa.³⁶ This project gives rise to challenging stochastic optimization and learning problems for optimal weather station site selection and for uncertainty quantification in the sensors and weather predictions. For example, precipitation, one of the most important variables for agriculture, is challenging to predict due to its heavy-tailed nature and the malfunctions of rain gauges. Dieterich and his collaborators are developing models for detecting instrument malfunctions and also conditional mixture models to capture the high variance of the phenomena. There are other

b United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*. Retrieved Aug. 25, 2018; http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/Lang=E
 c United Nations. *The Sustainable Development Goals Report*. Retrieved Aug. 25, 2018; <https://bit.ly/2WbeKNB>.

d We used 2013 data since the 2015 survey data coverage is too low and considerable Asia data are suppressed; <http://iresearch.worldbank.org/PovcalNet/povDuplicateWB.aspx>
 e United Nations. *A World That Counts: Mobilizing the Data Revolution for Sustainable Development*. Retrieved June 16, 2018; <http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>

Figure 1. On Sept. 25, 2015, under the auspices of the United Nations and as part of a wider 2030 Agenda for Sustainable Development, 193 countries agreed on a set of 17 ambitious goals, referred to as the Sustainable Development Goals (SDGs), to end poverty, protect the planet, and ensure prosperity for all.



challenges in agriculture, in particular, due to market failures and information asymmetries—a consistent problem in environmental policy.^{8,23}

There are also many challenges and opportunities in connection with social interventions in the U.S., where more than 40 million people live below the U.S. poverty threshold. The U.S. also has the highest infant mortality rate and the highest youth poverty rate in the Organization for Economic Cooperation and Development, which comprises 37 high-income economies regarded as the developed countries.^f For example, Los Angeles County has over 5,000 youth between the ages of 13 and 24 sleeping on the streets or living in emergency shelters on any given day. In the context of homeless youth drop-in centers in Los Angeles, Yadav et al.⁴⁰ propose novel influence maximization algorithms for peer-led HIV prevention, illustrating how AI algorithms can significantly improve dissemination of HIV prevention information among homeless youth and have real impact on the lives of homeless youth. Tambe and Rice³⁵ provide a compilation of other examples of AI for social work concerning HIV prevention, substance abuse prevention, suicide prevention, and other social work topics.

As a final example on balancing environmental and socioeconomic issues, consider the urban landscape, which is far more congested than it was 10, 20, or 50 years ago. There is a critical need to provide individualized transportation options that have smaller carbon footprints than the automobile. One emerging alternative is bike-sharing which allows for multimodal commute round trips, with a great degree of individual flexibility, as well as economic, environmental, and health benefits. These systems have given rise to a host of challenging logistical problems, whose computationally efficient solution is required to make this new alternative sustainable. The algorithmic requirements for these problems bring together issues from discrete optimization, stochastic modeling, and behavioral eco-

f Statement on visit to the U.S. by Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights (2018). Retrieved June 16, 2018; <http://socialprotection-humanrights.org/wpcontent/uploads/2018/06/G1812530.pdf>

Figure 2. Our research is focused on three general sustainability areas depicted as the faces of the triangle.

Solutions to problems in each of these areas draw on a combination of three broad computational themes, depicted as circles. Each sustainability application creates a transformative synthesis by incorporating a combination of computational techniques from any of these themes, while each computational technique is in turn applied to a variety of problems.



Figure 3. Subway lines highlight examples of general domain crosscutting computational problem classes identified in our research projects, which correspond to subway stops.

A junction, where one project appears on multiple lines, shows how one sustainability problem can bring together multiple computational problem classes. Similarly, each subway line shows how otherwise disparate sustainability applications are related through computational problem classes.

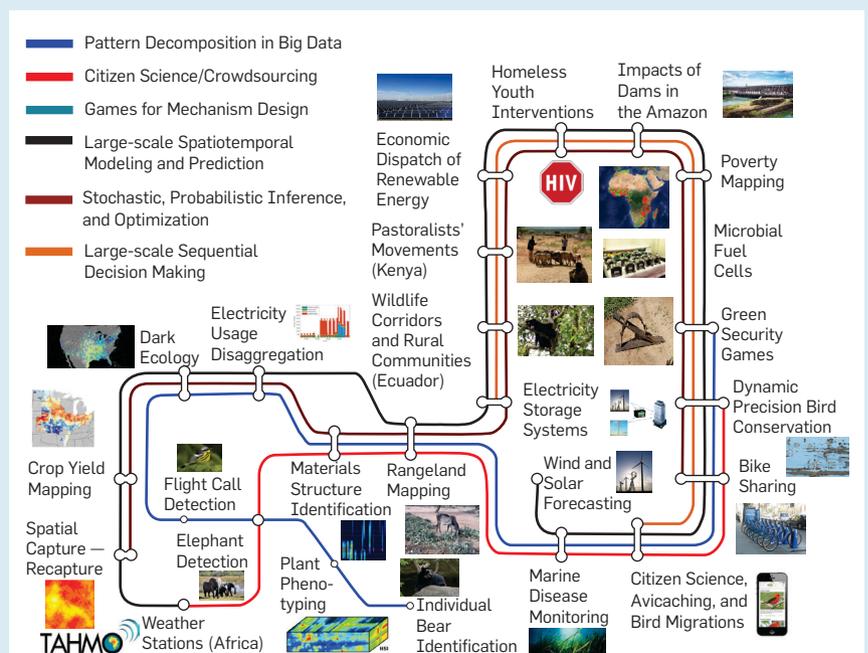


Figure 4. Transfer learning is an effective approach to model and predict socioeconomic indicators in data-scarce regions that takes advantage of satellite images that are globally available, updated frequently, and becoming increasingly more accurate.

The approach first trains a deep convolutional neural network to predict nighttime light intensity (a good proxy for economic activity) based on daytime satellite imagery. The model then estimates average household expenditures based on expenditure data from the World Bank's Living Standards Measurement Study surveys. This is done via semi-supervised learning, while enforcing spatial consistency using a Gaussian process on top of the neural network. The resulting model is surprisingly accurate, explains close to 70% of the variation in the data in some countries, and outperforms all previous methods including methods based on proprietary phone meta-data (not publicly available). This general approach has been adapted for large-scale spatial and temporal modeling and prediction of a variety of socioeconomic indicators.²⁰

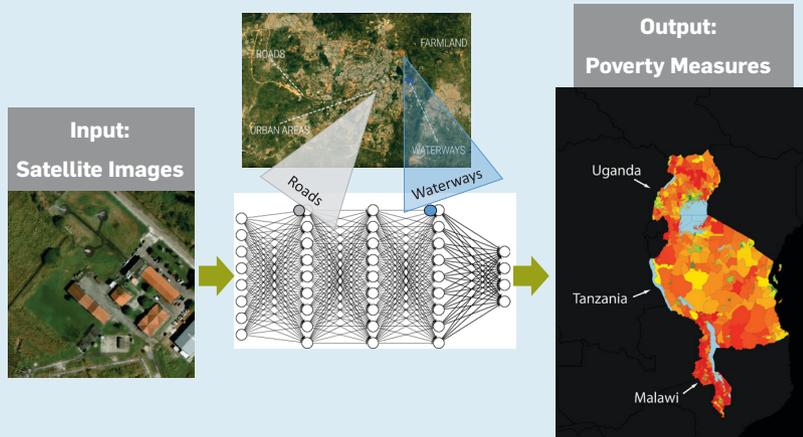


Figure 5. Games for mechanism design.

Games for mechanism design lead to challenging bi-level stochastic optimization and learning problems in which the game organizer must take into account the preferences of the agents (citizen scientists, bikers, or poachers) with respect to the organizer's actions, in order to identify the best incentive or protection strategy.



Games for bias reduction in observation collections and re-balancing activities

(a) The Avicaching game incentivizes citizen scientists to submit bird observations from undersampled areas;³⁹ Bike Angels incentivizes NYC bikers to rebalance Citi bikes.¹⁷



Games for combatting poaching and illegal activities

(b) Green security games strategically protect natural resources (forests, fish areas, and so on) against poaching and illegal activities.¹³

nomics, as well as mechanism design to appropriately incentivize desired collective behavior. One striking recent success is the crowdsourcing approach to rebalancing the shared bike fleet in NYC that contributes more to the effectiveness of Citi Bike than all motorized efforts (Figure 5); this and other computational challenges in this emerging domain are surveyed by Freund et al.¹⁷

Biodiversity and conservation. Accelerated biodiversity loss is another great challenge threatening our planet and humanity, especially considering the growing evidence of the importance of biodiversity for sustaining ecosystem services. The current rate of species extinction is estimated to be 100 to 1,000 times the background rates that were typical over Earth's history. Agriculture, urbanization, and deforestation are main causes of biodiversity reduction, leading to habitat loss and fragmentation. Climate change and introduction by humans of species to non-native ecosystems are further accelerating biodiversity loss.²⁸

A fundamental question in biodiversity research and conservation concerns understanding how different species are distributed across landscapes over time, which gives rise to challenging large-scale spatial and temporal modeling and prediction problems.^{15,25} Species distribution modeling is highly complex as we are interested in simultaneously predicting the distribution of hundreds of species, rather than a single species at a time, as traditionally done due to computational challenges. Motivated by this problem, Chen et al.⁷ developed the Deep Multivariate Probit Model (DMVP), an end-to-end learning approach for the multivariate probit model (MVP), which captures interactions of *any multi-entity process*, assuming an underlying Gaussian distribution⁷ (Figure 6), and scales considerably better than previous approaches.

Citizen science programs play a key role in conservation efforts and, in particular, in providing observational data. eBird, a citizen science program of the Cornell Lab of Ornithology, has over 450,000 members, who have gathered more than 650 million bird observations, corresponding to over 30,000,000 hours of field work.³⁴ Furthermore, to complement eBird observational data, other information sources are exploit-

ed. For example, Sheldon and collaborators' Dark Ecology project³³ extracts biological information from weather data. eBird data, combined with large volumes of environmental data and our spatiotemporal statistical and machine learning models of bird species occurrence and abundance, provide habitat preferences of the birds at a fine resolution, leading to novel approaches for bird conservation.²⁷ The results from the eBird species distribution models formed the basis for the 2011–2017 U.S. Department of Interior's State of the Birds (SOTB) reports.

The SOTB reports are generating tremendous interest from conservation organizations in using species distribution results to improve bird conservation. A good example is The Nature Conservancy's *Bird Returns* program.²⁷ The program uses reverse combinatorial auctions, in which farmers are compensated for creating habitat conditions for birds by keeping water in their rice fields for the periods that coincide with bird migrations. This novel market-based approach is only possible given the fine-grained bird habitat preference provided by the eBird-based models. Bird Returns has been tremendously successful and has created thousand of additional acres of habitat for migratory birds.

Other challenges concerning quantification and visualization of uncertainty in species prediction, multiscale data fusion and interpretation from multiple sensors, incorporation of biological and ecological constraints, and models of migration have also been addressed.^{30,32–34} Sheldon and collaborators introduced collective graphical models, which can model a variety of aggregate phenomena, even though they were originally motivated for modeling bird migrations^{6,32} (Figure 7).

Citizen science, while a valuable source of information for species distribution modeling, also poses several computational challenges with respect to imperfect detection, variable expertise in citizen scientists,²¹ and spatial and temporal sampling bias.^{34,39} The Avicaching game was developed to combat sample bias in eBird submissions (Figure 5).

To mitigate the various habitat threats encountered by species, several conservation actions are adopted. For example,

Figure 6. Multi-entity interactions.

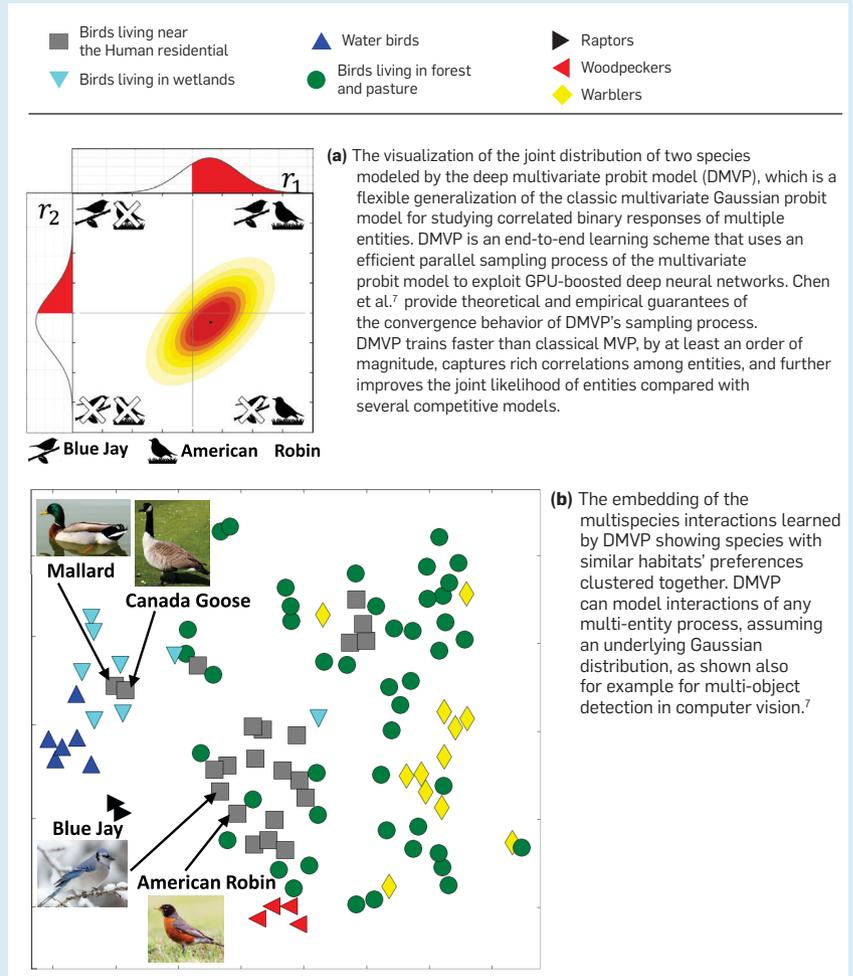


Figure 7. Collective graphical models (CGMs) are a general-purpose formalism for conducting probabilistic inference about a large population of individuals that are only observed in aggregate.

The generality of CGMs makes them suitable to model a range of aggregate phenomena, from bird migrations (the initial motivating application), to differential privacy.^{6,32} Formally, CGMs are a probabilistic model for the sufficient statistics of a graphical model, for which incomplete and noisy observations are available. Sheldon and collaborators have contributed a number of inference and learning algorithms and theoretical results about CGMs with surprising connections to the theory of belief propagation, and fast message-passing algorithms based on the Bethe entropy. The figure depicts a high-level representation of a collective graphical model. Noisy and incomplete observations \mathbf{y} (not shown) are made of the sufficient statistics through a noise model $p(\mathbf{y} | \mathbf{n})$, and the goals are to perform inference by computing the posterior distribution $p(\mathbf{n} | \mathbf{y})$ and to learn the parameters θ of the individual model.

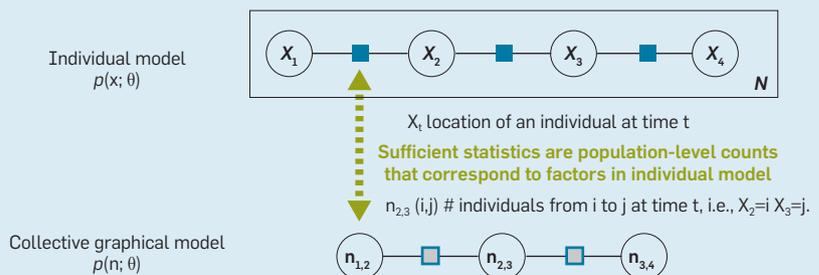


Figure 8. Robust planning of an efficient energy system to serve a load (building) from a wind farm (with variable wind speeds), the grid (with variable prices), and a battery storage device is challenging.

Energy storage provides a smooth, dispatchable flow of energy, matching energy when it is generated to loads when they arise. Motivated by his work in energy and other applications, Powell²⁶ proposes a unified modeling framework for sequential decision making, covering several distinct fields that deal with (sequential) decisions and uncertainty (dynamic programming, stochastic programming, stochastic control, simulation optimization, and bandit problems, among others) under a common umbrella. In this unified framework, there are four fundamental classes of policies consisting of policy function approximations (PFAs), cost function approximations (CFAs), policies based on value function approximations (VFAs), and look-ahead policies.



wildlife corridors have been shown to be an effective way to combat habitat fragmentation. The design of wildlife corridors, typically under tight conservation budgets, gives rise to challenging stochastic optimization problems. Current approaches to connect core conservation areas through corridors typically consider the movement of a *single* species at a time. Dilkina et al.⁹ propose new computational approaches for optimizing corridors considering benefit-cost and trade-off analysis for landscape connectivity conservation for multispecies. The results demonstrate economies of scale and complementarities conservation planners can achieve by optimizing corridor designs for financial costs and multiple species jointly. Another related work integrates spatial capture-recapture models into reserve design optimization. In yet another related effort, Fuller and collaborators are developing a program focused on Ecuador's Choco-Andean Biological Corridor, which comprises two of the world's most significant biodiversity hotspots, that integrates landscape connectivity for Andean bears and other species with economic, social and ecological information.

Prevention of wildlife crime is also important in conservation. In recent years there has been considerable AI research on devising wildlife monitoring strategies and simultaneously providing rangers with decision aids. The

approaches use AI to better understand patterns in wildlife poaching and enhance security to combat poaching (for example, see Fang et al.¹⁴). This work is leading to research advances at the intersection of computational and behavioral game theory and data-driven optimization. A notable example of this research developed so-called *green security games* (Figure 5) and has led to an application tool named Protection Assistant for Wildlife Security (PAWS),¹³ which has been tested and deployed in several countries, including Malaysia, Uganda, Botswana, and China.

Finally, we mention non-native invasive species, which invade both land and water systems and threaten ecosystems' ability to house biodiversity and provide ecosystem services. For example, the invasion of tamarisk trees in the Rio Grande valley in New Mexico has greatly reduced the amount of water available for native species and for irrigation of agricultural crops. Bio-economic models provide a basis for policy optimization and sensitivity analysis, by capturing the complex dynamics of the ecosystem, that is, the processes by which the invasive species is introduced to the landscape and spreads, as well the costs and effects of the available management actions. Unfortunately, often not much is known about these processes. Albers et al.² demonstrate the power of a stylized simulator-defined

Markov decision process approach for tamarisk, using a complex dynamical bio-economic model. A challenge is to scale up the approach and increase the realism of the bio-economic models.

Renewable and sustainable energy and materials. Renewables are being integrated into the smart grid in ever increasing amounts. Because renewables like wind and solar are non-dispatchable resources, they cannot be scheduled in advance, and alternative generation methods have to be scheduled to make up the difference. The variability and uncertainty of renewables have also raised the importance of energy storage (Figure 8). However, storage is expensive, and different storage technologies and settings are required to meet needs such as frequency regulation, energy shifting, peak shifting, and backup power. In general, controlling energy systems (generation, transmission, storage, investment) involves a number of new challenging learning and optimization problems.

For example, SMART-Invest²² is a stochastic dynamic planning model, which is capable of optimizing investment decisions in different electricity generation technologies. SMART-Invest consists of two layers. The first is an outer optimization layer that applies stochastic search to optimize investments in wind, solar, and storage. The objective function is non-convex, non-smooth, and only available via an expensive-to-evaluate black box function. The approach exploits approximate convexity to solve this optimization quickly and reliably. The second layer captures hourly variations of wind and solar over an entire year, with detailed modeling of day-ahead commitments, forecast uncertainties and ramping constraints. SMART-Invest produces a more realistic picture of an optimal mix of wind, solar, and storage than previous approaches, and therefore can provide more accurate guidance for policy makers.

In another example concerning the placement of hydropower dams in the Amazon basin (Figure 9), Wu et al.³⁸ propose new exact and approximation multi-objective optimization approaches, which are key to simultaneously consider different sustainability criteria.

In yet another example, Donti et al.¹⁰ propose task-based model learning, which was inspired by scheduling elec-

tricity generation. Task-based model learning is a general approach that combines data learning and decision making (for example, a stochastic optimization problem) in an end-to-end learning framework, specifying a loss function in terms of the decision-making objective. In this approach all components are differentiable, and therefore it is possible to learn the model parameters to improve the closed-loop performance of the overall system, which is a novel way to train machine learning models based upon the performance of decision-making systems.

Finally, we highlight new sustainable materials and processes. They provide a fundamental basis for solutions to some of the most pressing issues in energy, as well as more general issues in sustainability. In many cases, long-term solutions will depend on breakthrough innovations in materials, such as the development of new materials and processes for more efficient batteries, fuel cells, solar fuels, microbial fuel cells, or for CO₂ reduction. The high cost of conventional single-sample synthesis and analysis are driving the scientific communities to explore so-called high-throughput experimentation to accelerate the discovery process. This setup leads to computational challenges for designing and planning the experiments. Furthermore, the data analysis, integration, and interpretation process are key bottlenecks that are expert-labor intensive. Current state-of-the-art machine learning techniques are not able to produce physically meaningful solutions. Efficient computational methods are therefore urgently needed for analyzing the flood of high-throughput data to obtain scientific insights. A promising research direction is the development of generative models for unsupervised learning and for providing supervision using domain knowledge through theory-based models and simulators.

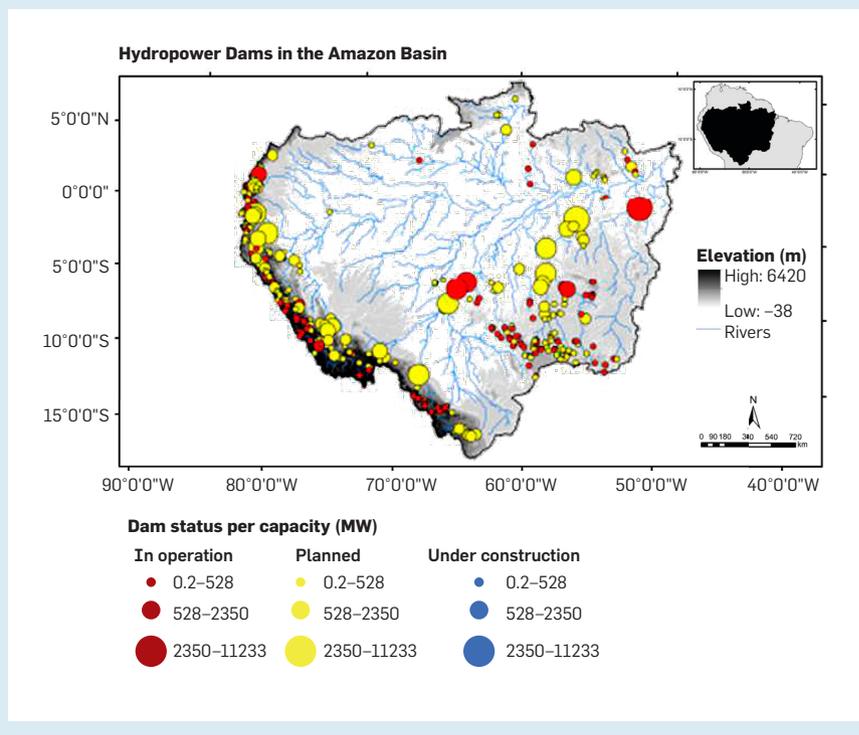
As an example, in high-throughput materials discovery, a challenging problem is the so-called phase-map identification problem, an inverse problem in which one would like to infer the crystal structures of the materials deposited onto a thin film based on the X-ray diffraction (XRD) patterns of sample points. This problem can be viewed as topic modeling or source separation with intricate physics constraints since the observed diffrac-

tion pattern of a sample point may consist of a mixture of several pure crystal patterns, and some of them may not be sampled. The task is further complicated by the inherent noise in the measurements. Human experts analyze the diffraction patterns by taking into account knowledge of the underlying physics and chemistry of materials, but it is a very labor-intensive task and often it is very challenging even for human experts. This is a good example that completely defies the current state of the art of machine learning. Phase-Mapper,⁴ is an AI platform that tightly integrates results from XRD experimentation with learning, reasoning, and human insights, to infer crystal structures from XRD data. In particular, Phase-Mapper integrates an efficient relaxed projection method for constrained non-negative matrix factorization that incorporates

physics constraints; prior knowledge based on known patterns from inorganic crystal structure databases, as well as human computation strategies. In addition Phase-Mapper uses theory-based models for incorporating prior knowledge. Since the deployment of Phase-Mapper at Caltech's Joint Center for Artificial Photosynthesis, thousands of XRD patterns have been processed, resulting in the discovery of new energy materials, such as a new family of metal oxide solar light absorbers. Gomes, Gregoire and collaborators are developing SARA (Scientific Autonomous Robotic Agent; <http://bit.ly/2M8efm9>) that encapsulates the scientific method for accelerating materials discovery substantially extending Phase Mapper. Finally, we point out a related source separation problem—hyper spectral plant phenotyping—that is

Figure 9. Multi-objective learning and optimization.

In many sustainability problems, it is critical to jointly consider multiple, often conflicting, objectives. This is the case for hydropower dam planning in the Amazon basin, with about 350 new hydropower dams proposed, which will dramatically affect a variety of Amazon ecosystem services, such as biodiversity, sediment transport, freshwater fisheries, navigation, besides energy production. The Pareto frontier captures the trade-offs between the multiple objectives with respect to the different non-dominated solutions. The non-dominated solutions also provide valuable information concerning the dams' ranking. We have developed exact dynamic-programming algorithms, fully polynomial time approximation schemes (FPTAS), and other approaches for computing the Pareto frontier for tree-structured networks, with application to the Amazon hydropower dam placement problem. For example, we can now approximate the Pareto frontier for the entire Amazon basin (~ 4M river segments), with respect to three criteria (energy; river connectivity, a good proxy for fish migrations and navigation; and sediment production) within 5% from the true optimal Pareto frontier in about 4 minutes (8 threads); or within 2% in about 1.2 hours (8 threads). The results, combined with visualization tools, help policymakers make more informed decisions concerning multiple criteria and different spatial scales.³⁸



tackled in Wahabzada³⁷ with probabilistic topic models.

Another area that can benefit dramatically from advanced AI and machine learning methods is the planning and design of scientific experiments. For example, Fern and collaborators are developing novel machine learning and constraint budgeted optimization techniques to help scientists design more efficient experiments for microbial fuels by allowing them to efficiently explore different nano-structures.³ They employ Bayesian optimization with resource constraints and production actions and have developed a new general Monte Carlo tree search algorithm with theoretical guarantees. This work also led to a large-scale empirical evaluation of Bayesian optimization algorithms, which was motivated by the confusing landscape of results in Bayesian optimization. The study involved implementing a number of top algorithms within a common framework and using cloud resources to run comparisons on a large number and variety of test functions. The main result of the study was to show the well-known Bayesian optimization heuristic—*expected improvement*—performed as well as any other approach in general and often won by significant margins. This includes beating methods such as the arguably more popular upper confidence bound (UCB) algorithm. The study found that algorithms such as UCB, which require setting a parameter for controlling exploration, are very sensitive to the parameters, making them difficult to apply widely. Expected improvement is parameter-free and appears to be quite robust. Krause and collaborators also apply Bayesian optimization for maximum power point tracking in photovoltaic power plants.¹

As a final example, Grover et al.¹⁹ model the search for the best charging policy for the Li-ion battery chemistry as a stochastic multi-arm bandit with delayed feedback. They found policies (functions for making decisions based on state variables) that considerably outperform current policies (by up to 35% in experimentation time).

Computational Synergies

We have highlighted how computational sustainability problems encompass a



Citizen science programs play a key role in conservation efforts, particularly in providing observational data. eBird, a citizen science program of the Cornell Lab of Ornithology, has over 450,000 members, who have gathered more than 650 million bird observations.



combination of distinguishing aspects that make them unique in scale, impact, complexity, and richness, posing new challenges and opportunities for computing and information science, leading to transformative research directions. One of our key goals has been to identify classes of computational problems that cut across a variety of sustainability (and other) domains. Given the universality of computational thinking, findings in one domain can be transferred to other domains. Examples of high-level cross-cutting computational problem classes, some of them depicted in Figure 3, include *spatiotemporal modeling and prediction* for bird conservation, poverty mapping, and weather mapping; *sequential decision making* for managing (renewable) resources, designing scientific experiments, managing invasive species, and pastoralism interventions; *pattern decomposition with complex constraints* for phase map identification in materials discovery, identification of elephant and bird calls from audio recordings, inferring plant phenotypes from hyper spectral data and scientific topic modeling; *active learning* (not shown in Figure 3), for scientific experimentation and sensor placement, including citizen science, and crowdsourcing, and *games for mechanism design* for providing incentives for citizen scientists, placing patrols and drones to combat poaching and illegal fishing, or incentivizing bikers to balance bike stations.

We believe that pursuing research in core or paradigmatic crosscutting computational problems is a sine qua non condition to ensure the cohesiveness and growth of computational sustainability as a field, so that researchers develop general models and algorithms with application in different sustainability and other domains. Our experience shows these *core problems naturally emerge out of real-world sustainability-driven projects, approached with the perspective of lifting solution methods to produce general methodologies, as opposed to only solving narrow problem scenarios.*

In this article, we focused on computational sustainability research examples from CompSustNet,^g a computational sustainability research network involving a large number of researchers. Unfortunately, we are not able to

^g <http://www.compsust.net/>

include many other exciting research contributions and computational challenges raised by sustainability questions, as identified in computer science, engineering, and social and natural sciences. Examples include the role of large-scale distributed systems and sensor networks, the Internet of Things, cyber-physical systems, cyber security, privacy, fairness, accountability, transparency for advanced computational systems, and also fundamental computational concepts such as reliability, modeling the hierarchical structure of socio-technical systems, and human-in-the-loop systems and intuitive, user-friendly interfaces. We also only touched on some of the 17 U.N. sustainable development goals. We point the reader to the increasing number of conferences and journals that are now starting to include tracks, workshops, and special issues focusing on tackling sustainability and societal issues, bringing together different computing and information science areas (HCI, systems, AI, and algorithms, among others).

Planning for a sustainable future encompasses complex interdisciplinary decisions for balancing environmental, economic, and societal needs, which involve significant computational challenges, requiring expertise and research efforts in computing and information science and related disciplines. Computational sustainability aims to develop new computational methodologies to help address such environmental, economic, and societal challenges. The continued dramatic advances in digital platforms, computer software and hardware, sensor networks and the Internet of Things continue to provide significant new opportunities for accelerating the pace of discovery to address societal and sustainability issues. Computational sustainability is a two-way street: it injects computational ideas, thinking, and methodologies into addressing sustainability questions but it also leads to foundational contributions to computing and information science by exposing computer scientists to new challenging problems, formalisms, and concepts from other disciplines. Just as sustainability issues intersect an ever-increasing cross-section of emerging scientific application domains, computational sustainability broadens the scope and diversity of

computing and information science while having profound societal impact.

Acknowledgments. We thank the CompSustNet members for their many contributions to computational sustainability and the support of two NSF Expeditions in Computing awards (CNS-0832782 and CCF-1522054). We thank the anonymous reviewers for their suggestions to improve the manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government. C

References

1. Abdelrahman, H., Berkenkamp, F., Poland, J., and Krause, A. Bayesian optimization for maximum power point tracking in photovoltaic power plants. In *Proceedings of the 2016 European Control Con. (Aalborg, Denmark, June 29–July 1, 2016)*, 2078–2083. <https://doi.org/10.1109/ECC.2016.7810598>
2. Albers, J.H., Dietterich, T., Hall, K., Katherine, L., and Taleghan, M. Simulator-defined Markov decision processes: A case study in managing bio-invasions. *Artificial Intelligence and Conservation* (2nd. ed.). F. Fang, M. Tambe, B. Dilkina, and A. Plumptre, (Eds.). Cambridge Univ. Press, 2018
3. Azimi, J., Fern, X., and Fern, A. Budgeted optimization with constrained experiments. *J. Artif. Intell. Res.* 56 (2016), 119–152.
4. Bai, J. et al. Phase mapper: Accelerating materials discovery with AI. *AI Mag.* 39, 1 (2018), 15–26.
5. Barrett, C., Garg, T., and McBride, L. Well-being dynamics and poverty traps. *Annual Review of Resource Economics* 8 (2016), 303–327.
6. Bernstein, G., McKenna, R., Sun, T., Sheldon, D., Hay, M., and Miklau, G. Differentially private learning of undirected graphical models using collective graphical models. In *Proceedings of the 34th International Conference on Machine Learning*, 2017, 478–487.
7. Chen, D., Xue, Y., and Gomes, C. End-to-end learning for the deep multivariate probit model. *ICML* (2018).
8. Coble, K., Mishra, A., Ferrell, S., and Griffin, T. Big data in agriculture: A challenge for the future. *Applied Economic Perspectives and Policy* 40, 1 (2018), 79–96. <https://doi.org/10.1093/aep/ppx056>
9. Dilkina, B. et al. Trade-offs and efficiencies in optimal budget-constrained multi-species corridor networks. *Conservation Biology* 31, 1 (2017), 192–202.
10. Donti, P., Kolter, J.Z., and Amos, B. Task-based end-to-end model learning in stochastic optimization. In *Proceedings of Advances in Neural Information Processing Systems Systems* (Long Beach, CA, USA, Dec. 4–9, 2017), 5490–5500.
11. Ermon, S. et al. Learning large-scale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in East Africa. In *AAAI*, 2015, 644–650.
12. Faghmous, J. and Kumar, V. A big data guide to understanding climate change: The case for theory-guided data science. *Big Data* 2, 3 (2014), 155–163.
13. Fang, F. et al. PAWS—A deployed game-theoretic application to combat poaching. *AI Magazine* 38, 1 (2017), 23–36.
14. Fang, F., Tambe, M., Dilkina, B., and Plumptre, A (eds.). *Artificial Intelligence and Conservation*. Cambridge University Press, 2018.
15. Fink, D. et al. Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications* 20, 8 (2010), 2131–2147.
16. Fisher, D.H. Recent advances in AI for computational sustainability. *IEEE Intelligent Systems* 31, 4 (2016), 56–61; <https://doi.org/10.1109/MIS.2016.61>
17. Freund, D., Henderson, S.G., and Shmoys, D.B. *Sharing Economy: Making Supply Meet Demand*. Springer, 2018.
18. Gomes, C.P. Computational sustainability: Computational methods for a sustainable environment, economy, and society. *The Bridge* 39, 4 (2009), 5–13.
19. Grover, A. et al. Best arm identification in multi-armed bandits with delayed feedback. In *Proceedings of the Inter. Conf. Artificial Intelligence and Statistics (Playa Blanca, Lanzarote, Canary Islands, Spain, April 9–11,*

- 2018), 833–842.
20. Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., and Ermon, S. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.
21. Kelling, S. et al. Can observation skills of citizen scientists be estimated using species accumulation curves? *PLoS one* 10, 10 (2015).
22. Khazaei, J. and Powell, W.B. SMART-Invest: A stochastic, dynamic planning for optimizing investments in wind, solar, and storage in the presence of fossil fuels. The case of the PJM electricity market. *Energy Systems* 9, 2 (2018), 277–303.
23. Kraus, S. Automated negotiation and decision-making in multiagent environments. *ECCAI Advanced Course on Artificial Intelligence*. Springer, 2001,150–172.
24. Lässig, J., Kersting, K., and Morik, K (Eds.). *Computational Sustainability. Studies in Computational Intelligence* 645 (2016). Springer.
25. Phillips, S.J., Anderson, R.P., and Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecological Modeling* 190, 3-4 (2006), 231–259.
26. Powell, W. A unified framework for stochastic optimization. *European J. Operational Research* 275, 3 (2019), 795–821.
27. Reynolds, M.D. et al. Dynamic conservation for migratory species. *Science Advances* 3, 8 (2017), e1700707.
28. Rockström, J. et al. Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and Society* 14, 2 (2009).
29. Rudin, C. and Wagstaff, K. Machine learning for science and society. *Machine Learning* 95, 1 (2014), 1–9; <https://doi.org/10.1007/s10994-013-5425-9>
30. Ruiz-Muñoz, J.F., You, Z., Reich, R., and Fern, X.Z. Dictionary learning for bioacoustics monitoring with applications to species classification. *Signal Processing Systems* 90, 2 (2018), 233–247.
31. Russell, S.J. et al. Letter to the editor: Research priorities for robust and beneficial artificial intelligence: An open letter. *AI Magazine* 36, 4 (2015).
32. Sheldon, D.R. and Dietterich, T.G. Collective graphical models. *Advances in Neural Information Processing Systems*, 2011, 1161–1169.
33. Sheldon, D.R. et al. Approximate Bayesian inference for reconstructing velocities of migrating birds from weather radar. In *Proceedings of the 27th AAAI Conf. Artificial Intelligence*. (Bellevue, WA, USA, July 14–18, 2013).
34. Sullivan, B.L. et al. The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation* 169 (2014), 31–40.
35. Tambe, M. and Rice, E (eds.). *Artificial Intelligence and Social Work*. Cambridge University Press, 2018.
36. Giesen, N., Hut, R., and Selker, J. The Trans-African Hydro-Meteorological Observatory (TAHMO). *Wiley Interdisciplinary Reviews: Water* 1, 4 (2014), 341–348.
37. Wahabzada, M., Mahlein, A.K., Bauchhage, C., Steiner, U., Oerke, E.C., and Kersting, K. Plant phenotyping using probabilistic topic models: Uncovering the hyperspectral language of plants. *Scientific Reports* 6 (2016).
38. Wu, X. et al. Efficiently approximating the Pareto frontier: Hydropower dam placement in the Amazon basin. In *AAAI* (2018).
39. Xue, Y., Davies, I., Fink, D., Wood, C., and Gomes, C.P. Avicaching: A two stage game for bias reduction in citizen science. In *Proceedings of the 2016 Intl. Conf. on Autonomous Agents & Multiagent Systems*.776–785.
40. Yadav, A. et al. Influence maximization in the field: The arduous journey from emerging to deployed application. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 2017, 150–158.

For questions regarding this article, contact **Carla Gomes** (gomes@cs.cornell.edu), a professor of computer science and director of the Institute for Computational Sustainability at Cornell University, Ithaca, NY, USA

Copyright held by authors/owners.



Watch the authors discuss this work in the exclusive *Communications* video. <https://caom.acm.org/videos/computational-sustainability>