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

CARLA GOMES, CORNELL UNIVERSITY; THOMAS DIETTERICH, OREGON STATE UNIVERSITY; BISTRA DILKINA, UNIVERSITY OF SOUTH 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, MASSACHUSSETS INSTITUTE OF TECHNOLOGY; ANGELA FULLER, UNITED STATES GEOLOGICAL SURVEY; JOHN GREGOIRE, CALTECH INSTITUTE OF TECHNOLOGY; JOHN HOPCROFT, CORNELL UNIVERSITY; ZICO KOLTER, CARNEGIE MELLON UNIVERSITY; WARREN POWELL, PRINCETON UNIVERSITY; NICOLE SANTOV, OHIO STATE UNIVERSITY; JOHN SELKER, OREGON STATE UNIVERSITY; BART SELMAN, CORNELL UNIVERSITY; DANIEL SHELDON, UNIVERSITY OF MASSACHUSSETS AMHERST; DAVID SHMOYS, CORNELL UNIVERSITY; MILIND TAMBE, UNIVERSITY OF SOUTH CALIFORNIA; CHRISTOPHER WOOD, CORNELL UNIVERSITY; WENG-KEEN WONG, OREGON STATE UNIVERSITY; XIAOJIAN WU, MICROSOFT; STEVE KELLING, CORNELL UNIVERSITY; YEXIANG XUE, PURDUE UNIVERSITY; AMULYA YADAV, PENNSYLVANIA STATE UNIVERSITY; AZIZ YAKUBU, HOWARD UNIVERSITY; MARY LOU ZEEMAN, BOWDOIN COLLEGE

Computational sustainability aims to develop computational methods to help solve environmental, economic, and societal problems and thereby facilitate a path towards a sustainable future. Sustainability problems are unique in scale, impact, complexity, and richness, offering challenges but also opportunities for the advancement of the state of the art of computing and information science.

Additional Key Words and Phrases: Computational Sustainability, Computing for Good, AI/Data Science for Social Good, Robust and Beneficial AI.

ACM Reference Format:
Carla Gomes, Cornell University; Thomas Dietterich, Oregon State University; Bistra Dilkina, University of South 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, Massachussets Institute of Technology; Angela Fuller, United States Geological Survey; John Gregoire, Caltech Institute of Technology; John Hopcroft, Cornell University; Zico Kolter, Carnegie Mellon University; Warren Powell, Princeton University; Nicole Santov, Ohio State University; John Selker, Oregon State university; Bart Selman, Cornell University; Daniel Sheldon, University of Massachussets Amherst; David Shmoys, Cornell University; Milind Tambe, University of South California; Christopher Wood, Cornell University; Weng-Keen Wong, Oregon State University; Xiaojian Wu, Microsoft; Steve Kelling, Cornell University; Yexiang Xue, Purdue University; Amulya Yadav, Pennsylvania State University; Aziz Yakubu, Howard University; Mary Lou Zeeman, Bowdoin College. 2018. Computational Sustainability: Computing for a Better World and a Sustainable Future. 1, 1 (August 2018), 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Author's address: Carla Gomes, Cornell University; Thomas Dietterich, Oregon State University; Bistra Dilkina, University of South 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, Massachussets Institute of Technology; Angela Fuller, United States Geological Survey; John Gregoire, Caltech Institute of Technology; John Hopcroft, Cornell University; Zico Kolter, Carnegie Mellon University; Warren Powell, Princeton University; Nicole Santov, Ohio State University; John Selker, Oregon State university; Bart Selman, Cornell University; Daniel Sheldon, University of Massachussets Amherst; David Shmoys, Cornell University; Milind Tambe, University of South California; Christopher Wood, Cornell University; Weng-Keen Wong, Oregon State University; Xiaojian Wu, Microsoft; Steve Kelling, Cornell University; Yexiang Xue, Purdue University; Amulya Yadav, Pennsylvania State University; Aziz Yakubu, Howard University; Mary Lou Zeeman, Bowdoin College.

Fig. 1. On September 25th 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 [23, 36].

INTRODUCTION

These are exciting times for computional sciences with the digital revolution permeating a variety of areas and radically transforming businesses, science, and the daily activities of the general public. The Internet and the World-Wide-Web, GPS, satellite communications, remote sensing, and smart-phones, are accelerating dramatically 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.g., e-commerce, social networks, personalized medicine, IBM Watson and AlphaGO, and self-driving cars.

Unfortunately, humanity is also facing tremendous challenges. Millions of people live in poverty and humans' activities and climate change are threating 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.

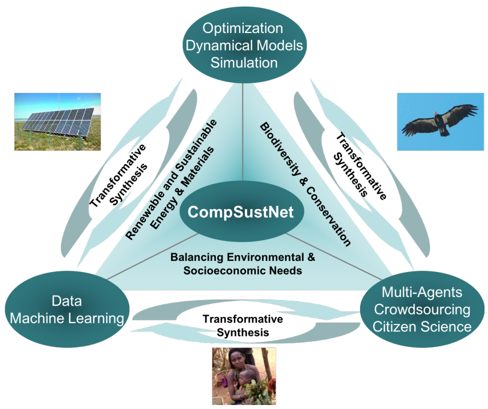

Fig. 2. Sustainability areas and computational themes of our research.

> *Our vision is that computational scientists can and should play a key role in helping address societal and environmental challenges in pursuit of a sustainable future, which will also advance computational sciences.*

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 [16]. 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 an increasing number of computing and information scientists has engaged in research efforts focused on social good and sustainability (see e.g., [10, 12, 14, 22, 28, 30, 34]), 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.

TOWARDS A SUSTAINABLE FUTURE

In 1987, *Our Common Future*, a United Nations report by the World Commission on Environment and Development, raised serious concerns about the state of our planet, the livelihood of current and future generations, and introduced the groundbreaking notion of *Sustainable Development*[40].

> Sustainable Development is development that meets the needs of the present without compromising the ability of future generations to meet their needs.

The Sustainable Development Goals (SDGs) 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 (Fig. 1)[36].

COMPUTATIONAL RESEARCH IN SUSTAINABILITY

In this section we illustrate some of our computational sustainability research, which has focused on three main general sustainability themes: (1) Balancing Environmental and Socioeconomic Needs, (2) Biodiversity and Conservation, and (3) Renewable and Sustainable Energy and Materials, centered around three main broad computational topics: (1) Decision Making, Constraint Reasoning and Optimization, Dynamical Models, and Simulation, (2) Small and Big Data and Machine Learning, and (3) Multi-Agent Systems, Crowdsourcing, and Citizen Science (Fig. 2). The section is organized in terms of our three sustainability themes, highlighting cross-cutting computational themes, as depicted in the subway lines of Fig. 3.

Balancing Environmental and Socio-economic Needs

The elimination of poverty is one of the most challenging sustainable development goals. Globally, over 750 million people live below the international poverty line. Rapid population growth, ecosystem conversion, and new threats due to conflicts and climate change are further pushing several regions into chronic poverty [37].

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 [35]. An exciting computational research direction to mitigate this challenge uses a novel transfer learning approach that takes advantage of the advances machine learning to obtain large-scale spatial and temporal socioeconomic indicators from publicly available high-resolution satellite imagery (Fig. 4). The approach, is 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 it is currently being used by the World Bank [19].

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 semi-annual dry seasons they must migrate from their home villages to remote pastures and waterpoints. Understanding the spatio-temporal resource preferences of herders is paramount in the design of policies for sustainable development. Unfortunately, such preferences are often unknown to policy-makers and must be inferred from data. Ermon et al. [9] developed generative models, based on (inverse) reinforcement learning and dynamic discrete choice models, to infer the spatio-temporal preferences of migratory pastoralists, which provide key information to policy makers concerning, e.g., locations for adding new watering points for the herders.

Access to insurance is critical for development since uninsured losses can lead to a vicious cycle of poverty. 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 [38]. This project gives rise to challenging stochastic optimization and learning problems for selecting optimal sites for the weather stations and for quantifying

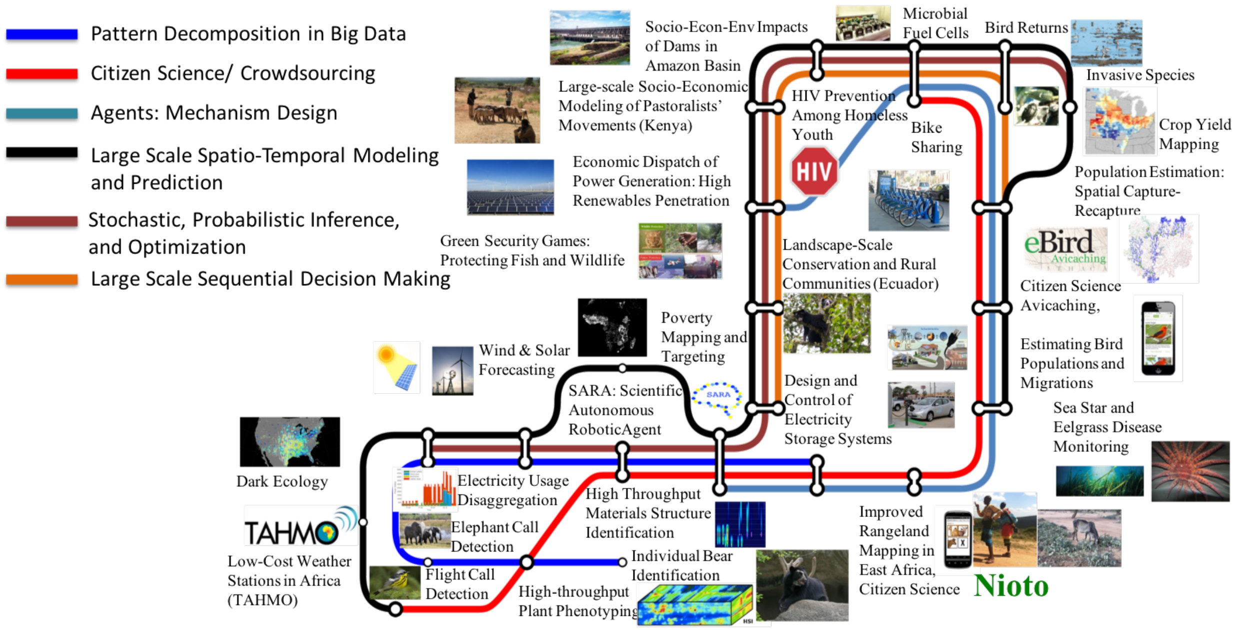

Fig. 3. Subway line highlighting cross-cutting computational themes of some of our research projects.

uncertainty in the sensors and weather predictions. For example, precipitation, the most important variable for agriculture, is challenging to predict due to its heavy-tailed nature and the malfunctions of rain gauges. Dietterich 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 also many challenges and opportunities for computing and information science researchers in connection with social interventions in the United States, where more than 40 million people live below the US poverty threshold. The US 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 [37]. 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, [43] proposes 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 Rice [34] provide a compilation of other examples of AI for social work, concerning e.g., HIV prevention, substance abuse prevention, suicide prevention and other social work topics.

As a final example on balancing environmental and socio-economic 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, even in contrast to owning a bike, allows for multi-modal commute round-trips, with a much greater 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 economics, as well as mechanism design to appropriately incentivize desired collective behavior. One striking recent success is the crowd-sourcing approach to rebalancing the shared bike fleet in NYC that contributes more to the effectiveness of Citi Bike than all motorized efforts; this and other computational challenges in this emerging domain are surveyed by Freund et al. [15].

Biodiversity and Conservation

Accelerated biodiversity loss is another great challenge threating 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 introductions of species by humans to non-native ecosystems are further accelerating biodiversity loss [27].

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 [13, 24]. Species distribution modeling is highly complex as we are interested in simultaneously predicting the distribution of hundreds of species, rather than a single species, as traditionally done. Motivated by

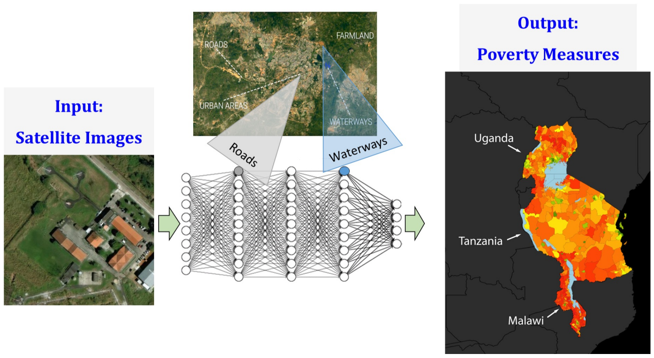

Fig. 4. Transfer learning is an effective approach to model and predict socio-economic indicators in data scarce regions that takes advantage of satellite images that are globally available, updated frequently, and becoming increasingly more accurate. The approach first trains a deep convolutional neural 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, 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 socio-economic indicators [19].

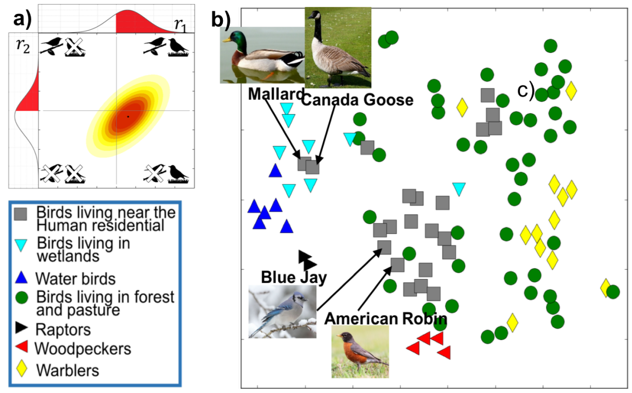

Fig. 5. Multi-entity interactions: **a)** The visualization of the joint distribution of two species modeled by the deep multivariate probit model (DMVP), which is a flexible generalization of the classic multivariate Gaussian probit model for studying correlated binary responses of multiple entities. DMVP is an end-to-end learning scheme that uses an efficient parallel sampling process of the multivariate probit model to exploit GPU-boosted deep neural networks. We have provided theoretical and empirical guarantees of the convergence behavior of DMVP's sampling process. DMVP trains faster than classical MVP, by at least an order of magnitude, captures rich correlations among entities, and further improves the joint likelihood of entities compared with several competitive models. **b)** The embedding of the multi-species interactions learned by DMVP. DMVP can model interactions of any multi entity process, assuming a multi-variate Gaussian model, as we showed also for e.g., multi-object detection in computer vision [6].

this problem, we developed the Deep Multivariate Probit Model (DMVP) [6], an end-to-end learning approach for the multivariate probit model (MVP), which captures interactions of *any multi entity process*, assuming a multi-variate Gaussian model [6] (Fig. 5).

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 400,000 members, who have gathered more than 500 million bird observations, corresponding to over 30,000,000 hours of field work [33]. Furthermore, to complement eBird observational data, other information sources are exploited. For example, the *Dark Ecology* project [32] extracts biological information from weather data. eBird data, combined with large volumes of environmental data and our spatio-temporal 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 [26]. 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 richness and success of these reports is generating tremendous interest from governmental and non-governmental conservation organizations in using species distribution results to improve bird conservation.

For example, the Nature Conservancy's *Bird Returns* program [26] uses reverse combinatorial auctions, in which farmers are compensated for creating habitat conditions for birds, e.g., 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 habit for migratory birds.

In our work we are also addressing other challenges concerning quantification and visualization of uncertainty in species prediction, multi-scale data fusion and interpretation from multiple sensors, incorporation of biological and ecological constraints, and models of migration (see e.g., [29, 31–33]). Fig. 6 depicts collective graphical models, which can model a variety of aggregate phenomena, even though they were originally motivated for modeling bird migrations. [5, 31]. Concerning citizen science, while it is a valuable source of information for species distribution modeling, it also poses several computational challenges with respect to imperfect detection, variable expertise in citizen scientists [20], and spatial and temporal sampling bias [33, 42]. Avicaching is a game that was developed to combat sample bias in eBird submissions (Fig. 7).

To mitigate the various habitat threats encountered by species, several conservation actions are adopted. For example, wildlife corridors have been shown effective as a 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 connecting core conservation areas through corridors typically consider simplistic strategies and often the movement of a single species across a network of protected areas. Dilkina et al. [7] propose new computational approaches for optimizing corridors considering benefit-cost and trade-off analysis for landscape connectivity conservation for multi-species. The

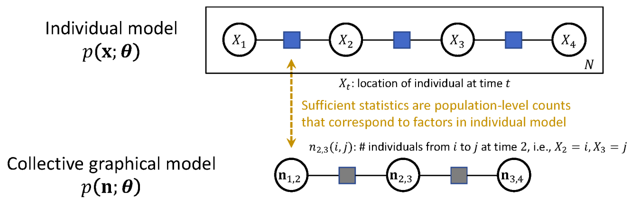

Fig. 6. Collective graphical models (CGMs) are a general-purpose formalism for conducting probabilistic inference about a large population of individuals that are only observed in aggregate. The generality of CGMs makes them suitable to model a range of aggregate phenomena, from bird migrations (the initial motivating application), to differential privacy [5, 31]. Formally, CGMs are a probabilistic model for the sufficient statistics of a graphical model, for which incomplete and noisy observations are available. We have contributed a number of inference and learning algorithms and theoretical results about CGMs with surprising and beautiful connections to the theory of belief propagation, and fast message-passing algorithms based on the Bethe entropy have been developed. The figure depicts a high-level a representation of a collective graphical model. Noisy and incomplete observations $\mathbf{y}$ (not shown) are made of the sufficient statistics through a noise model $p(\mathbf{y} \mid \mathbf{n})$, and the goals are to perform inference by computing the posterior distribution $p(\mathbf{n} \mid \mathbf{y})$ and to learn the parameters θ of the individual model.

Fig. 7. Games for mechanism design: a) The Avicaching game incentivizes citizen scientists to submit bird observations from under-sampled areas [42]; Bike Angels incentivizes NYC bikers to re-balance Citi bikes [15]. b) Green security games strategically protect natural resources (forests, fish areas, etc) against poaching and illegal activities [11]. These games lead to challenging bi-level stochastic optimization and learning problems in which the game organizer needs to take into account the preferences of teh agents (citizen scientists, bikers, or poachers) with respect to the organizer's actions, in order to identify the best incentive or protection strategy.

results demonstrate economies of scale and complementarities conservation planners can achieve by optimizing corridor designs for financial costs and for multiple species connectivity jointly. Another effort, integrates spatial capture-recapture models into reserve design optimization [18]. In a related effort, Fuller and collaborators are developing a program focused on Ecuador's Choco-Andean Biological Corridor that integrates landscape connectivity for Andean bears and other species with economic, social and ecological information. Ecuador's Choco-Andean Biological Corridor comprises two of the world's most significant biodiversity hot-spots.

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 provide rangers with decision aids. The approaches use AI to better understand patterns in wildlife poaching and enhance security to combat poaching (see e.g., [12]). 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* (Fig. 7) and has led to an application tool named Protection Assistant for Wildlife Security (PAWS) [11], which has been tested and deployed in several countries, including in Malaysia, Uganda, and Botswana.

Finally we mention non-native invasive species (IS), which invade both land and water systems and dramatically 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, i.e., the processes by which the invasive species is introduced to the landscape and it spreads, as well the costs and effects of the available management actions. Unfortunately, often not much is known about these processes, which makes it challenging to develop realistic IS bio-economic models. [2] demonstrates the power of a stylized simulator-defined MDP approach for tamarisk, using a complex dynamical bio-economic model. A key 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 (Fig. 8). However, storage is expensive, and different storage technologies are required to meet needs such as frequency regulation, energy shifting, peak shifting and backup power. Storage can also be used in a variety of settings, including grid-level storage (using central control), dedicated pairs with utility-scale wind or solar farms, and behind-the-meter applications for companies and residences. In general, controlling energy systems (generation, transmission, storage, investment) involves a number of new challenging learning and optimization problems that need to be solved in the presence of different types of uncertainty.

For example, SMART-Invest [21] 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.

Task-based model learning [8] is a general approach that combines data learning and decision making (e.g, 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 elements of the pipeline are differentiable, and therefore

Fig. 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.

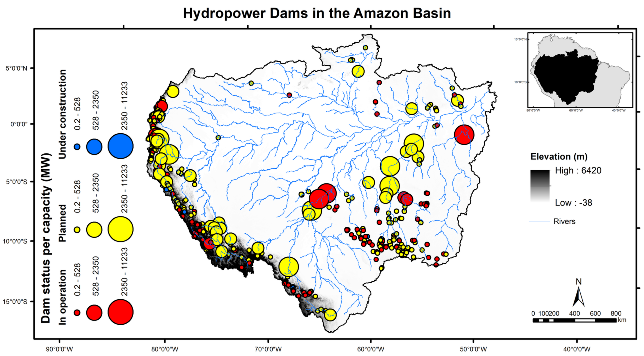

Fig. 9. Multiobjective learning and optimization: In many sustainability problems, it is critical to jointly consider multiple, often conflicting, objectives. This is the case for hydropower dam planning in the Amazon basin, with over 300 new hydropower dams proposed, which will dramatically affect a 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 mutiple 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 ($\sim 5M$ river segments), with respect to four criteria (energy, river connectivity, a good proxy for fish migrations and transportation, sediment production, and seismic risk) within 10% fron the true optimal Pareto frontier containing about 500K non-dominated solutions in about 6 hours; or within 5% containing about 2M non-dominated solutions in about 5 days. The results, combined with visualization tools, help policy makers make more informed decisions concerning multiple criteria and different planning geographic scales [41].

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. Task-based model learning was inspired by scheduling electricity generation.

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 set up 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. We are developing techniques e.g., based on 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 patterns of sample points. This problem can be viewed as topic modeling or source separation, with intricate physics constraints, since the observed diffraction pattern of a sample point may consist of a mixture of several pure crystal patterns, 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 X-Ray diffraction (XRD) experimentation, AI problem solving, machine learning, and human computation, 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 we are also developing crystallography theory-based generative models for incorporating prior knowledge. Since the deployment of Phase-Mapper at the Joint Center for Artificial Photosynthesis, at Caltech, thousands of XRD diffraction patterns have been processed, resulting in the discovery of new energy materials, such as a new family of metal oxide solar light absorbers in the previously unsolved Nb-Mn-V oxide system. We are also developing SARA (Scientific Autonomous Robotic Agent) for automating and encapsulating the scientific method for discovery of new materials for clean energy. Finally, we point out a related source separation problem, hyperspectral plant phenotyping, which is tackled in [39] with probabilistic topic models to uncover the hyperspectral language of plants.

Another area that can benefit dramatically from advanced AI and machine learning methods is the planning and design of scientific

Fig. 10. Sequential Decision Making: Motivated by his work in energy and other applications, Powell [25] proposes a unified modeling framework, covering several distinct fields that deal with (sequential) decisions and uncertainty (dynamic programming, stochastic programming, stochastic control, simulation optimization, bandit problems, etc) under a common umbrella. The modeling framework is centered on an optimization problem that involves searching over policies (functions for making decisions that depend on state variables), as opposed to the more traditional problem in deterministic optimization of searching for the best deterministic decision. 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. When applied to an energy storage problem, each of the four classes of policies might work best depending on the data.

experiments. For example, Azimi et al. [3] are developing novel machine learning and constraint budgeted optimization techniques to help scientists desgin more efficient experiments for microbial fuels by allowing them to efficiently explore different nano-structures. 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 code for the study is publicly available (https://github.com/Eiii/opt_cmp). The main result of the study was to show that 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 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. Abdelrahman et al. [1] also apply Bayesian optimization for maximum power point tracking in photovoltaic power plants.

As a final example, Grover et al. [17] model the search for the best charging policy for the Li-ion battery chemistry as a stochastic multi-armed bandit with delayed feedback. They found policies that considerably outperform current policies.

COMPUTATIONAL SYNERGIES

The previous sections highlight how computational sustainability problems encompass a combination of distinguishing aspects that make them unique in scale, impact, complexity, and richness, posing new challenges and opportunities to 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 themes, some of them depicted in Fig. 3, include *spatio-temporal modeling, and prediction* for, e.g., 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, e.g., phase map identification in materials discovery, identification of elephant and bird calls from audio recordings, inferring plant phenotypes from hyperspectral data and scientific topic modeling; *active learning and optimal learning*, e.g., for scientific experimentation and sensor placement, including citizen science, and crowdsourcing, and *game theory and 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 cross-cutting 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 that these core problems naturally emerge out of *real-world sustainability-driven* projects, approached with the perspective of lifting solution approaches to produce general methodologies, as opposed to only solving specific problems.

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. 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 sciences 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 (www.compsustnet.org) for their many contributions to computational sustainability and the support of two NSF Expeditions in Computing awards (1234567890 and 1234567890). Unfortunately, given the tight space limitations, we are forced to omit a significant amount of other exciting research contributions from this group and other researchers.

REFERENCES

[1] Hany Abdelrahman, Felix Berkenkamp, Jan Poland, and Andreas Krause. 2016. Bayesian optimization for maximum power point tracking in photovoltaic power plants. In *2016 European Control Conference, ECC 2016, Aalborg, Denmark, June 29 - July 1, 2016*. 2078–2083. https://doi.org/10.1109/ECC.2016.7810598

[2] Jo H. Albers, Tom Dietterich, Kim Hall, Lee Katherine, and Majid Taleghan. 2018. Simulator-Defined Markov Decision Processes: A Case Study in Managing Bio-invasions. In *Artificial Intelligence and Conservation* (2nd. ed.), Fei Fang, Milind Tambe, Bistra Dilkina, and Andrew Plumptre (Eds.). Cambridge University Press, New York, NY.

[3] Javad Azimi, Xiaoli Z. Fern, and Alan Fern. 2016. Budgeted Optimization with Constrained Experiments. *J. Artif. Intell. Res.* 56 (2016), 119–152. https://doi.org/10.1613/jair.4896

[4] Junwen Bai, Yexiang Xue, Johan Bjorck, Ronan Le Bras, Brendan Rappazzo, Richard Bernstein, Santosh K. Suram, Robert Bruce van Dover, John M. Gregoire, and Carla P. Gomes. 2018. Phase Mapper: Accelerating Materials Discovery with AI. *AI Magazine* 39, 1 (2018), 15–26. https://www.aaai.org/ojs/index.php/aimagazine/article/view/2785

[5] Garrett Bernstein, Ryan McKenna, Tao Sun, Daniel Sheldon, Michael Hay, and Gerome Miklau. 2017. Differentially Private Learning of Undirected Graphical Models Using Collective Graphical Models. In *Proceedings of the 34th International Conference on Machine Learning, ICML*. 478–487.

[6] Di Chen, Yexiang Xue, and Carla P. Gomes. 2018. End-to-End Learning for the Deep Multivariate Probit Model. *ICML* abs/1803.08591 (2018).

[7] Bistra Dilkina, Rachel Houtman, Carla P Gomes, Claire A Montgomery, Kevin S McKelvey, Katherine Kendall, Tabitha A Graves, Richard Bernstein, and Michael K Schwartz. 2017. Trade-offs and efficiencies in optimal budget-constrained multi-species corridor networks. *Conservation biology* 31, 1 (2017), 192–202.

[8] Priya L. Donti, J. Zico Kolter, and Brandon Amos. 2017. Task-based End-to-end Model Learning in Stochastic Optimization. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*. 5490–5500.

[9] Stefano Ermon, Yexiang Xue, Russell Toth, Bistra N Dilkina, Richard Bernstein, Theodoros Damoulas, Patrick Clark, Steve DeGloria, Andrew Mude, Christopher Barrett, et al. 2015. Learning Large-Scale Dynamic Discrete Choice Models of Spatio-Temporal Preferences with Application to Migratory Pastoralism in East Africa.. In *AAAI*. 644–650.

[10] James H Faghmous and Vipin Kumar. 2014. A big data guide to understanding climate change: The case for theory-guided data science. *Big data* 2, 3 (2014), 155–163.

[11] Fei Fang, Thanh Hong Nguyen, Rob Pickles, Wai Y. Lam, Gopalasamy R. Clements, Bo An, Amandeep Singh, Brian C. Schwedock, Milind Tambe, and Andrew Lemieux. 2017. PAWS - A Deployed Game-Theoretic Application to Combat Poaching. *AI Magazine* 38, 1 (2017), 23–36.

[12] Fei Fang, Milind Tambe, Bistra Dilkina, and Andrew Plumptre (eds.). 2018. *Artificial Intelligence and Conservation*. Cambridge University Press.

[13] Daniel Fink, Wesley M Hochachka, Benjamin Zuckerberg, David W Winkler, Ben Shaby, M Arthur Munson, Giles Hooker, Mirek Riedewald, Daniel Sheldon, and Steve Kelling. 2010. Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications* 20, 8 (2010), 2131–2147.

[14] Douglas H. Fisher. 2016. Recent Advances in AI for Computational Sustainability. *IEEE Intelligent Systems* 31, 4 (2016), 56–61. https://doi.org/10.1109/MIS.2016.61

[15] D. Freund, S.G. Henderson, and D.B. Shmoys. 2018. *Sharing Economy: Making Supply Meet Demand*. Springer.

[16] Carla P Gomes. 2009. Computational sustainability: Computational methods for a sustainable environment, economy, and society. *The Bridge* 39, 4 (2009), 5–13.

[17] Aditya Grover, Todor Markov, Peter Attia, Norman Jin, Nicolas Perkins, Bryan Cheong, Michael Chen, Zi Yang, Stephen Harris, William Chueh, and Stefano Ermon. 2018. Best arm identification in multi-armed bandits with delayed feedback. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*. 833–842.

[18] Amrita Gupta, Bistra Dilkina, Dana Morin, Angle Fuller, Andy Royle, Christopher Shutherland, and Carla Gomes. 2018. Trade-offs and efficiencies in optimal budget-constrained multispecies corridor networks. Conservation biology.

[19] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.

[20] Steve Kelling, Alison Johnston, Wesley Hochachka, Marshall Iliff, Daniel Fink, Jeff Gerbracht, Carl Lagoze, Frank La Sorte, Travis Moore, Andrea Wiggins, Weng-Keen Wong, Chris Wood, and Jun Yiu. 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? *PloS one* 10, 10 (2015), e0139600.

[21] Javad Khazaei and Warren B Powell. 2018. 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.

[22] Jörg Lässig, Kristian Kersting, and Katharina Morik (Eds.). 2016. *Computational Sustainability*. Studies in Computational Intelligence, Vol. 645. Springer.

[23] Unite Nations. 2017. The Sustainable Development Goals Report. Retrieved on June 16, 2018 https://unstats.un.org/sdgs/files/report/2017/TheSustainableDevelopmentGoalsReport2017.pdf.

[24] Steven J Phillips, Robert P Anderson, and Robert E Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological modelling* 190, 3-4 (2006), 231–259.

[25] Warren Powell. 2019. A Unified Framework for Stochastic Optimization. To appear in European J. Operational Research.

[26] Mark D Reynolds, Brian L Sullivan, Eric Hallstein, Sandra Matsumoto, Steve Kelling, Matthew Merrifield, Daniel Fink, Alison Johnston, Wesley M Hochachka, Nicholas E Bruns, et al. 2017. Dynamic conservation for migratory species. *Science advances* 3, 8 (2017), e1700707.

[27] Johan Rockström, Will Steffen, Kevin Noone, Åsa Persson, F Stuart Chapin III, Eric Lambin, Timothy M Lenton, Marten Scheffer, Carl Folke, Hans Joachim Schellnhuber, et al. 2009. Planetary boundaries: exploring the safe operating space for humanity. *Ecology and society* 14, 2 (2009).

[28] Cynthia Rudin and Kiri L. Wagstaff. 2014. Machine learning for science and society. *Machine Learning* 95, 1 (2014), 1–9. https://doi.org/10.1007/s10994-013-5425-9

[29] José Francisco Ruiz-Muñoz, Zeyu You, Raviv Raich, and Xiaoli Z. Fern. 2018. Dictionary Learning for Bioacoustics Monitoring with Applications to Species Classification. *Signal Processing Systems* 90, 2 (2018), 233–247.

[30] Stuart J. Russell, Tom Dietterich, Eric Horvitz, Bart Selman, Francesca Rossi, Demis Hassabis, Shane Legg, Mustafa Suleyman, Dileep George, and D. Scott Phoenix. 2015. Letter to the Editor: Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter. *AI Magazine* 36, 4 (2015).

[31] Daniel R Sheldon and Thomas G Dietterich. 2011. Collective graphical models. In *Advances in Neural Information Processing Systems*. 1161–1169.

[32] Daniel R. Sheldon, Andrew Farnsworth, Jed Irvine, Benjamin Van Doren, Kevin F. Webb, Thomas G. Dietterich, and Steve Kelling. 2013. Approximate Bayesian Inference for Reconstructing Velocities of Migrating Birds from Weather Radar. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, July 14-18, 2013, Bellevue, Washington, USA.*

[33] Brian L Sullivan, Jocelyn L Aycrigg, Jessie H Barry, Rick E Bonney, Nicholas Bruns, Caren B Cooper, Theo Damoulas, André A Dhondt, Tom Dietterich, , et al. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation* 169 (2014), 31–40.

[34] Milind Tambe and Eric Rice (eds.). 2018. *Artificial Intelligence and Social Work*. Cambridge University Press.

[35] United Nations. 2014. A World That Counts: Mobilising the Data Revolution for Sustainable Development. Retrieved on June 16, 2018 http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf.

[36] United Nations. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. Retrieved June 16, 2018 from http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

[37] United Nations. 2018. Statement on visit to the USA, by Professor Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights. Retrieved on June 16, 2018, from http://socialprotection-humanrights.org/wp-content/uploads/2018/06/G1812530.pdf.

[38] Nick van de Giesen, Rolf Hut, and John Selker. 2014. The Trans-African Hydro-Meteorological Observatory (TAHMO). *Wiley Interdisciplinary Reviews: Water* 1, 4 (2014), 341–348.

[39] M. Wahabzada, A.-K. Mahlein, C. Bauckhage, U. Steiner, E.-C. Oerke, and K. Kersting. 2016. Plant Phenotyping Using Probabilistic Topic Models: Uncovering the Hyperspectral Language of Plants. *Scientific Reports* 6 (2016).

[40] WCED. 1987. Our Common Future. Published as annex to the General Assembly document A/42/427, Development and International Cooperation: Environment. Report of the World Commission on Environment and Development. Nairobi, Kenya: UNEP..

[41] Xiaojian Wu, Jonathan M. Gomes-Selman, Qinru Shi, Yexiang Xue, Roosevelt Garcia-Villacorta, Suresh Sethi, Scott Steinschneider, Alexander Flecker, and Carla P. Gomes. 2018. Efficiently Approximating the Pareto Frontier: Hydropower Dam Placement in the Amazon Basin.. In *AAAI*.

[42] Yexiang Xue, Ian Davies, Daniel Fink, Christopher Wood, and Carla P Gomes. 2016. Avicaching: A two stage game for bias reduction in citizen science. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. 776–785.

[43] Amulya Yadav, Bryan Wilder, Eric Rice, Robin Petering, Jaih Craddock, Amanda Yoshioka-Maxwell, Mary Hemler, Laura Onasch-Vera, Milind Tambe, and Darlene Woo. 2017. 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*. 150–158.