Computational sustainability meets materials science (*)

Carla P. Gomes[1,†], Daniel Fink[2], R. Bruce van Dover[3], & John M. Gregoire[4]

1 *Cornell University, Department of Computer Science and Institute for Computational Sustainability*
2 *Cornell University, Laboratory of Ornithology*
3 *Cornell University, Department of Materials Science and Engineering*
4 *California Institute of Technology, Division of Engineering and Applied Science*
† *Corresponding author: gomes@cs.cornell.edu*

(*) pre-print, published *in Nature Rev. Materials,* **6**, pages 645–647 (2021),
https://www.nature.com/articles/s41578-021-00348-2

Computational sustainability harnesses computing and artificial intelligence for human well-being and the protection of our planet. Materials science is central to many sustainability challenges. Exploiting synergies between computational sustainability and materials science advances both fields, furthering the ultimate goal of establishing a sustainable future.

Humanity's consumption of Earth's resources endangers our planet and the livelihood of current and future generations. Our Common Future, the 1987 seminal report by the United Nations World Commission on Environment and Development led by Gro Brundtland, highlighted the interconnectedness of environmental, economic, and societal issues pertaining to sustainability, and introduced the notion of sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their needs." In 2015, as part of the United Nations 2030 Agenda for Sustainable Development, 193 countries agreed on 17 ambitious goals, referred to as the Sustainable Development Goals.

Computational sustainability and materials

Planning for sustainable development encompasses complex interdisciplinary decisions spanning a range of questions concerning human well-being, infrastructure (such as sustainable cities and smart grids), and the environmental protection of the Earth and its species. Such complex decision-making is challenging and requires expertise and research efforts in computational sciences and related disciplines. Computational Sustainability is a nascent interdisciplinary field harnessing computing and artificial intelligence (AI) to provide solutions to computational problems concerning balancing environmental, economic, and societal needs for a sustainable future.[1-4]

Materials have been instrumental in humanity's progress and have been key technology enablers. Materials discovery and development will play a critical role in providing greener, more efficient, and accessible technologies for a more sustainable future.[4-6] For example, the Materials Genome Initiative is truly visionary, using computing and AI methods to accelerate materials design and discovery to address key challenges in clean energy, sustainable materials, and human prosperity. The initiative has been very productive and has had significant impact throughout materials research by generating of open-source materials data and computational infrastructure.

Over the past decade, the Institute for Computational Sustainability has developed AI and machine learning (ML) approaches for various computational sustainability applications which have also been harnessed to advance materials discovery, and vice versa. Herein we describe synergies resulting from ecology and materials science applications. We highlight the importance of model interpretability and of the incorporation of prior scientific

knowledge to better condition the models, to produce scientifically meaningful solutions, and to compensate for the paucity of training data.

Multi-entity prediction

Biodiversity is critical for sustaining ecosystem services. Biodiversity loss has increased dramatically since the Industrial Revolution. For example, since the 1970s, North America has lost nearly 30% of the total number of birds.[7] A fundamental question in biodiversity research is how different species are distributed across landscapes over time. The ability to capture interactions between species and their local environments and interactions among different species is essential for addressing many of the core questions in ecology and conservation today. In particular, it is critical to be able to predict the distribution of hundreds of species simultaneously. However, it is computationally challenging for traditional joint species distribution approaches to simultaneously handle numerous species.

To address this challenge, we have developed a general end-to-end deep-learning framework with prior knowledge based on the multivariate probit (MVP) model, a multi-Gaussian model called Deep Multivariate Probit Model (DMVP).[8] DMVP is a general multi-entity model, suitable for predicting the distribution of multiple species and other entities, such as the different types of land cover across geographic regions over time, or the different objects in images. DMVP automatically learns the relative importance of a potentially large number of input features and generates a structured and interpretable latent space to express the associations between entities (such as species or land cover types) and features (such as environmental features), as well as the interactions among the entities. To tackle the computational challenge of integrating the MVP likelihood, DMVP uses a novel parallel sampling process, which allows for efficient deep net training. We applied DMVP to harness high-dimensional data, combining bird observations from eBird, a global citizen-science bird monitoring platform, with remote sensing data. The integration of these data in a unified model uncovered species interactions and habitat associations for the entire North American avifauna, providing key information on species richness at the fine-grain resolution needed for conservation efforts, which is beyond the scale of previous approaches.

This work provided the impetus for tackling multi-property prediction in materials science, resulting in the Hierarchical Correlation Learning of Materials Properties (H-CLMP) model.[9] Analogously to DMVP, H-CLMP (pronounced H-CLAMP) encodes a structured interpretable latent space to capture the materials features' associations and materials properties' interactions. A multi-Gaussian distribution for pair-wise correlation modelling is coupled to a graph attention model for capturing higher-order correlations, collectively providing a sufficiently expressive model for multi-property prediction that remains trainable with modest datasets, because the model design facilitates its conditioning. H-CLMP also incorporates prior scientific knowledge via transfer learning, using physics-based computational data from The Materials Project to encode materials chemistry in a generative model, which is then used to augment the multi-property prediction. H-CLMP outperforms state-of-the-art methods for the prediction of optical absorption properties for solar energy materials in new composition spaces for which there is no available training data.

Pattern demixing

Pattern demixing involves decomposing an observed signal into its constituent parts and is used, for example, for the identification of animal calls in audio recordings used for monitoring wildlife. Human experts perform signal demixing by reasoning about prior knowledge. By prior knowledge, we mean anything that is known about the structure of the problem, such as rules that characterize valid solutions. Crystal-structure phase mapping, a core challenge in materials science, requires demixing noisy mixtures of X-ray diffraction (XRD) patterns into the source XRD signals of the corresponding crystal structures. Crystal-structure phase mapping is a challenging unsupervised pattern demixing task, given that labeled training data are typically not available. Materials scientists tackle phase mapping by reasoning about known crystal structure patterns from existing

databases combined with knowledge about thermodynamic rules of phase diagrams. Nevertheless, often phase mapping is beyond materials scientists' capabilities due to complexities such as the presence of dozens of phase mixtures and alloy-dependent variations in diffraction patterns in high-dimensional composition spaces, creating a major bottleneck in high-throughput materials discovery. We developed deep reasoning networks (DRNets)[10] to address these challenges. DRNets are a general end-to-end framework with seamless integration of data-driven learning with knowledge-driven reasoning about prior knowledge for unsupervised pattern demixing. DRNets produce an interpretable latent space to encode the domain rules, coupled with a generative decoder that encapsulates prior knowledge about known crystal structures patterns. The interpretable latent space of DRNets was inspired by our DMVP model for joint species distributions. In turn, DRNets inspired the generative model of H-CLMP for transferring prior materials chemistry knowledge. DRNets outperform human capabilities and previous methods for crystal-structure phase-mapping problems, solving previously unsolved chemical systems to facilitate the discovery of solar energy materials.

Computational synergies

The dramatic advances in information and computing technology critically rely on the digital representation of information and the universal computing capabilities of digital devices. Additionally, the generality of computational models enables the transferability of findings across domains. Computational sustainability research is a two-way street: it uses computational thinking and methodologies to address sustainability questions and it also leads to foundational contributions to computing and AI, by exposing computational scientists to new challenging problems and formalisms from other disciplines. The resulting general AI methodologies can then be applied to problems across different fields, as discussed herein for materials science. Examples of cross-cutting computational problems, some discussed above, include: multi-entity prediction for species distributions, land cover, image object detection, and materials properties; pattern demixing for crystal-structure phase mapping, identification of animal calls from audio recordings, and inference of plant phenotypes from hyperspectral data; active learning for scientific experimentation and sensor placement, including citizen science and crowdsourcing; and optimization of sequential decision making for managing (renewable) resources and invasive species and for designing science experiments. Computational synergies can be leveraged for exploring the vast materials space, both in the realm of theory and for planning, designing, executing, and interpreting experiments, using autonomous or semi-autonomous systems. At a higher level, there are opportunities for computational synergies to tackle the many other issues concerning materials and their interconnectedness with environmental, economic, societal, and technological aspects, ranging from materials life cycle analysis and supply chain optimization to mechanism and policy design for incentivizing human behaviors for a sustainable future.

Acknowledgements Computational sustainability research has been supported by an Expedition in Computing NSF award (CCF-1522054). eBird has been supported by the Leon Levy Foundation, the Wolf Creek Foundation, and NSF (DBI-1939187). Materials science research has also been supported by the AFOSR Multidisciplinary University Research Initiative (MURI) Program FA9550-18-1-0136, US DOE Award No.DE-SC0020383, and an award from the Toyota Research Institute.

References

1. Carla Gomes, Thomas Dietterich, Christopher Barrett, Jon Conrad, Bistra Dilkina, Stefano Ermon, Fei Fang, Andrew Farnsworth, Alan Fern, Xiaoli Fern, et al. Computational sustainability: Computing for a better world and a sustainable future. *Communications of the ACM*, 62(9):56–65, 2019.

2. David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*,

2019.

3. Deya Chatterjee and Shrisha Rao. Computational sustainability: A socio-technical perspective. *ACM Computing Surveys (CSUR)*, 53(5):1–29, 2020.

4. Philip Ball. Learning from the big picture. *Nature materials*, 17(12):1062–1062, 2018.

5. Martin L Green, Laura Espinal, Enrico Traversa, and Eric J Amis. Materials for sustainable development. *MRS bulletin*, 37(4):303–309, 2012.

6. Daniel P Tabor, Loïc M Roch, Semion K Saikin, Christoph Kreisbeck, Dennis Sheberla, Joseph H Montoya, Shyam Dwaraknath, Muratahan Aykol, Carlos Ortiz, Hermann Tribukait, et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nature Reviews Materials*, 3(5):5–20, 2018.

7. Kenneth V Rosenberg, Adriaan M Dokter, Peter J Blancher, John R Sauer, Adam C Smith, Paul A Smith, Jessica C Stanton, Arvind Panjabi, Laura Helft, Michael Parr, et al. Decline of the north american avifauna. *Science*, 366 (6461):120–124, 2019.

8. Di Chen, Yexiang Xue, and Carla Gomes. End-to-end learning for the deep multivariate probit model. In *International Conference on Machine Learning*, pages 932–941. PMLR, 2018.

9. Shufeng Kong, Dan Guevarra, Carla P. Gomes, and John M. Gregoire. Materials representation and transfer learning for multi-property prediction. *Applied Physics Reviews,* https://doi.org/10.1063/5.0047066, 2021.

10. Di Chen, Yiwei Bai, Wenting Zhao, Sebastian Ament, Dan Guevarra, Lan Zhou, Bart Selman, Bruce van Dover, John Gregoire, and Carla Gomes. Automating crystal-structure phase mapping by combining deep learning with constraint reasoning. *Nature Machine Intelligence*, in press.

Related links

Our Common Future: **https://sustainabledevelopment.un.org/milestones/wced**

Sustainable Development Goals: https://sdgs.un.org/goals

Materials Genome Initiative: https://www.mgi.gov

Institute for Computational Sustainability: http://computational-sustainability.cis.cornell.edu/

eBird: https://ebird.org

Materials Project: https://materialsproject.org/

Liquid Sunlight Alliance: https://www.liquidsunlightalliance.org/