System Integration with Multiscale Networks (SIMoN): A Geospatial Model Transformation Framework for a Sustainable Future

Marisa J. Hughes, Michael T. Kelbaugh, Elizabeth P. Reilly, Victoria J. Campbell, Susama Agarwala, Miller L. Wilt, Andrew R. Badger, Dillon Ponzo, Evan Fuller, Ximena Calderon Arevalo, Alex Fiallos, Daniel Fiume, Lydia Fozo, Jalen Jones, and Praagna Kashyap

ABSTRACT

A team at the Johns Hopkins University Applied Physics Laboratory (APL) developed SIMoN (System Integration with Multiscale Networks), a framework that connects predictive resource models from domains such as water, electricity, climate, and population, and passes data about resource usage and availability between models. SIMoN is useful for interfacing models with different native environments and geospatial definitions and can potentially be adapted to many other applications.

With climates changing; populations growing; resources such as food, water, and energy becoming more and more scarce; and globalization creating complex dependencies, new modeling techniques that can account for dynamic and interdependent systems are needed to evaluate resource sustainability. These techniques must be able to adapt to many highly coupled domains and process the assortment of ever-evolving data and models that often use different units, definitions, and geotemporal scales.

For the independent research and development project described in this article, an APL team developed a new modeling framework that addresses these challenges. Called System Integration with Multiscale Networks (SIMoN), the framework connects predictive resource models from different domains, such as water, electricity, and population, and passes resource utilization and availability data between models. Each model takes discrete time steps and runs in its own Docker container. These containers enable a modular design and customized model environments and translate easily to a computing cluster for scalability. The initial SIMoN modeling framework and models was made open source in April 2020. SIMoN's flexibility allows the user to introduce new domain models, compare results of joint model runs with different configurations, and adapt to new model types or domains. This is in contrast to stovepiped domain models or traditional integrated assessment models, where domain models are inexorably tied together.

SIMoN's novelty stems from its treatment of geospatial regions. Geospatial translation is a challenging modeling problem in integrated models, with each model operating within different geospatial scales and definitions such as counties, states, watersheds, and power regions. In SIMoN, geographies are structured by a partially ordered set of geospatial partitions, with a corresponding directed acyclic graph to organize the complex interplays between regions in numerous compatible shapefiles. SIMoN enables users to define consistent aggregation and disaggregation maps for data transformation between disparate notions of geography. Consistency of data aggregation/disaggregation pairs is defined using a set of mathematical axioms, including a right inverse property. This limits the propagation of error that can be introduced by the repeated data exchange between geography types at each time step. Once carefully defined, geospatial data transformations are performed automatically as needed by wrappers around each model. This unique approach to geospatial transformation can be extended to provide flexible tools to tie models together across domains. See Hughes et al.[1] for a detailed description of this approach.

The focus of SIMoN is to provide tools for combining existing models. Initial SIMoN experiments have focused on publicly available data sets and models with basic but representative functionality for proof of concept. Models currently implemented for demonstration include:

1. Two county-level population models extrapolating US census data[2] developed for comparison: one implemented using the Holt's linear trend method[3] and the other based on logistic population growth. Each model predicts county-level populations from the previous years' data.

2. A simple model of per-capita power demand based on total state-level power sales in 2016.[4] This model is modified by the change in global mean surface temperature to scale the demand for power generation in a first-order approximation.

3. A model of power generation that assumes generation always meets demand for each power region, specifically the North American Electric Reliability Corporation (NERC) regions, then constructs a power production profile based on the generation-weighted production of power plants within each NERC region. This profile is generated using US Energy Information Administration (EIA) plant-level emissions and water consumption data.[5] Important outputs of this model include greenhouse gas emissions and water consumption for power plant cooling.

4. Data from climate models run by the scientific community, including Global Climate Model (GCM) runs such as GFDL-CM3.[6] Results of these runs are based on representative concentration pathways, which represent scenarios of time-dependent greenhouse gas concentration projections and can have outputs of near-surface air temperature, precipitation, and evaporation.

5. An open-source climate model integrated into the framework, FAIR: Finite Amplitude Impulse Response simple climate model.[7] This model takes in global emissions and converts them into greenhouse gas concentrations and effective radiative forcings. Based on the effective radiative forcings, the model calculates the global mean surface temperature change.[8]

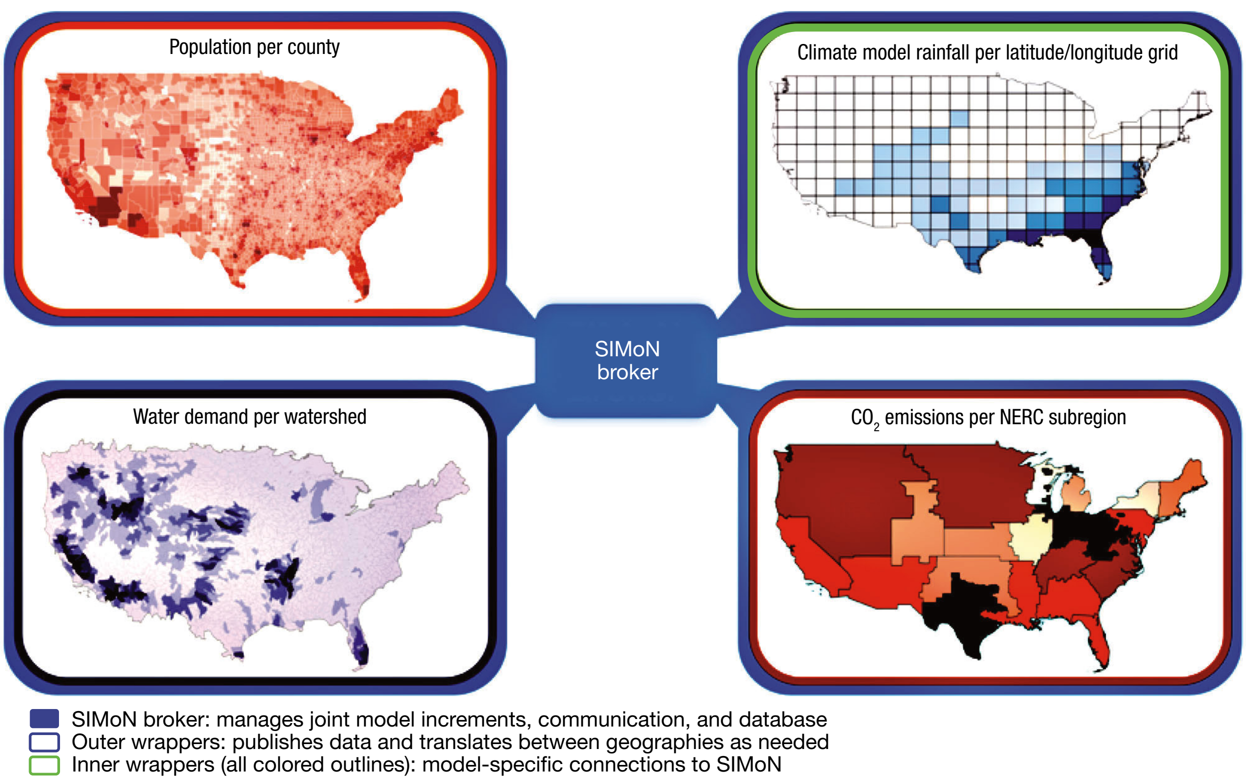


Figure 1. SIMoN architecture diagram with model output snapshots from the year 2050. Note: These results are intended to demonstrate SIMoN modeling communication and should not be taken at face value until model validation efforts are completed. (© 2020 IEEE. Adapted, with permission, from Hughes et al.[1])

6. A new Python model for accounting for available water in the contiguous United States. The model follows the general guidelines used to determine water resource budgets for each area as described by the Michigan Water Resources Division assuming a steady state water budget.[8] Water availability is derived from input (rainfall) and consumption (evaporation, evapotranspiration, and human use) for each Hydrologic Unit Code 8 (HUC 8) watershed[10] and transfers any unused water for a given year into the next downstream watershed.
7. A water demand model that calculates each US county's water demand[11] by multiplying its population by its water consumption per capita and then adding its thermoelectric water usage from the power supply model's output. To calculate water consumption rates for each county, values for "thermoelectric recirculating, total consumptive use, fresh in Mgal/d" were subtracted from values for "irrigation and thermoelectric water, total consumptive use, fresh in Mgal/d." To convert this daily rate to an annual rate, the difference was divided by the county's 2015 population and then multiplied by 365.

It is up to the modeler running the SIMoN framework to decide which models to use and how these models should interact within the framework. An example joint modeling result is given below, representing predicted resource supply and demand levels in the continental United States in the year 2050. These results demonstrate the capability of the system to integrate disparate models. You can see how the different overlapping geographies would present a challenge to data exchange between models as these models exchange data after each yearly time step, an issue addressed in the SIMoN wrappers shown in Figure 1.

Resource modeling predictions from SIMoN were compared to other state-of-the-art models for validation. These comparisons indicate that SIMoN's integrated resource models are on the right track, for example yielding water shortages in the same regions.[12,13] However, validation efforts via backcasting (i.e., predicting the present with past data) indicate that SIMoN's population projections tend to grow too quickly, hastening the shortfall of critical resources. We theorize that this disparity stems from the emerging phenomenon of population growth slowing in response to technological and economic advancement, making historical populations alone a poor predictor of future growth. More sophisticated population models are currently in development.

SIMoN shows promise as a general tool for interfacing models with different native environments and geospatial definitions. Since its purpose is to build a bridge between disparate domains, it has the potential to be adapted to many other applications. Next steps for SIMoN could include automatic consistency checks, tools for defining new geographies, metrics for model interdependency and joint uncertainty, native model validation support, and integration with machine learning methods. SIMoN's ability to calculate metrics that span systems is the first step in a generalized decision-making framework, which could optimize the development and placement of new mitigation and adaptation technologies to address increasing demand, resource depletion, and the increasing pressures on human–Earth systems induced by climate change.

ACKNOWLEDGMENTS: We acknowledge support from APL's Research and Exploratory Development Mission Area. We also thank the organizers of APL's Cohort-based Integrated Research Community for Undergraduate Innovation and Trailblazing program—or CIRCUIT—for funding support, training. and coordination of the students on this project. The CIRCUIT program aims to help high-potential undergraduate students overcome barriers to STEM careers. It has expanded from one sponsored summer project to eight yearlong efforts involving four APL mission areas.

REFERENCES

[1]M. Hughes, M. Kelbaugh, V. Campbell, E. Reilly, S. Agarwala, et al., "System Integration with Multiscale Networks (SIMoN): A modular framework for resource management models," in *Proc. 2020 Winter Simulation Conf.*, K.-H. Bae et al., Eds., Dec. 2020, pp. 656-667, https://doi.org/10.1109/WSC48552.2020.9383983.

[2]"Population, population change, and estimated components of population change: April 1, 2010, to July 1, 2018," County Population Totals: 2010–2019. United States Census Bureau. 2018. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html.

[3]R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. OTexts, 2018. https://otexts.com/fpp2/.

[4]"Formeia-861annualelectricitysales." Annual Electric Power Industry Report, Form EIA-861 detailed data files. US Energy Information Administration. https://www.eia.gov/electricity/data/eia861/.

[5]"Formeia-923." Form EIA-923 detailed data with previous form data (EIA-906/920). US Energy Information Administration. https://www.eia.gov/electricity/data/eia923/.

[6]Z. Hausfather, "SimMod: A simple Python based climate model," *Berkeley Earth*, Jun. 7, 2016, http://berkeleyearth.org/simmod-a-simple-python-based-climate-model/.

[7]X Smith, C. J., P. M. Forster, M. Allen, N. Leach, R. J. Millar, G. A. Passerello, and L. A. Regayre, "FAIR v1. 3: A simple emissions-based impulse response and carbon cycle model," *Geosci. Model Develop.*, vol. 11, no. 6, pp. 2273–2297, 2018, https://doi.org/10.5194/gmd-11-2273-2018.

[8]S. M. Griffies, M. Winton, L. J. Donner, L. W. Horowitz, S. M. Downes, et al., "The GFDL CM3 coupled climate model: Characteristics of the ocean and sea ice simulations," *J. Climate*, vol. 24, no. 13, pp. 3520–3544, 2011, https://doi.org/10.1175/2011JCLI3964.1.

[9]Michigan Water Resources Division, "General guidelines for calculating a water budget," Mar. 2010. https://www.michigan.gov/documents/deq/wrd-water-budget_565040_7.pdf.

[10]"National Hydrography Products." National Hydrography. USGS. https://www.usgs.gov/core-science-systems/ngp/national-hydrography/access-national-hydrography-products.

[11]C. A. Dieter, K. S. Linsey, R. R. Caldwell, M. Harris, T. I. Ivahnenko, et al., "Estimated use of water in the United States county-level data for 2015" (ver. 2.0, Jun. 2018), US Geological Survey data release, https://doi.org/10.5066/F7TB15V5, accessed Jul. 3, 2019.

[12]V. C. Tidwell, P. H. Kobos, L. A. Malczynski, W. E. Hart, and G. T. Klise, "Decision support for integrated water-energy planning," Technical report, Sandia National Laboratories, Albuquerque, NM, 2009.

[13]J. Abatzoglou, S. Dobrowski, S. Parks, and K. C. Hegewisch, "TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015," *Sci. Data*, vol. 5, article 170191, pp. 1–12, 2018, https://doi.org/10.1038/sdata.2017.191.

Marisa J. Hughes, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Marisa J. Hughes is assistant supervisor of the Reasoning Machines Section and the assistant program manager for environmental resilience in APL's Research and Exploratory Development Department. She has a BS in mathematics from Binghamton University and an MS and a PhD in mathematics from Cornell University. Marisa's research interests include methods to understand, mitigate, adapt, and reverse environmental threats including climate change, resource shortages, and threatening waste. She has experience in data science, data modeling, systems analysis, machine learning, and management, as well as in Earth systems, intelligence, missile defense, health, neuroscience, and submarine maintenance domains. Her email address is marisa.hughes@jhuapl.edu.

Michael T. Kelbaugh, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Michael T. Kelbaugh is a member of the Associate Professional Staff in APL's Research and Exploratory Development Department. He has a BS in computer science and a BA in economics from the University of Maryland, Baltimore County. His email address is michael.kelbaugh@jhuapl.edu.

Elizabeth P. Reilly, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

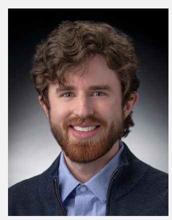
Elizabeth P. Reilly is a project manager and assistant group supervisor of the Artificial Intelligence Group in APL's Research and Exploratory Development Department. She has a BS in mathematics from Wake Forest University, an MA in mathematics from the University of South Carolina, and a PhD in applied mathematics and statistics from Johns Hopkins University. Elizabeth has a special interest in graph theory, connectomics, graph signal processing, and artificial intelligence for climate change. She is the co-lead of the Motifs to Models efforts and has focused on probabilistic approaches to identifying motifs in noisy networks. Her email address is elizabeth.reilly@jhuapl.edu.

Victoria J. Campbell, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Victoria J. Campbell is a mechanical engineer in APL's Research and Exploratory Development Department. She has a BS and an MEng in mechanical engineering from Cornell University. Her work has ranged from detailed mechanical design and drawings for electronics packaging to finite element analysis for high-speed impact. Other research interests include augmented reality (AR) and its development, applications, and integration, as well as climate change and its impact on Earth systems, renewable energy, and disasters, in association with geoengineering concepts. Her email address is victoria.campbell@jhuapl.edu.

Susama Agarwala, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Susama Agarwala is a theoretical mathematician in APL's Research and Exploratory Development Department. She has a BS in physics and a BS in mathematics from the Massachusetts Institute of Technology (MIT) and an MA and a PhD in mathematics from Johns Hopkins University. Susama brings a deep knowledge of many fields of mathematics to a diversity of projects and subject areas including the Defense Advanced Research Projects Agency (DARPA) programs CASCADE (Complex Adaptive System Composition And Design Environment) and L2M (Lifelong Learning Machines). She mathematically investigates issues of artificial intelligence safety using, among other things, ideas from Riemannian geometry enriched connectomics. Her email address is susama.agarwala@jhuapl.edu.



Miller L. Wilt, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Miller L. Wilt is software engineer in APL's Research and Exploratory Development Department. He has a BS in computer engineering from Lehigh University and an MS in computer science from Johns Hopkins University. Miller has a diverse set of programming and DevOps skills, including machine learning, full-stack web development, Android development, and software deployment. He is part of a cross-sector research and development effort investigating the use of machine learning to classify the type of data transmitted over a wireless communications link; leads a team assessing the architecture, design, code quality, and security of a government-developed application; and has architected, designed, and developed various systems. Miller has been awarded three internal intellectual property (IP) disclosures resulting in multi-million-dollar follow-on contracts and received three APL special achievement awards for outstanding technical excellence. His email address is miller.wilt@jhuapl.edu.

Andrew R. Badger, Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Andrew R. Badger is a member of the Associate Professional Staff in APL's Research and Exploratory Development Department. He has a BS in computer engineering from the University of Maryland, College Park and an MS in applied biomedical engineering from Johns Hopkins University. His email address is andrew.badger@jhuapl.edu.

Dillon Ponzo, Amazon Web Services, Seattle, WA

Dillon Ponzo received a BS in environmental engineering from Johns Hopkins University in 2016, as well as an MSE in computer science in 2019. He completed this work as a graduate student intern in APL's Research and Exploratory Development Department. He currently works as a software development engineer at Amazon Web Services. Previously, he worked as a software developer and researcher in APL's Intelligent Systems Center. His email address is dponzo18@gmail.com.

Evan D. Fuller, Department of Defense, Washington, DC

Evan D. Fuller is a data scientist with the US Department of Defense in the greater Washington, DC, area. He received a BS in mathematics from the University of Texas at Austin in 2004, followed by a PhD in combinatorics from the University of California, San Diego in 2009. Following that, he worked as a professor at Montclair State University in New Jersey. His email address is fuller.evan@gmail.com.

Ximena Calderon Arevalo, APL CIRCUIT Intern

Ximena Calderon Arevalo is a junior studying applied mathematics and statistics at APL. She began interning at APL in May 2019 with the CIRCUIT Program. Her interests include optimization and data analysis. Her email address is ximenacalderon0917@gmail.com.

Alex Fiallos, APL CIRCUIT Student

Alex Fiallos is a graduating senior from Johns Hopkins University with a BS in biomedical engineering and applied mathematics. He joined APL as a CIRCUIT intern. His research interests include data science, computer vision, and software development. His email address is afiallo1@jhu.edu.

Daniel Fiume, APL CIRCUIT Intern

Daniel Fiume is a freshman at Brown University. He is interested in studying applied math, computer science, and physics. In the future, he plans to apply these skills to some of the biggest challenges our society faces today, including climate change, income inequality, and public health.

Lydia Fozo, CIRCUIT Intern

Lydia Fozo is studying chemical and biomolecular engineering with a minor in computational medicine at Johns Hopkins University. She is an intern at APL in the CIRCUIT program. Her interests include machine learning, data analysis, and personalized medicine. Her email address is lfozo1@jhu.edu.

Jalen Jones, APL CIRCUIT Intern

Jalen Jones is studying public health at Johns Hopkins University. She is an intern at APL in the CIRCUIT program. Her interests include health equity, health surveillance, population demographics, and epidemiology. She is a Bloomberg Scholar. Her email address is jjone264@jhu.edu.

Praagna Kashyap, APL ASPIRE Intern

Praagna Kashyap is a rising high school senior at Centennial High School in Howard County, Maryland. She is interested in pursuing international affairs and public health in her undergraduate studies. Currently, she operates an online tutoring program for ESOL students and students affected by COVID-19. Praagna conducted background research on crops and the food supply in Uganda and collaborated with Johns Hopkins University undergraduate students using Python to build the SIMoN framework.