Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

HYBRID SIMULATION DEVELOPMENT – IS IT JUST ANALYTICS?

David Bell
Derek Groen

Department of Computer Science
Brunel University London
Kingston Lane,
Uxbridge, UB8 3PH, UK

Navonil Mustafee

Centre for Simulation, Analytics and Modelling
University of Exeter Business School
Streatham Court, Rennes Drive
Exeter, EX4 4ST, UK

Jonathan Ozik

Decision and Infrastructure Sciences
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439, USA

Steffem Strassburger

Information Technology in Production & Logistics
Technische Universität Ilmenau
P.O. Box 100 565
Ilmenau, 98684, GERMANY

ABSTRACT

Hybrid simulations can take many forms, often connecting a diverse range of hardware and software components with heterogeneous data sets. The scale of examples is also diverse with both the high-performance computing community using high-performance data analytics (HPDA) to the synthesis of software libraries or packages on a single machine. Hybrid simulation configuration and output analysis is often akin to analytics with a range of dashboards, machine learning, data aggregations and graphical representation. Underpinning the visual elements are hardware, software and data architectures that execute hybrid simulation code. These are wide ranging with few generalized blueprints, methods or patterns of development. This panel will discuss a range of hybrid simulation development approaches and endeavor to uncover possible strategies for supporting the development and coupling of hybrid simulations.

1 INTRODUCTION

The ability to analyze an ever growing sea of data is clearly important in many domain islands, from financial analysis and subsequent trading to migrant movement analysis and humanitarian response. The computational capability that can be used to analyze this 'big'-data has also facilitated the level at which this analysis can take place. Simulations are now more likely to operate at near realistic levels of representation and generate vast amounts of data. Analytics is a term that is widely used to describe the process of data analysis, with Davenport (2006) describing a process of adapting systems that produce data, making data available, selecting analytical approaches and deploying architectures to operationalize this process. Lycett terms this new data-centric approach as 'datafication' and highlights a number of challenges including the manner in which data is conceptualized, analyzed and visualized. Here we extend 'datafication' challenges to include coupling and the normalization of consumed and generated data. Simulation is also able to respond in part to these challenges and also offer causation and sensitivity analysis. Hybrid simulation typically applies a number of techniques in the model implementation phase of a simulation study in order to gain further insights (Mustafee and Powell 2018). These insights may emerge from the use or generation of additional input or output data and additional analytical perspectives. Figure 1 positions this operationalization of hybrid simulation as a number of models are implemented, integrated and verified. Importantly, the integration also requires a robust coupling strategy and design in order to manage the complex operation.

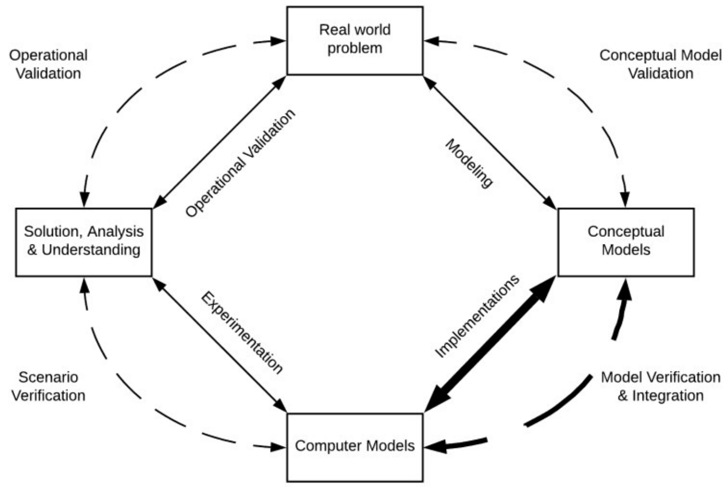

Figure 1: Stages of a hybrid simulation study - adapted from Mustafee and Power (2018).

Integration and coupling of simulation models is particularly complex and covers a varied number of patterns. Five popular approaches include: 1) Semantic integration, 2) Multi-scale integration, 3) multi-paradigm integration, 4) Multi-platform integration and 5) Multi-processing integration (including parallelization). Semantic integration provides the underpinning of simulation models with common or related conceptual language. Alowayyed et al. (2017) highlighted computational challenges when coupling across a range of scales that include near realistic representations. Multi-paradigm is more typical in the hybrid community, where a number of techniques (e.g. System Dynamics and Discrete Event Simulations) are applied in model implementation phase of a simulation study in order to gain further insights (Mustafee and Powell 2018). Multi-platform connect across hardware or software platforms (typically between commercial software packages). Finally, multi-processing coupling covers both parametric studies with looser coupling and MPI style tight coupling.

Implementation(s) is the translation from more abstract conceptual models into computer executable models - enabling data analysts or scientists to experiment with different perspectives of the problem in a dynamical way (Eldabi et al. 2018). The focus of this panel is on this translation from both process and resulting simulation model perspectives. Eldabi et al. (2018) previously examined: 1) Commercial packages, 2) programming languages, 3) model integration process and 4) model input sources. Model integration was categorized from more manual techniques through to the tools and programming interfaces. It is this integration process that underpins positions presented in this paper. The panelists will be invited to answer the following questions:

1. How does hybrid simulation differ from analytics?
2. What does a hybrid simulation design look like and how does it link to more conceptual Modeling?
3. How can machine learning better support hybrid simulation development?
4. What are the challenges in hybrid experimentation?

The four panelist present a range of views. Mustafee starts with a focus on types of analytics and their relationship to Hybrid Simulation (HS) and Hybrid Modelling (HM). Strassburger discusses the opportunities offered by machine learning (ML) and argues the benefit of a single execution environment.

Both Groen and Ozik then explore the scaling and coupling of required computation in high performance settings.

2 HYBRID MODELING WITH SIMULATION AND DATA-DRIVEN ANALYTICS (NAVONIL MUSTAFEE)

The panelists were asked the following question – *"How does HS differ from Analytics?*". In responding to this question, I will first present my definition of analytics and distinguish between hybrid simulation and hybrid Modeling. These definitions will then frame my response on the intersection of analytics with conventional one-technique simulations (e.g., DES model, ABS model), hybrid simulation (e.g., SD with DES) and hybrid Modeling.

Analytics Definition: Analytics, as a term, is gaining increasing prevalence in areas such as HR (people analytics), marketing (marketing analytics, behavioral analytics), social media (social learning analytics) and teaching/Higher Education (learning analytics). In most cases, analytics encompasses data-driven decision making that is reliant on historical datasets, Big Data (Big Data analytics), and increasingly, real-time data (real-time analytics); the specific method of analysis could be classical OR/statistical techniques (e.g., mathematical models, regression Modeling), methods from data science and AI (e.g., machine learning approaches), computer science (e.g., distributed computing approaches and frameworks for large-scale data processing), or indeed computer simulation. I am using the terms 'analytics' and 'data analytics' as synonyms, categorizing the latter as either descriptive analytics, predictive analytics or prescriptive analytics. Mustafee et al. (2018) have presented the following definitions:

- Descriptive analytics analyses and presents data using techniques such as descriptive statistics, data summaries and real-time reporting;
- Predictive analytics, in its most general sense, refers to any method which can support predictions about what might happen including data mining, forecasting, mathematical approaches (Delen and Demirkan 2013; Waller and Fawcett 2013; Shao et al. 2014) and machine learning (Barga et al. 2015);
- Prescriptive tools inform decision making by suggesting a solution path, for example, simulation can anticipate the consequences of unforeseen interactions and prescribe interventions on the basis of tested scenarios (Marshall et al. 2015).

Hybrid Simulation and Hybrid Modelling Definition: HS is the combined application of multiple simulation techniques (e.g., SD, ABS, DES) for the purposes of achieving a better representation of the system under scrutiny. Powell and Mustafee (2017) introduce a wider definition of hybrid simulation, which is characterized by its focus on interdisciplinary research, where techniques and methods from various disciplines may be combined with traditional simulation methods in different stages of the simulation life-cycle (conceptual Modeling, input/output data analysis, model development, experimentation, V&V). This is referred to as Hybrid Modeling (HM). Unlike HM, the focus of HS is mostly confined to the M&S community. Although HS has made a notable contribution in the development of our discipline, in particular, it has been successful in reducing the silos specific to Modeling methodologies (From: *"I am an SD modeller and my expertise is in holistic analysis"* and *"I do DES and I excel in the detailed analysis of the system"*; To: *"I offer analysis at multiple resolutions and often choose a combination of techniques that best answers the questions at hand"*), it can be critiqued as it fails to engage with other disciplines. For our community to be more "external" focused and to learn from research approaches and paradigms, methodologies and techniques that have existed/being developed in other fields of study, it is important for the M&S community to take a leap from *"Hybrid Simulation to Hybrid Modeling"* (Mustafee and Powell 2018).

How does Hybrid Simulation differ from Analytics? In my view, this will depend on whether descriptive, predictive or prescriptive analytics, or indeed a combination of them, is being considered.

Descriptive analytics is usually defined as Business Intelligence (BI) (Chen et al. 2012). One key element of BI applications is the use of real-time data. The focus of my panel piece is also on analytics-based approaches that churn real-time or near real-time data (Big Data analytics, real-time analytics), and how they differ from HS. Also, and perhaps more importantly and in keeping with the stated future direction of research on HM (rather than only HS), how do we conceptualize an HM that combines data-driven analytics and computer simulation?

Simulation models (including HS) use statistical distributions that are usually derived from historical data. For a simulation to benefit from real-time data (for example, to change input distributions based on data feeds that are continuously updated), would generally require integration of a real-time data-acquisition engine with the simulation model (Mustafee et al. 2018; Ongoo et al. 2018). Such integration is arguably outside the core expertise of OR/simulation researchers and practitioners (who are experts at model development, experimentation and analysis). Thus, in relation to both descriptive and predictive analytics, neither the conventional one-technique simulation nor HS will benefit from real-time data and the analytics-based approaches supporting it (Table 1). A combined application of data-driven analytics with simulation can, however, be achieved through an HM, wherein simulation researchers extend their models using methodologies, standards, tools and software from the field of Computer Science, Applied Computing, Industrial Engineering, Data Science, etc. (based on the choice of specific descriptive and predictive techniques). A simulation is a form of prescriptive analysis, and therefore any form of simulation (whether hybrid or not) can be considered as prescriptive analytics (Table 1; last row).

Table 1: Intersection of data-driven analytics (with a focus on real-time data) with simulation approaches**.**

Data-Driven Analytics	**Conventional Simulation**	**Hybrid Simulation**	**Hybrid Model**
Descriptive Analytics	No	No	Yes *(e.g., real-time data feeds can be integrated with the model)*
Predictive Analytics	No	No	Yes (e.g., *forecasting approaches from CS, OR and Data Sciences can be used for better representation of future scenarios as distributions based on forecasts will better reflect what is likely to happen; forecasts can trigger automatic execution of simulation experiments for short-term decision making)*
Prescriptive Analytics	Yes	Yes	Yes

<u>HM model in practice</u>: A conceptualization of an HM that combines the various forms of analytics with computer simulation is the *Right Hospital-Right Time (RH-RT)* framework for urgent care services (Mustafee et al. 2018). An application of this framework is the NHSquicker digital platform (https://nhsquicker.co.uk). Near real-time data on Emergency Department (ED) and Urgent Care Centre (UCC) wait times are received from ~24 centers of urgent care (including six hospitals with 24/7 ED) operating in Cornwall, Devon and Somerset. The descriptive analytics combines wait time information with travel times (retrieved from Google Maps API), to “nudge” users to visit centers of urgent care that may further away but where they will be seen quicker. As ED departments tend to be very busy, the descriptive analytics approach tries to level demand for urgent care, by nudging non-acuity patients (non-life-threatening conditions) away from hospitals to other centers of urgent care. The predictive analytics component of RH-RT uses forecasting models on time-series data (downloaded every 30 minutes through a back-end system) to make short-term predictions (1-4 hours) on how busy an ED/UCC is likely to be (Harper and Mustafee 2019). The prescriptive analytics component is a DES model of ED/UCC. Figure 2

identifies several feedback elements (represented as dashed circles) that underline the need for a joined-up approach when using the descriptive, predictive and prescriptive analysis and the potential of using the output from one form of analysis as the input to subsequent analysis. Yet another feedback loop intersects the input (near real-time data) and the analysis components of RH-RT. This loop highlights the need to constantly monitor the predictions with real-time data, to calibrate the analytical models.

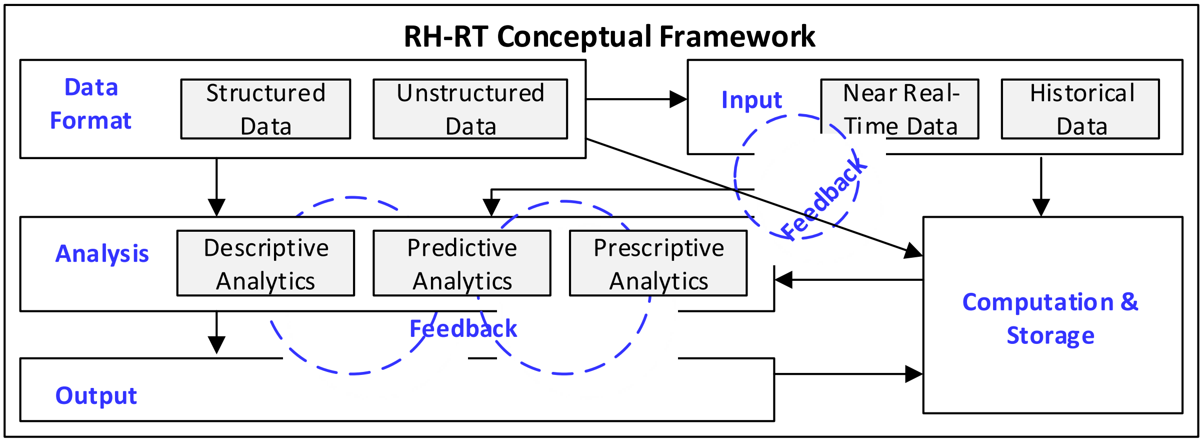

Figure 2: The RH-RT framework for conceptualization of HM comprising of data-driven analytical approaches and computer simulation (Mustafee et. al. 2018).

3 MACHINE LEARNING AND SIMULATION (STEFFEN STRASSBURGER)

Simulation is a well-established tool for planning, monitoring, and controlling complex systems. For a computer simulation of a real system it is indispensable to create a model of this system. Simulation models are generally abstractions of the real-world system under observation and focus on its most relevant components and attributes. In traditional, monolithic modeling approaches, the modeler is bound to and potentially limited by the chosen modeling paradigm. As these monolithic modeling approaches have been explored and investigated for many years, no groundbreaking innovations in the respective methodologies themselves are to be expected.

On the other hand, simulation as an analysis and prediction tool is nowadays often challenged by emerging new technologies like machine learning and artificial intelligence (AI). These technologies are often used to handle massive amounts of data and to answer complex analysis questions.

In order to stay relevant, simulation has to address these challenges. We foresee two paths for achieving this. Both can be considered “hybrid”.

The first path is the combination of simulation with machine learning techniques to improve the analysis of simulation output. The concept of Data Farming is the prototype of this idea. Within Data Farming, a simulation model serves as data generator (Brandstein and Horne 1998; Horne and Meyer 2005). By using efficient experimental design alongside high performance computing, one is able to maximize the data yield and corresponding information gain. The goal is to cover the whole experimental bandwidth of the possible model behavior. The farming metaphor describes how the data output can be maximized by experimental designs like a farmer that cultivates his land to maximize his crop yield (Sanchez and Sanchez 2017). Originally developed for military applications, Data Farming has far more potential to reinvent how simulation studies are conducted and has been proven successful for example in domains like manufacturing (Feldkamp et al. 2015). It enables us to connect the simulation model with a wide range of tools and analysis methods in order to investigate the large amount of simulation data. For this purpose we developed a process called Knowledge Discovery in Simulation Data (Feldkamp et al. 2015). This process allows us to use the Data Farming concept alongside Data Mining and Visual Analytics (Keim et al. 2008) in order to discover complex relations in the model and ultimately gain knowledge. We have shown in several real world case studies (Feldkamp et al. 2016; Feldkamp et al. 2017) that this approach can indeed uncover hidden relations in the model. While this approach constitutes a hybrid method in the sense that multiple analysis methods are used in combination with simulation, the simulation model itself does not necessarily have to be a hybrid system model (HSM) for this; on the contrary it is most often a traditional model using a monolithic modeling approach.

The second path for advancing the area of simulation is the idea of combining different modeling approaches to build a simulation model. The general idea of this concept is not new: combining discrete-event simulation and continuous simulation has long been known as "combined simulation" and could be considered the prototype of a hybrid simulation method.

In the best case the construction of a hybrid system model can be done within a single simulation tool that supports multiple modeling paradigms or *world views*. An example of such a hybrid system modeling approach for depicting the energy consumption in production simulations is given in (Wörrlein et al. 2019). Here, we use the simulation system Anylogic and all three Modeling methods that Anylogic offers to build a model of a production system and its energy consumption. Using a process-oriented world view we model the general production process with its resources. For the detailed operation logic of machines with their different operation modes (warmup, fast-warm-up, standby, idle, …) we use the agent based modeling approach with its state charts for modeling machine behavior. For depicting energy consumption, we apply systems dynamics with its stock and flow diagrams to replay previously measured fine-granular energy consumption data of each individual energy consumer. For building such a hybrid system model, it is beneficial to have some kind of blueprint that guides the user in the decision which modeling method should be used for which aspects of the model. For modeling manufacturing systems with their energy consumption, such a blueprint has been suggested in (Wörrlein et al. 2019).

In general, it will be highly beneficial to stay within the boundaries of a single simulation system even when performing multi-method modeling. Although distributed simulation based on the High Level Architecture (HLA) (Strassburger 2001) in principle offers a solution for interoperability between heterogeneous simulation systems (Taylor et al. 2012), and thus also enables hybrid systems modeling, the ease-of-use argument still weighs heavily when deciding which modeling tools to use. For some scenarios, especially when existing and complex model components shall be reused and brought together, distributed simulation may be the best choice for enabling hybrid simulation, for others, especially when building models from scratch, it may not.

We currently see another important trend for hybrid systems modeling that typically cannot be performed within the boundaries of a single simulation system: the combination of traditional simulation modeling approaches with machine learning. Such a combination is well suited for problems in which some behavioral aspects of the real system are difficult to model within the modeling paradigm of the chosen simulation system, but where enough data from the real system exists to have some machine learning method try to emulate the behavior represented in the data. We applied this approach in several scenarios. In (Bergmann et al. 2017) a proof of concept shows how manufacturing control rules can be obtained from adequately trained artificial neural networks (ANN). Here, all behavioral aspects except a decision rule were modeled in a process-oriented world view in the simulation system. Only a single type of decision was obtained from an ANN that was coupled online via appropriate interfaces to the simulation system. This approach is highly relevant, because the determination of exact decision rules (such as FIFO, LIFO, …) is in practical scenarios often very difficult. Using an ANN to approximate the decision rule based on observed data sets is an elegant solution to solve this problem.

A more complex scenario is introduced in (Wörrlein et al. 2019). Here, entire time series data of energy consumption of machines is predicted by deep neural networks and injected into a traditional discrete-event simulation model. While the setup and the training of these networks is highly complex, the combination of simulation and deep learning brings together the best of both worlds concerning state of the art analysis and prediction tools.

In summary, we argue that in the era of big data and machine learning, and in order to stay relevant, simulation must evolve. Hybrid modeling and simulation offers a good way to enhance the capabilities of simulation. It can even be used to bring simulation modeling and machine learning together in various ways.

4 HYBRID SIMULATION PROTOTYPING (DEREK GROEN)

Simulations are a simplified representation of reality, capturing a subset of its enormous range of physical processes and features, constructed and applied for a specific purpose, using a computer (though exception exist (Holmberg 1941). Though simulations can be used to represent a single process, it is much more common for simulations to approximate a behavior where multiple processes contribute and/or interact, possibly across different time and space scales (e.g, as formalized by Borgdorff et al. 2013). As an example, see Figure 3, which provides an overview of the process couplings we are trying to establish as part of the EU-funded VECMA and HiDALGO projects.

In a simulation, processes may be approximated using mathematical models, heuristics, data collections, or even hardware devices, e.g., GRAPE (Makino and Daisaka 2012) or ANTON (Shawe al. 2014). While hybrid analytics may concern the synthesis of data sources and learning or analysis procedures to achieve a particular purpose, hybrid simulation extends beyond that. It encapsulates the human scientific knowledge of how processes work in computational sub models, combines this with other observations or models, and uses the knowledge embedded in these models to pursue new scientific discoveries. When aggregating sub models, a key factor is the contribution of an individual model's output to the overall outcome of the simulation in the regime of scientific interest. That contribution is not always known in advance, and the developer needs to be aware of the need to incorporate new models previously unthought of, due to emerging behaviors or requirements. As a result, a good way to look at hybrid simulation design and development is from an exploratory research perspective. Simply put: when developing an approach, expect to rewrite it a few dozen times, and adapt your design and implementation approach accordingly.

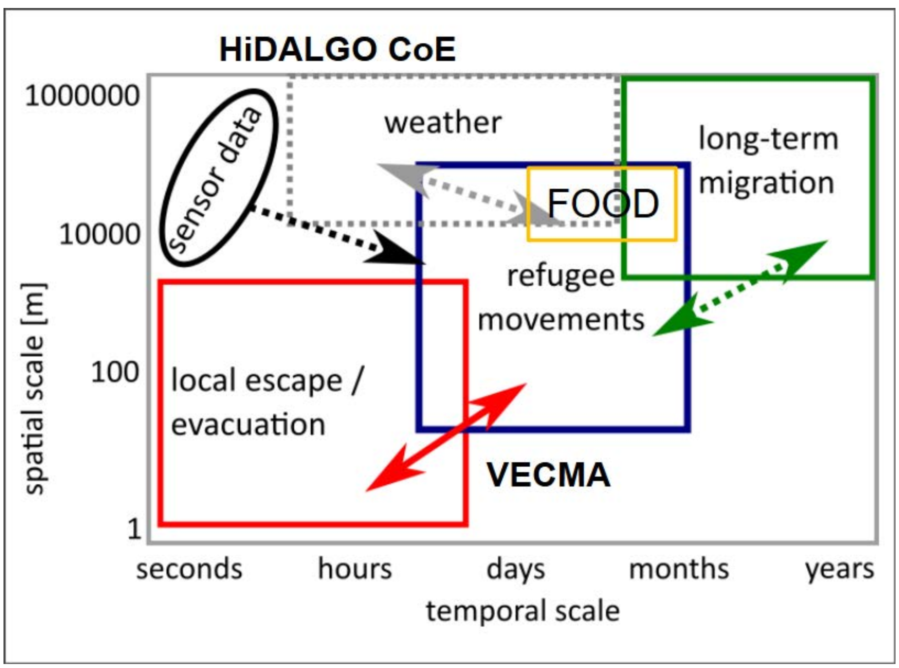

Figure 3: Envisioned model couplings for a full-featured hybridized model of forced migration, mapped to a scale separation map (Borgdorff et al. 2013). This image is an evolution of the diagram originally presented in Groen (2016).

Much of the current simulation literature focuses on the execution and analysis around a single simulation code, and this can be a limiting factor when re-using such works in a hybrid context. By accounting for all the major activities required to develop a simulation (e.g., as we have done in our simulation development approach (SDA) here (Suleimenova et al. 2017) we explicitly position the simulation code in the wider context of research, validation and simulation development. This wider awareness is useful, because one can easily observe whether the existing validation setting is preserved when integrating a second model, and to what extent the data sources and the problem definitions need to be modified. For instance, when combining a migration model with a conflict evolution model in a hybrid, one may be still be able to construct a simulation using the inputs for a migration model, but will have to define a new validation setting to validate the conflict evolution model against historical conflicts. In summary, to improving our ability to create hybrid simulations I suggest to: 1) Take an evolutionary prototyping approach to simulation building (Boehm et al. 1988), keeping individual models small and anticipating a frequent re-composition and re-definition of the hybrid problem, 2) keep careful track of the relevance of individual sub models to the overarching challenge, and be prepared to leave them out (or add other ones) if the need arises and 3) use simulation development approaches to position simulation codes in the context of research, validation and simulation development - this will help new researchers to use codes appropriately, and gives them a starting point from which they can evolve their own approach.

In the case of our migration simulations (see Figure 3), this is also the approach we apply. We initially developed an agent-based model of people escaping violent conflicts in search of safety abroad, and are currently prototyping models for e.g., food security and local escape. We fit these into our existing SDA, and will perform a two-fold validation. For instance, we will first check whether any food security mode [e.g., as first investigated in Vanhille et al. (2019)] produces outcomes that are close to the observed outcomes, and then check whether the incorporation of food security models provides a significant accuracy benefit to the application overall.

As a final remark, in terms of technical design and development I advocate a light-weight approach, but what that entails is largely in the eye of the beholder. Having coded in Python for 15 years, light-weight constitutes for me a slim program with 100-1000 lines of core logic while ready-made tools with abstraction layers and GUIs introduce needless overheads and restrictions in my case. But when I consult one of my students, or indeed some of my colleagues who have interacted in GUI environments for decades, the opposite may well be true for them. But whether your weapon of choice is code or click, one thing is certain: any coupling approach that works is likely to be simple and at least reasonably flexible. And fortunately for us, a lot of such approaches are now in existence, catering for the full spectrum of developer needs (Groen et al. 2019).

5 MODEL EXPLORATION OF AGENT-BASED (AND OTHER BLACK-BOX) MODELS AT HPC-SCALES (JONATHAN OZIK)

Advances in high performance agent-based modeling (ABM) have enabled the simulation of a variety of complex systems. By developing increasingly realistic models, intricate aspects of the modeled systems can be encapsulated, including agent activities and decision-making, agent interactions over social networks, demographic and geographic heterogeneity, and agent adaptation and learning. For ABMs to be trusted as research and policy-relevant electronic laboratories, it is necessary to robustly characterize their possible model behaviors, a process we have referred to as model exploration (ME), through heuristic methods. As ABMs have become more complex and their parameter spaces have grown in size, brute force methods for model exploration have necessarily given way to iterative and adaptive approaches, that strategically explore the large model behavior spaces, e.g., Genetic Algorithms (Holland 1992), Active Learning (Settles 2012; Binois et al. Ludkovski 2018) or sequential Approximate Bayesian Computing (ABC) (Beaumont 2010). Implementing dynamic model exploration at the requisite scale is a challenge, and this has led to a proliferation of smaller scale and ad hoc approaches due to the expertise impedance mismatch that exists between general computational science and the HPC realm. In many cases, this mismatch creates off-limit areas of research which are abandoned due to the perceived prohibitive time, effort, or computational expenditures required.

The Extreme-scale Model Exploration with Swift (EMEWS) framework (https://emews.github.io) is aimed at lowering the barriers for utilization of advanced, large-scale model exploration (Ozik et al. 2016). It does this by allowing the direct integration of multi-language ME algorithms to control the evolution of ME workflows that can run on the largest computing resources, while coordinating existing, native-code models (i.e., without the need to port or re-code models). The rise of artificial intelligence (AI) has been instrumental in driving the development of statistical and machine learning software that can be applied to ME, resulting in vibrant ecosystems of free and open source libraries that are continually added to and updated as research frontiers are expanded. EMEWS enables the use of these sophisticated heuristic algorithms (written in R or Python) to control large and complex ME workflows. Furthermore, since EMEWS is built on top of the Swift/T parallel scripting language (Wozniak et al. 2013), which is a multi-language-enabled workflow engine, analytics pipelines, e.g., for simulation data pre and post processing, are easily integrated. In Figure 4 we demonstrate an example of the high-level architecture of an EMEWS workflow. Here a Python-based GA algorithm that was developed with the use of the DEAP toolkit (Fortin et al. 2012) is the ME component and C++-based Repast HPC models (Collier and North 2013) are being dynamically launched and their outputs evaluated. We note that the Repast HPC models themselves can be distributed (via MPI) and therefore we exploit the available concurrency at multiple scales.

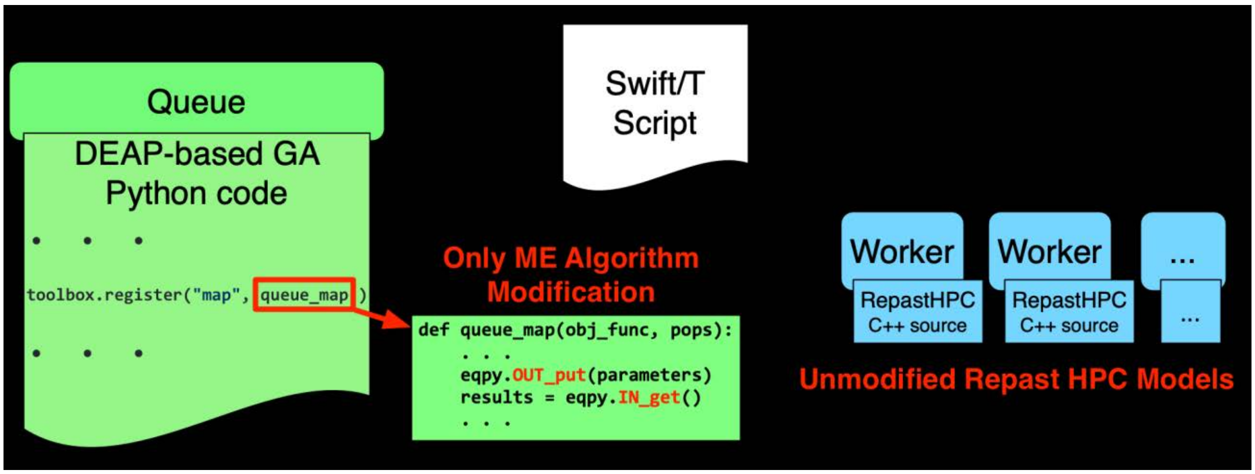

Figure 4: EMEWS workflow showing the Model Exploration (ME) components in green and Models components in blue for a workflow applying a Python-based GA algorithm to the C++-based Repast HPC model. The only ME algorithm modification for integrating with EMEWS is highlighted, while the Repast HPC model is unmodified.

We have applied EMEWS to ABM in many areas of research, including infectious disease modeling (Ozik et al. 2018), healthcare interventions (Kaligotla et al. 2018), and in silico cellular modeling for cancer (Ozik et al. 2018; Ozik et al. 2019). However, these large-scale ME methods are applicable to any black-box modeling paradigm and so we have applied EMEWS to other modeling approaches as well. These include microsimulation (Rutter et al. 2018) and large-scale deep learning (Wozniak et al. 2018a; Wozniak, et al. 2018b; Zaki et al. 2018).

An important thrust of our work is focused outward, i.e., in improving the capabilities and ease-of-use of the framework to enable the adoption of large-scale ME approaches across the computational modeling community. One aspect of this is our philosophy of integration through aggressive reuse of existing software and cutting-edge open-source tools. We are also creating user on-boarding resources such as immersive tutorials (https://www.mcs.anl.gov/~emews/tutorial/) and use-case repositories (https://github.com/emews/mela). Finally, recognizing that access to computational resources varies, we are developing methods for simple deployment of EMEWS workflows to cloud computing platforms.

6 SUMMARY

This paper presents views of four panelist, uncovering the range and scope of hybrid simulation in an era of big data and high performance computing. The paper itself focuses on the integration and coupling challenges in a range of contexts. The benefits of hybrid simulation with a single platform are clear, with integration support embedded. Complexities around scale are also presented – typically leading to distributed or high performance architecture choice. Near realistic representation of a domain is driving this requirement for exascale computational power. Software architecture that support this migration are discussed from both domain specific and technical perspectives. The scale and scope of near realism in modelling and simulation requires new methods of analyzing and utilizing data at different points in the workflow.

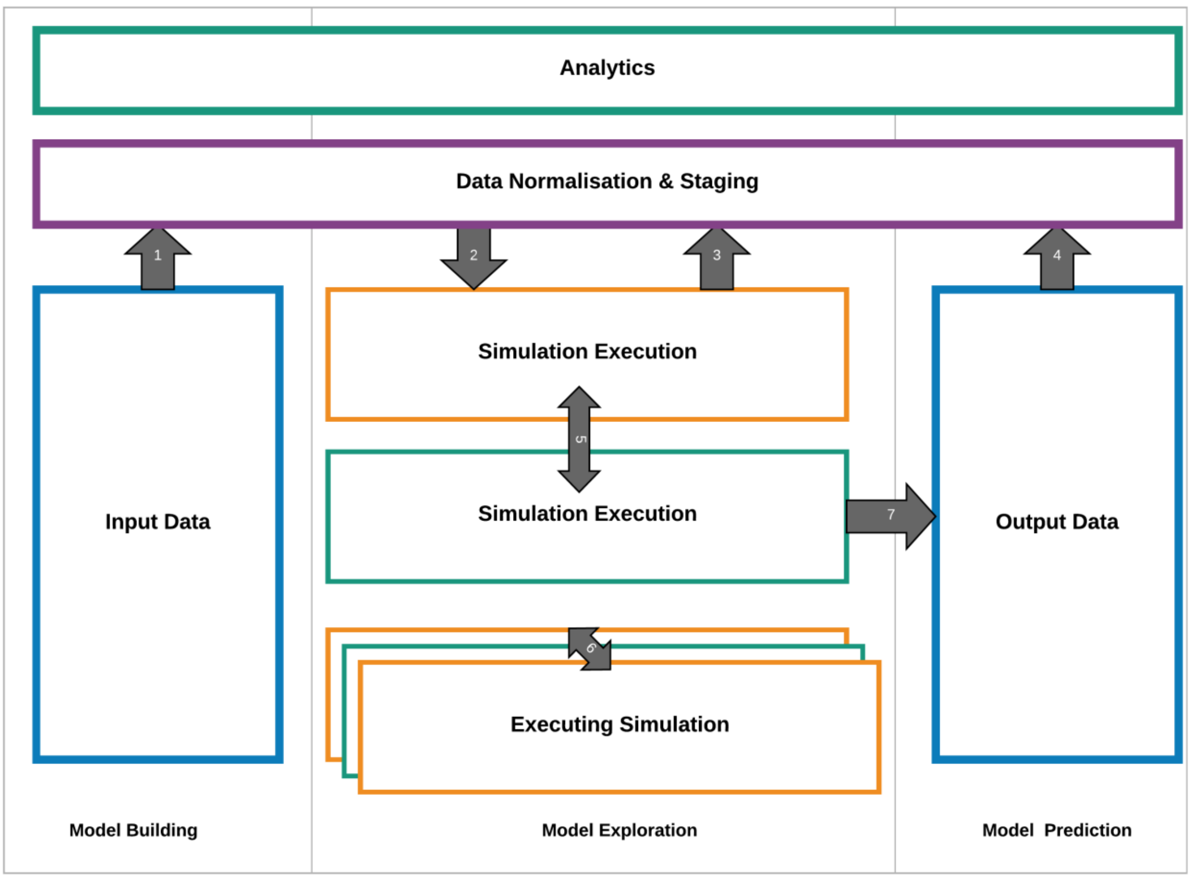

Figure 5: Analytics across a simulation and modeling workflow.

Figure 5 synthesizes the varied viewpoints presented into a hybrid simulation-analytics platform. Traditional analytical approaches that use empirical data have a long tradition in model building. Artificial Intelligence, machine learning and other statistical approaches (termed analytics in Figure 5) are able to process larger data sets in order to support a model building process (1), including heuristic rule induction. Consequently, once built and deployed, analytics can be used to explore the execution as it progresses and interim results as they are generated – directing the exploration in near real-time (2 & 3). Finally, analytics are able to process the substantial data resulting from simulations (7) and both present results to the user or scientist and also direct further exploration (or hybrid interaction between simulations – 5 & 6). Unsurprisingly, the normalization and staging of data underpins much of the looser coupling approaches (1-4), but also important within tighter coupling between executing simulations (5 & 6).

ACKNOWLEDGEMENTS

Ozik's material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357.

REFERENCES

Alowayyed, S., D. Groen, P. V. Coveney, and A. G. Hoekstra, 2017. “Multiscale Computing in the Exascale Era”. *Journal of Computational Science* 22(1):15-25.

Barga, R., V. Fontama, W. H. Tok, and L. Cabrera-Cordon. 2015. *Predictive Analytics with Microsoft Azure Machine Learning*. Berkely, CA: Apress.

Beaumont, M. A. 2010. “Approximate Bayesian Computation in Evolution and Ecology”. *Annual Review of Ecology, Evolution, and Systematics* 41(1):379–406.

Bergmann, S., N. Feldkamp, and S. Strassburger. 2017. "Emulation of Control Strategies through Machine Learning in Manufacturing Simulations". *Journal of Simulation* 11(1):38-50.

Binois, M., R. B. Gramacy, and M. Ludkovski. 2018. “Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments.” *Journal of Computational and Graphical Statistics* 27(4):808–21.

Boehm, B. W., 1988. “A Spiral Model of Software Development and Enhancement”. Computer 5(1):61-72.

Borgdorff, J., J. L. Falcone, E. Lorenz, C. Bona-Casas, B. Chopard, and A. G. Hoekstra, 2013. “Foundations of Distributed Multiscale Computing: Formalization, Specification, and Analysis”. *Journal of Parallel and Distributed Computing* 73(4):465-483.

Brandstein, A. and G. Horne. 1998. “Data Farming: A Meta-technique for Research in the 21st Century”. *Maneuver Warfare Science* 1998:93-99.

Brooks, R. J. and S. Robinson, 2000. *Simulation Studies: Key Stages and Processes*. Hampshire, UK: Palgrave Macmillan.

Chen, H., R. H. L. Chiang, and V. C. Storey. 2012. “Business Intelligence and Analytics: From Big Data to Big Impact”. *MIS Quarterly* 36(4):1165-1188.

Collier, N. and M. J. North. 2013. “Parallel Agent-Based Simulation with Repast for High Performance Computing”. *Simulation* 89(10):1215–35.

Davenport, T. H. 2006. “Competing on Analytics”. *Harvard Business Review 84*(1):98.

Eldabi, T., S. Brailsford., A. Djanatliev, M. Kunc, N. Mustafee, and A. F. Osorio, 2018. “Hybrid Simulation Challenges and Opportunities: A Life-cycle Approach”. In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1500-1514. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Feldkamp, N., S. Bergmann, and S. Strassburger. 2015. "Knowledge Discovery in Manufacturing Simulations". In *Proceedings of the 3rd ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, edited by

Feldkamp, N., S. Bergmann, S. Strassburger, and T. Schulze. 2016. "Knowledge Discovery in Simulation Data: A Case Study of a Gold Mining Facility". In *Proceedings of the 2016 Winter Simulation Conference,* edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick. 1607-1618. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Feldkamp, N., S. Bergmann, S. Strassburger, T. Schulze, P. Akondi, and M. Lemessi. 2017. "Knowledge Discovery in Simulation Data – a Case Study for a Backhoe Assembly Line". In *Proceedings of the 2017 Winter Simulation Conference,* edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page. 4456-4458. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Fortin, F-A., F-M. Rainville, M-A. Gardner, M. Parizeau, and C. Gagné. 2012. “DEAP: Evolutionary Algorithms Made Easy.” *Journal of Machine Learning Research* 13(1):2171–2175.

Groen, D. 2016. “Simulating Refugee Movements: Where Would You Go?”. *Procedia Computer Science* 80:2251-2255.

Groen, D., J. Knap., P. Neumann, D. Suleimenova., L. Veen,. and K. Leiter, 2019. “Mastering the Scales: A Survey on the Benefits of Multiscale Computing Software”. *Philosophical Transactions of the Royal Society A* 377(2142):p.20180147.

Harper, A. and N. Mustafee. 2019. “Proactive Service Recovery in Emergency Departments: A Hybrid Modelling Approach using Forecasting and Real-time Simulation”. In *Proceedings of the ACM SIGSIM Conference on Principles of Advanced Discrete*

Simulation (PADS) Conference, edited by Z. li, 201-204. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Holland, J. H. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, Mass: A Bradford Book.

Holmberg, Erik. 1941. "On the Clustering Tendencies among the Nebulae. II. a Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure." The Astrophysical Journal 94(3): 385-395.

Horne, G. E. and T. E. Meyer. 2005. "Data Farming: Discovering Surprise". In *Proceedings of the 2005 Winter Simulation Conference,* ", edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 1500-1514. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kaligotla, C., J. Ozik, N. Collier, C. M. Macal, S. Lindau, E. Abramsohn, and E. Huang. 2018. "Modeling an Information-based Community Health Intervention on the South Side of Chicago". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 2600-2611. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Keim, D. A., F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. 2008. "Visual Analytics: Scope and Challenges". In *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*, 1st ed, edited by Simoff, S., Böhlen, M.H. and Mazeika, A.. Berlin, Heidelberg: Springer.

Lycett, M., 2013. "'Datafication': Making Sense of (Big) Data in a Complex World." *European Journal of Information Systems* 22(4):381–386.

Makino, J. and H. Daisaka, 2012. "GRAPE-8--An Accelerator for Gravitational N-body Simulation with 20.5 Gflops/W Performance". In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (SC12)*, edited by H. Wasserman, 1-10. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Marshall D, L. Burgos-Liz, M. Ijzerman, W. Crown, W. Padula, P. Wong, K. Pasupathy, M. Higashi, and N. Osgood. 2015. "Selecting a Dynamic Simulation Modeling Method for Healthcare Delivery Research – Part 2." Report of the ISPOR Dynamic Simulation Modeling Emerging Good Practices Task Force. *Value in Health* 18(2):147-160.

Mustafee, N. and J. H. Powell, 2018. "Towards a Unifying Conceptual Representation of Hybrid Simulation and Hybrid Systems Modelling". In *Proceedings of the UK Operational Research Society Simulation Workshop (SW18)*, edited by A. Anagnostou and M. Fakhimi, 19-21. Birmingham, UK: UKORS.

Mustafee, N. and J. H. Powell. 2018. "From Hybrid Simulation to Hybrid Systems Modelling". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1430-1439. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Mustafee, N., J. H. Powell, and A. Harper. 2018. "RH-RT: A Data Analytics Framework for Reducing Wait-time at Emergency Departments and Centres for Urgent Care". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 100-110. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Onggo, B. S., N. Mustafee, A. Smart, A. A. Juan, and O, Molloy. 2018. "Symbiotic Simulation System: Hybrid Systems Model meets Big Data Analytics". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1358-1369. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ozik, J., N. Collier, J. M. Wozniak, C. M. Macal, C. Cockrell, S. H. Friedman, A. Ghaffarizadeh, R. Heiland, G. An, and P. Macklin. 2018. "High-Throughput Cancer Hypothesis Testing with an Integrated PhysiCell-EMEWS Workflow". *BMC Bioinformatics* 19(18):483.

Ozik, J., N. Collier, R. Heiland, G. An, and P. Macklin.. 2019. "Learning-Accelerated Discovery of Immune-Tumour Interactions". *Molecular Systems Design & Engineering*. 4(4), 747-760.

Ozik, J., N. Collier, J. M. Wozniak, and C. Spagnuolo. 2016. "From Desktop to Large-Scale Model Exploration with Swift/T." In *2016 Winter Simulation Conference (WSC).*

Ozik, J., N. Collier, J. M. Wozniak, C. M. Macal, and G. An. 2018. "Extreme-Scale Dynamic Exploration of a Distributed Agent-Based Model With the EMEWS Framework." *IEEE Transactions on Computational Social Systems* 5(3):884–95.

Powell, J. H. and N Mustafee. 2017. "Widening Requirements Capture with Soft Methods: An Investigation of Hybrid M&S Studies in Healthcare". *Journal of the Operational Research Society* 68(10):1211-1222.

Roemer, A. C. and S. Strassburger, S. 2019. "Hybrid System Modeling Approach for the Depiction of the Energy Consumption in Production Simulations". In *Proceedings of the 2019 Winter Simulation Conference,* edited by N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y-J. Son, To Appear. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rutter, C., J. Ozik, M. DeYoreo, and N. Collier. 2018. "Microsimulation Model Calibration Using Incremental Mixture Approximate Bayesian Computation". *ArXiv:1804.02090 [Stat]*, April. http://arxiv.org/abs/1804.02090.

Sanchez, S. and P. J. Sanchez. 2017. "Better Big Data via Data Farming Experiments". In *Advances in Modeling and Simulation, 1st edition,* edited by A. Tolk, J. Fowler, G. Shao and E. Yücesan, 159–179. Springer International Publishing, Cham.

Settles, B. 2012. "Active Learning". *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6(1):1–114.

Shao, G., S. Shin, and S. Jain. 2014. "Data Analytics Using Simulation for Smart Manufacturing". In *Proceedings of the 2014 Winter Simulation Conference,* edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 2192-2203. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Shaw, D. E., J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M Deneroff, R. O. Dror, A. Even, C. H. Fenton and A. Forte. 2014. "Anton 2: Raising the Bar for Performance and Programmability in a Special-purpose Molecular Dynamics Supercomputer". In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, edited by F. Johnson and K. Wohlever. 41-53. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Strassburger, S. 2001. "Distributed Simulation Based on the High Level Architecture in Civilian Application Domains". Society for Computer Simulation International, Ghent, Belgium.

Suleimenova, D., D. Bell, and D. Groen, 2017. "A Generalized Simulation Development Approach for Predicting Refugee Destinations. *Scientific Reports* 7(1):13377.

Taylor, S., S. Turner, S. Strassburger, and N. Mustafee. 2012. "Bridging The Gap: A Standards-Based Approach to OR/MS Distributed Simulation". *ACM Transactions on Modeling and Computer Simulation* 22(4):Article No. 18.

Vanhille, C., D. Suleimenova, and D. Groen, 2019, "A Coupled Food Security and Refugee Movement Model for the South Sudan Conflict". In Proceedings of the International Conference on Computational Science, edited by P.M.A. Sloot, 725-732. Springer, Cham.

Waller, M. A. and S. E. Fawcett. 2013. "Data Science, Predictive Analytics and Big Data: A Revolution that will Transform Supply Chain Design and Management". *Journal of Business Logistics* 34(2):77-84.

Wörrlein, B., S. Bergmann, N. Feldkamp, and S. Straßburger. 2019. "Deep-Learning-basierte Prognose von Stromverbrauch für die hybride Simulation". In Proceedings of the 18th ASIM Dedicated Conference on Simulation in Production and Logistics, Chemnitz, Germany, September 18-20, 2019. To Appear.

Wozniak, J. M., T. G. Armstrong, M. Wilde, D. S., Katz, E. Lusk, and I. T. Foster. 2013. "Swift/T: Large-Scale Application Composition via Distributed-Memory Dataflow Processing". In *Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, edited by P. Balaji. 95–102. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wozniak, J. M., P. E. Davis, T. Shu, J. Ozik, N. Collier, M. Parashar, I. Foster, T. Brettin, and R. Stevens. 2018a. "Scaling Deep Learning for Cancer with Advanced Workflow Storage Integration". In *Proceedings of the 2018 Fourth Workshop Machine Learning in HPC Environments. Dallas, TX, USA. 114-123.* Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wozniak, J. M., R. Jain, P. Balaprakash, J. Ozik, N. T. Collier, J. Bauer, F. Xia, T. Brettin, R. Stevens, J. Mohd-Yusof, C. Garcia Cardona, B. Van Essen, M. Baughman. 2018b. "CANDLE/Supervisor: A Workflow Framework for Machine Learning Applied to Cancer Research". *BMC Bioinformatics* 19(18):491.

Zaki, G., J. M. Wozniak, J. Ozik, N. Collier, T. Brettin, and R. Stevens. 2018. "Portable and Reusable Deep Learning Infrastructure with Containers to Accelerate Cancer Studies." In *Proceedings of the 2018 Fourth International Workshop on Extreme Scale Programming Models and Middleware (ESPM2),* edited by H. Subramoni, K. Schulz and D. Panda. 54-61. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIBLIOGRAPHIES

DAVID BELL is a Reader in the Department of Computer Science at Brunel University London where he carries out research into digital service design – modeling and simulating user interactions, decision making and behavioral change. Research domains have included healthcare, heritage and cyber security. Dr. Bell also spent 15 years working in the IT industry (primarily within Investment Banking) in roles from developer to technology director. He is the Education Chair for the ACM SIGSIM where he is involved in the promotion and support of doctoral simulation researchers. His email address is david.bell@brunel.ac.uk.

DEREK GROEN is a Lecturer in Simulation and Modeling at Brunel University London, and a visiting Lecturer at the Centre for Computational Science, University College London. He is principal investigator for Brunel in the EU-funded HiDALGO and VECMA HPC projects, and has published over 50 peer-reviewed papers. He specializes in multiscale computing, agent-based simulation, and high-performance computing. His email address is derek.groen@brunel.ac.uk

NAVONIL MUSTAFEE is Associate Professor of Operations Management and Analytics and the Deputy Director of The Centre for Simulation, Analytics and Modelling (CSAM) at the University of Exeter Business School, UK. He developed and led the 'Hybrid Simulation' track at the Winter Simulation Conference from 2014 to 2018. His email address is n.mustafee@exeter.ac.uk. His website is https://sites.google.com/site/navonilmustafee/.

JONATHAN OZIK is a Computational Scientist at Argonne National Laboratory and Senior Scientist in the Consortium for Advanced Science and Engineering at the University of Chicago. He develops applications of large-scale agent-based models, including models of infectious diseases, healthcare interventions, biological systems, water use and management, critical materials supply chains, and critical infrastructure. He also applies large-scale model exploration across modeling methods, including ABM, microsimulation and machine/deep learning. Dr. Ozik leads the Repast project (repast.github.io) for agent-based modeling toolkits and the Extreme-scale Model Exploration with Swift (EMEWS) framework for large-scale model exploration capabilities on high performance computing resources (emews.org). His email address is jozik@anl.gov.

STEFFEN STRASSBURGER is a professor at the Technische Universität Ilmenau and head of the Group for Information Technology in Production and Logistics. Previously he was head of the "Virtual Development" department at the Fraunhofer Institute in Magdeburg, Germany. His research interests include distributed simulation, automatic simulation model generation, and general interoperability topics within the digital factory and Industry 4.0 context. He is also an active member of the Simulation Interoperability Standards Organization (SISO). His web page can be found via http://www.tu-ilmenau.de/itpl. His email is steffen.strassburger@tu-ilmenau.de.