Final white paper: MIRACLE project (MIning Relationships Among variables in large datasets from CompLEx systems)

Dawn Cassandra Parker, Michael J. Barton, Tatiana Filatova, J. Gary Polhill, Xiongbing Jin, Allen Lee, Ju-Sung Lee, Kirsten Robinson (Wright), and Calvin Pritchard.

January 27, 2019

This report is written from the perspective of the senior institutional faculty at the collaborating organizations (Michael Barton and Marco Janssen (Arizona State University), Tatiana Filatova (University of Twente), Dawn Parker (University of Waterloo), and Gary Polhill (James Hutton Institute). Technical details of grant activities are outlined in detail in respective final grant reports, linked below. In this white paper, we seek to supplement those reports, by providing a history of our collaboration, an explanation of our motivation and goals for undertaking the project, a review of the project goals and anticipated challenges, and an assessment of the successes and lessons learned from the project.

Who are we?

We are all relatively early adopters of a social science simulation methodology called "agent-based modelling". Agent-based models (ABMs) are simulation models, most often implemented in computer code, which represent individual decision-making entities and their interactions. ABMs are most often used to represent complex systems, where analytical mathematics may not be practical or are not a useful representation of the dynamics of interest. While early ABM work occurred in the later part of the 20th century, ABM modelling increased in visibility and popularity through the early part of this century, with growth in fields such as coupled natural-human systems modelling growing exponentially (Matthews, Gilbert, Roach, Polhill, & Gotts, 2007; Janssen, 2017).

We were all part of that growth trajectory, and each of us has devoted a substantial part of their scholarship towards efforts to support ABM and help it develop scientifically mature methods. Janssen, Parker, Polhill met in 2001 at the seminal workshop on agent-based models of land-use and land-cover change (Parker, Berger, & Manson, 2002; Parker, Manson, Janssen, Hoffmann, & Deadman, 2003). They have continued to collaborate since. Barton and Janssen were founding members CoMSES Net (Network for Computational Modeling in Social and Ecological Sciences), an international scientific network, supported by the National Science Foundation, that provides a publication and archive platform for ABM code and other resources for modelers (Janssen, Alessa, Barton, Bergin, & Lee, 2008). Parker and Polhill have been actively involved in CoMSES Net institutional development and governance. Filatova has an active collaboration history with

Parker and Polhill. Further, as an officer of the Environmental Modelling and Software Association, she has been instrumental in establishing agent-based modeling research within the association's conferences and publications (Cite spatial ABM EMS special issue?).

Why did we undertake this project?

Throughout our careers, we have all been actively involved in identifying barriers to the scientific evolution of ABM, and in trying to find solutions. Since 2001, we have been discussing the problem of code sharing. Most scholars who develop ABMs, either alone or in research groups, create their own models from scratch, without reusing existing code. We have always known that this approach is inefficient and can inhibit scientific communication, transparency, and replicability. Even with the CoMSES Net archive available, code re-use has been minimal. Even when a researcher might try to replicate others' work, key details on workflow and dependencies might be missing. We have also recognized that much of the scientific scepticism regarding ABM stems from difficulties in completing global sensitivity analysis on model output. Unlike a straightforward, analytical equilibrium condition, one cannot know with ease or certainty what causal relationships existing within output data, or how robust these relationships might be.

Our patterns of collaboration are somewhat more predictable by now than analysis of output data from complex systems. One or two of us identifies a key issue or methodological challenge and discusses this issue in invited presentations for various international audiences, as well as among ourselves. We ultimately move beyond speculation and rhetoric and tackle the issue via grant proposals, a publication stream, or both. Joint work between Parker, Polhill, and Filatova on agent-based land market models has followed this trajectory, as have Barton and Janssen's efforts to develop and expand the ABM model library. Parker, Pohill, and other ABM colleagues have also had long-standing discussions regarding model ontology and model communication protocols (Grimm et al., 2010; Parker, Brown, Polhill, Manson, & Deadman, 2008).

The challenges presented by analysis of ABM output data was a theme of Parker's thinking for some time before this collaborative grant. On the one hand, the problem of model sharing remained unsolved. Lively discussions in CoMSES Net planning meetings focused on a key related issue—should every scholar doing ABM know how to program their own models? While the position of many founding members was "yes," Parker envisioned a world where this barrier to entry for the field was dismantled, as it has been for regression methods. In an unrelated note, the challenge of how to conduct global sensitivity analysis lurked behind the scenes in her work, as well as it seemed, in the work of others. She and Filatova applied standard regression techniques to model output to try to reveal testable hypothesis, knowing that the statistical assumptions of such models did not hold for complex systems output data. While colleagues were exploring alternatives, there was no

clear answer to the problem—even when asking scholars from other fields, from chemistry to ecology to physics.

Putting all of these pieces together, the Digging into Data call appeared to offer a reasonable opportunity for our group to take on this new challenge—provided we could interest the review panel in 1) the proposition that simulation outputs represented a novel and challenging new source of "big" social science data and 2) project support that was essentially support to develop computational social science infrastructure. (Such funding is virtually non-existent). In our favour, we also sought to investigate clear scientific questions as part of the research. What methods were scholars using to analyze their ABM outputs? What worked and did not work? What methods might be shared among researchers, and what new insights might be gained?

As with all interdisciplinary multi-national grant applications, proposal preparation was intense and complex. In Parker's view, the availability of dedicated personnel support at the University of Waterloo was essential to the grant's successful preparation. Xiongbing Jin, a post-doc at UW, was highly interested in the topic and highly motivated to obtain additional funding support to allow him to pursue the project. Meagan Bean, administrative coordinator for the Waterloo Institute for Complexity and Innovation, provided key administrative support to keep preparation on track.

We got the grant. Yeah!

The grant's main achievements are outlined in the attached annual reports. Here, using the risk table and further elaboration from our perspectives, we outline the challenges faced (anticipated and not), how we addressed them, and we also highlight what we feel were successful strategies for this model of grant.

Table 2. Risk table (modified from grant proposal)

Risk	*Likelihood*	*Impact*	*What actually happened*
Failure to recruit post-docs	Medium	High	We all managed to recruit post-docs, but at different times with different start dates, creating a major challenge for coordination of project work. *The problem was exacerbated by different funding start and end dates at the different universities.* As a result, some post-docs ended their term before other began.
Low level interest in workshops	Medium	Medium	Interest and attendance at workshops were very good. As part of our strategy to coordinate grant activities with conferences on related themes, we were able to hold formal workshops with both IEMSS and the Social Simulation conferences, which were advertised as part of conference activities. *Access to conferences with flexible session structure helped facilitate grant activities.*
No demo	High	Low	Our first two workshops targeted information

available for workshops			gathering—what are members of the community now doing, and what tools would help them? A discussion format was fine. For our final capstone workshop, we had a downloadable interactive demo prepared, and we even prepared logins for registered participants. Then, the eduroam internet at the conference venue failed. We had a video back-up demonstration prepared, and did manage to have a very informative discussion with participants in spite of the failure. However, we probably missed an opportunity to have participants adopt the prototype.
Low level uptake by community	Medium	Medium	While we have disseminated the prototype via github and an overview publication in EMS, others have not adopted the prototype. However, we do not see that as a failure. The prototype now serves a basis for new tools being developed by CoMSES, which have a very high likelihood of use and adoption (Barton et al 2018).

Without a doubt, the most serious challenges to the project occurred because of differing start and end dates to funding at each location, coupled with different dates for post-docs and senior research personnel. Much of the work for the project—software development--was planned to occur in coordinated, concurrent fashion. This was not possible—and thus, the model for software development had to be reworked. Work that was originally planned to occur in early stages at Arizona State University (ASU) occurred instead at the University of Waterloo. To facilitate this, a PhD student (Robinson) who had substantive prior experience managing software development was moved into a planning and project coordination role. When the personnel (hired as permanent research staff at that point) at ASU came on line, they took over major software development tasks. As a result, the software prototype was not finished in time for the extensive cross-site comparisons that we planned. However, the ASU team managed to capture substantial funding resources from NSF, which have allowed them to build on lessons learned from the software prototype to build a next-generation version.

In spite of these challenges, we see many points of success from the collaboration:

- We have placed overview publications in key outlets:
    - A JASSS article (Lee et al., 2015) offers an overview of state-of-the-art approaches in analysing and reporting ABM outputs, while highlighting challenges and outstanding issues. The article's continuing relevance is highlighted by the fact that it currently ranks among the most viewed articles in the journal over the past 8 weeks (ranked 7th), despite its publication over three years ago.
    - An EMS publication (Jin et al., 2017) gives an overview of the prototype with links to the user manual, project website and code archive
- Highly qualified personnel who received training through the grant have moved into permanent research positions:

    - Xiongbing Jin, post-doc at UW, is now Senior Specialist, Modelling at Canada Mortgage and Housing Corporation.
    - Ju-Sung Lee, post-doc at the University of Twente, is now Assistant Professor in the Department of Media and Communication at Erasmus University, Rotterdam
    - Calvin Pritchard, former UW undergraduate co-op student, is now an Academic Professional Researcher at ASU.
- CoMSES has obtained additional funding through the NSF Big Data Spoke for next-stage development of the MIRACLE tools as part of an integrated platform to support all stages of computational models of human-environment interactions.
- The UW team, in collaboration with the ASU team, have obtained a Compute Canada resource allocation grant to support further development of the new CoMSES tools.
- Individual country teams accomplished their main project goals, given the responsibility and workflow modifications.
- International collaborative networks were substantively strengthened. The project brought together teams who were well connected through bilateral collaboration networks—UW and ASU, UW and Scotland, UW and Twente, Scotland and Twente. The project created a successful quadrilateral collaboration network. In particular, the nascent links between ASU and Scotland/Twente have been substantially strengthened.
- The resulting project, we think is an exemplar example of social-science computing. Products, presentations, publications, and tools contribute to improved scientific methodology but also represent forward steps in the computer science realm (for instance, cloud-based interactive tools, metadata standards for both workflow and output data).

We do have some suggestions for program modifications:

- Allow for larger-resourced grants with a longer time frame, or for application for follow-up funds for projects off to a successful start;
- Ensure concurrent funding start and end dates for projects, with sufficient start-up time to hire personnel and accomplish pre-project strategic coordination and planning.

We also have suggestions for teams considering applications to similar programs:

- The short time frames and relatively small budgets of programs like this limit the collaborative work that can realistically be accomplished. Thus collaborative work goals should focus on comparison of work already in progress, or development of a first-stage prototype.
- Given this point, be sure to have project objectives that can be accomplished largely independently at partner institutions, with testing and feedback from other partners on a flexible schedule.

- Our strategy of planning project meetings and outreach sessions at professional conferences worked very well. It allowed us to conserve financial resources, but also to maintain a high profile for the project.
- Be adaptive—especially with personnel and their planned roles. Our flexibility to adapt roles and responsibilities allowed us to develop our software prototype in spite of obstacles.

From a scientific standpoint, our experience in this grant highlighted that work on the topic, in all fields, is in very early stages:

- Current analysis methods in use were even more simplistic than anticipated; overview article was an important contribution.
- In conversations with scholars from many other disciplines spanning social to natural science, it's clear that the issue of appropriate analysis algorithms remains an open and unsolved problem. New AI approaches may be appropriate for forecasting for complex data, but they do not offer insights into system structure, causal relationships, and parameter sensitivity.
- This grant gave us a unique opportunity to have these conversations, within and outside our fields. We are highly appreciative of that sort of opportunity, and we encourage funding agencies to continue to participate in such collaborative funding models, especially around cutting-edge, frontier scientific topics.

References

Barton, C.M., Lee, A., Janssen, M.A., Pritchard, C., Buetow, K.H. (2018). Discoverable, Accessible, Documented, Reusable, Interoperable: the Imperative to Publish Code in Modern Science. Presented at the Paper presented at the American Geophysical Meeting Fall Meeting, Washington, DC.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, *221*, 2760–2768. https://doi.org/10.1016/j.ecolmodel.2010.08.019

Janssen, M. A., Alessa, L. N., Barton, M., Bergin, S., & Lee, A. (2008). Towards a community framework for agent-based modelling. *Journal of Artificial Societies and Social Simulation*, *11*(2), 6.

Janssen, M.A. (2017). The Practice of Archiving Model Code of Agent-Based Models. Journal of Artificial Societies and Social Simulation 20. https://doi.org/10.18564/jasss.3317

Jin, X., Robinson, K., Lee, A., Polhill, J. G., Pritchard, C., & Parker, D. C. (2017). A prototype cloud-based reproducible data analysis and visualization platform for outputs of agent-based models. *Environmental Modelling & Software*, *96*, 172–180.

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., … Parker, D. C. (2015). The complexities of agent-based modeling

output analysis. *Journal of Artificial Societies and Social Simulation*, *18*(4), 4.

Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G., & Gotts, N. M. (2007). Agent-based land-use models: a review of applications. *Landscape Ecology*, *22*(10), 1447–1459.

Parker, D. C., Berger, T., & Manson, S. M. (2002). Agent-Based Models of Land-Use/Land-Cover Change: Report and Review of an International Workshop. Bloomington, In : LUCC Focus 1 Publication 6.

Parker, D. C., Brown, D., Polhill, J. G., Manson, S. M., & Deadman, P. (2008). Illustrating a new 'conceptual design pattern' for agent-based models and land use via five case studies: the MR POTATOHEAD framework. (A. L. Paredes & C. H. Iglesias, Eds.), *Agent-Based Modelleling in Natural Resource Management*. Valladolid, Spain: Universidad de Valladolid.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review. *Annals of the Association of American Geographers*, *93*(2), 314–337. https://doi.org/10.1111/1467-8306.9302004

Complete bibliography of project outputs

Publications (total = 6)

Lee, J.-S., & Filatova, T. (2017). A Network Analytic Approach to Investigating a Land-Use Change Agent-Based Model. In W.Jager (Ed.), *Advances in Social Simulation 2017* (pp. 231–240). Springer: Heidelberg.

Jin, X., Robinson, K., Lee, A., Polhill, J. G., Pritchard, C., & Parker, D. C. (2017). A prototype cloud-based reproducible data analysis and visualization platform for outputs of agent-based models. *Environmental Modelling & Software*, *96*, 172–180.

Lee, J.-S., & Filatova, T. Agent-Based Model Outcome Reducibility as an Indicator of Sensitivity (under review). *Ecological Complexity*.

Polhill, G., Milazzo, L., Dawson, T., Gimona, A., & Parker, D. (2017). Lessons learned replicating the analysis of outputs from a social simulation of biodiversity incentivisation. In *Advances in Social Simulation 2015* (pp. 355–365). Springer.

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., … Parker, D. C. (2015). The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation*, *18*(4), 4.

Polhill, J. G., Dawson, T. D., Parker, D. C., Jin, X., Robinson, K., Filatova, T., … Edwards,

E. (2014). Towards metadata standards for sharing simulation outputs. In Miguel, Amblard, Barceló, & Madella (Eds.), *Advances in Computational Social Science and Social Simulation*

Conference presentations (total = 17)

Lee, J.-S., & Filatova, T. (2015). A Network Analytic Approach to Investigating a Land-Use Change Agent-Based Model. In *Social Simulation Conference 2015*. Groningen, The Netherlands. 14-18 September 2015.

Polhill, G. (2014). Analysing output data from an agent-based model of biodiversity incentive schemes. In *Workshop G2 (Analyzing and Synthesizing Results from Complex Socio-ecosystem Models with High-dimensional Input, Parameter and Output Spaces), 7th International Congress on Environmental Modelling and Software*. San Diego, California, USA. 15-19 June 2014.

Parker, D. (2014). Hoping for a MIRACLE: Mining Relationships Among Variables in Large Datasets from Complex Systems (Digging into Data research initiative). In *SSHRC Western "Big Data in the Social Sciences" Conference*. London, Ontario, Canada. March 15, 2014.

Filatova, T., & Lee, J.-S. (2015). Housing markets and climate change: non-linearities and abrupt crashes from the bottom-up. In *Conference on Complex Systems*. Tempe, Arizona, USA. 30 September 2015.

Filatova, T., Lee, J.-S., & Naimi, B. (2015). Housing markets and economic choices under uncertainty: exploring non-linear market dynamics in a spatial agent-based model. In *The 20th Workshop on the Economic Science with Heterogeneous Interacting Agents (WEHIA)*. Sophia Antipolis. 21-23 May 2015.

Parker, D., Filatova, T., Barton, M., Dawson, T., Polhill, G., Milazzo, L., … Pritchard, C. (2016). MIRACLE. In *Digging into Data Challenge Round 3 End of Project Conference*. The Lighthouse, Glasgow, UK. 27-28 January 2016.

Polhill, G., Milazzo, L., Parker, D., Jin, X., Pritchard, C., Lee, J.-S., & Salt, D. (2015). MIRACLE simulation outputs metadata specification version 1.1.6. In *Technical Report to MIRACLE project team*.

Jin, X., Robinson, K., Lee, A., Polhill, G., Pritchard, C., & Parker, D. (2016). MIRACLE: A Prototype Cloud-Based Reproducible Data Analysis and Visualization Platform for Outputs of Agent-Based Models. In D. Brown, E.-K. Kim, L. Perez, & R. Sengupta (Eds.), *Workshop at GIScience. Rethinking the ABCs: Agent-Based*

Models and Complexity Science in the age of Big Data, CyberGIS, and Sensor Networks. Montreal, Canada. 27 September 2016.

Filatova, T., Horstman, E., Hulscher, S. J. M. H., Lee, J.-S., & de Waard, J. (2016). Modelling adaptive behaviour in spatial agent-based models: coastal cities and climate change. In *8th International Congress on Environmental Modelling and Software (iEMSs)*. Toulouse, France. July 2016.

Dawson, T., Milazzo, L., & Polhill, G. (2016). Modelling Dynamic Social-ecological Systems: Towards an Integrated, Multiscale Approach. In *International Environmental Modelling and Software Society, iEMSs 2016*. Toulouse, France. July 2016.

Lee, J.-S., Filatova, T., & Voinov, A. (2014). Presentation "Analysing and synthesising results from agent-based models of complex socio-ecosystems" at Workshop G2 (Analyzing and Synthesizing Results from Complex Socio-ecosystem Models with High-dimensional Input, Parameter and Output Spaces). In *7th International Congress on Environmental Modelling and Software*. San Diego, California, USA. 15-19 June 2014.

Filatova, T. (2017). Spatial agent-based models: introducing individual interacting actors in environmental models. In *CSDMS Annual Meeting 2017: Modeling Coupled Earth and Human Systems - The Dynamic Duo*. University of Colorado, Boulder, USA. May 23-25, 2017.

Parker, D., Barton, M., Dawson, T., Filatova, T., Jin, J. L., Polhill, G., … Voinov, A. (2015). The MIRACLE project: A community library to archive and document analysis methods for output from agent-based models. In *Association of American Geographers Annual Meeting*. Chicago, USA. 21-25 April 2015.

Parker, D., Barton, M., Dawson, T., Filatova, T., Jin, X., Lee, A., … Voinov, A. (2015). The MIRACLE project: cyberinfrastructure for visualizing model outputs. In *Satellite Session on Computational Transparency in Modeling Complex systems, Conference on Complex Systems*. Tempe, Arizona, USA. 30 September 2015

Parker, D., Barton, M., Dawson, T., Filatova, T., Jin, X., Lee, A., … Voinov, A. (2015). The MIRACLE project: Tools and analysis methods for output from agent-based models of coupled human-natural systems. In *The 2015 AMMCS-CAIMS Congress*. Waterloo, Ontario, Canada. 7-12 June 2015.

Parker, D., Polhill, G., & Lee, A. (2016). The MIRACLE prototype: A hands-on demo of a new tool for visualization, analysis, and workflow management for agent-based models of socioeconomic systems. In *8th International Congress on Environmental Modelling and Software (iEMSs)*. Toulouse, France. July 2016.

Polhill, J. G., Dawson, T. D., Parker, D. C., Jin, X., Robinson, K., Filatova, T., … Edwards, E. (2014). Towards metadata standards for sharing simulation outputs. In Miguel, Amblard, Barceló, & Madella (Eds.), *Advances in Computational Social Science and Social Simulation (ESSA 2014)*. Barcelona, Spain: Autònoma University of Barcelona. 1-5 September 2014.

Barton, C.M., Lee, A., Janssen, M.A., Pritchard, C., Buetow, K.H. (2018). Discoverable, Accessible, Documented, Reusable, Interoperable: the Imperative to Publish Code in Modern Science. Presented at the Paper presented at the American Geophysical Fall Meeting, Washington, DC.

Workshops (total = 3)

Workshop Organization by D.C. Parker, J.G. Polhill, T.Filatova, M. Barton and T. Dawson: "Analyzing and Synthesizing Results from Complex Socio-ecosystem Models with High-dimensional Input, Parameter and Output Spaces" at the 7th International Congress on Environmental Modelling and Software (iEMSs), San Diego, CA, June 15-19, 2014. See workshop G2 at http://www.iemss.org/sites/iemss2014/Session%20and%20Workshop%20Proposals%20Final.pdf

Workshop Organization by J.G. Polhill and T. Filatova: "Metadata for Social Simulation Outputs" at the Social Simulation Conference 2014, Barcelona, Spain, September 1-5 2014. Web: http://www.essa2014.eu/SSC2014_Program.html

Satellite Session on Computational Transparency in Modeling Complex systems, Conference on Complex Systems, Tempe, Arizona, 30 September 2015. Contributors D.C. Parker and T.Filatova.