A Computational Pipeline for Modeling and Predicting Wildfire Behavior

Nuno Fachada[a]
Lusófona University, COPELABS, Campo Grande, 376, Lisboa, Portugal

Keywords: Agent-based Modeling, High-performance Computing, Computational Intelligence, Verification and Validation, Wildfires.

Abstract: Wildfires constitute a major socioeconomic burden. While a number of scientific and technological methods have been used for predicting and mitigating wildfires, this is still an open problem. In turn, agent-based modeling is a modeling approach where each entity of the system being modeled is represented as an independent decision-making agent. It is a useful technique for studying systems that can be modeled in terms of interactions between individual components. Consequently, it is an interesting methodology for modeling wildfire behavior. In this position paper, we propose a complete computational pipeline for modeling and predicting wildfire behavior by leveraging agent-based modeling, among other techniques. This project is to be developed in collaboration with scientific and civil stakeholders, and should produce an open decision support system easily extendable by stakeholders and other interested parties.

1 INTRODUCTION

Wildfires represent a major socioeconomic burden in affected regions, often resulting in widespread devastation. This is particularly true in the case of Portugal, which has the highest density of wildfire ignitions among southern European countries (Catry et al., 2010). While several scientific and technological approaches have been used for understanding, predicting and mitigating the problem (Collins et al., 2013), recent tragic wildfires show this is very much an open problem (Gómez-González et al., 2018).

Agent-based modeling (ABM) is a modeling approach where each entity of the system being modeled is represented as an independent decision-making agent. It is a useful technique for exploring systems that can be modeled in terms of interactions between individual components (Fachada et al., 2015), and is thus a good fit for modeling the behavior of wildfires (Niazi et al., 2010).

This position paper proposes a scientific project to study wildfire propagation by developing a complete agent-based modeling and simulation pipeline. It aims to advance the state of the art by simultaneously offering the following novel specifications: i) simulations should be able to run on commodity hardware; ii) models should be retrainable with new data sources, as these become available; iii) models should work robustly with missing data; and, iv) the pipeline should be a template for others to build upon.

The project is to be developed in collaboration with several stakeholders, namely scientific and civil authorities. Its main output will be a wildfire modeling and simulation pipeline with two applications in mind: i) act as a decision support system for civil protection stakeholders; and, ii) rather than a final product, serve as a platform to be extended by interested parties.

2 BACKGROUND

Agent-based modeling (ABM) is a bottom-up modeling approach where each entity of the system being modeled is represented as an independent decision-making agent. It is a useful technique for simulating and exploring systems that can be modeled in terms of interactions between individual components, such as cell cultures, military units in a battlefield or epidemiology scenarios (Fachada et al., 2015). ABM shares characteristics with, and adds capabilities to, discrete-event simulation, system dynamics and Monte Carlo methods (Macal, 2016), and in a broad sense, generalizes techniques such as cellular automata (CA) and network models (Trucchia et al., 2020). ABM is thus a good fit for modeling the behavior of wildfires (Yassemi et al., 2008; Widyastuti et al., 2020). Al-

---

[a] https://orcid.org/0000-0002-8487-5837

Fachada, N.
A Computational Pipeline for Modeling and Predicting Wildfire Behavior.
DOI: 10.5220/0011073900003197
In *Proceedings of the 7th International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2022)*, pages 79-84
ISBN: 978-989-758-565-4; ISSN: 2184-5034

though a few modeling approaches have been used to characterize Portugal's wildfires (Catry et al., 2010; Collins et al., 2013), the use of ABM is relatively uncommon given the socioeconomic impacts of the problem (Gómez-González et al., 2018).

Large scale emergent behavior in ABMs is population sensitive, and is thus desirable that the number of agents in a simulation is able to reflect the reality of the system being modeled (Fachada et al., 2017c). Additionally, stochastic models in general, and ABMs in particular, usually require the exploration of large parameter spaces (Hill et al., 2013). Consequently, simulating realistic models can be computationally demanding (Fachada and Rosa, 2017), often requiring parallel and/or distributed implementation approaches in order to leverage the high core count in modern processors (Xiao et al., 2019). This has indeed been the case for a number of wildfire ABMs (Smith et al., 2016).

While parallelizing ABMs is in itself a complex task, other difficulties exist. ABMs are very sensitive to implementation details, and the impact that seemingly unimportant aspects such as data structures, algorithms or order of events can have on results is tremendous (Wilensky and Rand, 2007; Fachada et al., 2017b). By definition, parallelization requires changes in many of these aspects. Furthermore, the partitioning of pseudo-random number generator (PRNG) streams in parallel simulations may bring problems such as hidden correlations or overlaps in different substreams, potentially invalidating simulation results (Hill et al., 2013).

In any case, verification and validation (V&V) is crucial for ABMs used in making real-world decisions (David et al., 2017). This is especially the case for complex dynamical systems in general (Williams, 2018), and wildfires in particular (Niazi et al., 2010; Achtemeier, 2013). Knowledge gaps are often uncovered by the V&V iterative process (David et al., 2017; Fachada et al., 2017b; Fachada et al., 2016), notably when integrating external data (e.g., satellite data (Cardil et al., 2019; Han et al., 2021)) in the model—a requirement for making accurate fire propagation forecasts. In such cases, computational intelligence techniques can be used to fill such gaps (Pathak et al., 2018; Fernandes et al., 2019; Fernandes et al., 2020) and aid in data integration (Ntinas et al., 2017; Denham et al., 2012).

Here, we propose a scientific project to study wildfire propagation by developing a computational simulation pipeline which includes, but is not to limited to, parallel high-performance ABMs, V&V tools, machine learning (ML) techniques and data fusion methods, with the goal of successful V&V of models against real world wildfire propagation scenarios in Portugal.

3 RESEARCH PLAN AND METHODS

In this project, we propose to develop a complete wildfire simulation pipeline, with the aim of addressing the scientific question of whether it is possible to accurately and efficiently predict the propagation of such fires using computational modeling and computational intelligence methods. The project has the following novel specifications: i) model simulation and visualization should be able to run on commodity hardware; ii) models should be easily retrainable with new data sources, as these become available; iii) models should work robustly in the absence of one or more data streams (i.e., the best possible forecast should be generated given the available data); and, iv) the pipeline should be easily extendable by third parties, i.e., the project aims to provide a template for others to build upon.

Collaborations are essential for a project of this type. Concerning basic research, collaboration will be sought with IPMA (Portuguese Institute for the Sea and Atmosphere) and possibly NOAA (National Oceanic and Atmospheric Administration). For applied research and field tests, cooperation with the ANPC (Portuguese Civil Protection Authority) is critical.

The project is divided into the following objectives/tasks:

1. Parallel ABM for conjectural wildfire simulations (18 months). Develop a parallel ABM framework aimed at wildfire simulation with multiple layers, namely, but not limited to, CA, mobile agents and network/graph layers. Framework should be highly performant and capable of incorporating different types of data (e.g., GIS for geographic features, GPS for location of wildfire response means, drone video fire tracking (Costa et al., 2021)), though not able to directly account for external data sources. Tentative models should broadly predict fire behavior given initial conditions in some scenarios, allowing users to pose "what-if" questions, but are not fit for actual decision making. The following questions will have to be answered:
    - What is the best shared-memory architecture to run simulations on these models? Given highly heterogeneous model components, are GPUs a feasible target architecture, or do multicore

  CPUs offer a more viable prospect in terms of man–hour/performance balance? Will it be necessary to use a distributed memory architecture (e.g., multiple computers, supercomputers) to run these models?
  - Given the different types of data to be accounted for, how and where should these be integrated in the various model layers?
  - Does the massive partition of models (and consequently, of PRNGs), for the purpose of concurrent simulations, skew simulation results one would obtain with serial executions (i.e., with no PRNG partitioning)? If so, what strategies can be employed to minimize the issue?
2. Data integration and offline model training (18 months). Integrate publicly available data in the framework, namely from satellites such as SMAP (moisture), Suomi NPP (visible and infrared imaging), Meteosat (visible and infrared imaging), Sentinel-2 (Han et al., 2021) or the various sources available from LSA SAF, such as land surface temperature, wildfire monitoring, vegetation and land cover. Use ML techniques to train models using historical data (from the same sources), focusing on a limited number of wildfire-prone patches of land in Portugal. Models should be data-driven, retrodictively valid and robust in their forecasts when not all of the driving data is available. Questions to be answered:
   - It is possible to automate the process of data fusion? What types of data transformations are required?
   - What are the most appropriate ML methods for offline model training? Is it necessary to develop custom ML approaches?
   - How is model robustness affected when suppressing different types of driving data?
3. Automatic model optimization (12 months). An automated V&V procedure is to be implemented. Models should be tuned using stochastic optimization methods in the presence of new information. Framework improvements based on basic research stakeholders feedback. Models are not yet expected to be used for decision making in the field. Questions to be answered:
   - It is possible to consolidate the V&V process in the overall modeling and simulation pipeline?
   - What kind of stochastic optimization methods work best for online model tuning?
4. Field trials (12 months). Simulations tested during wildfires in the modeled land patches in the presence of project stakeholders (IPMA, ANPC, others). Further framework improvements based on stakeholders feedback. The following questions will have to be answered:
   - Can the models accurately and efficiently predict the propagation of wildfires? This is essentially the main scientific question of the project.
   - Can the framework be integrated in the stakeholders workflow (e.g., ANPC) as a decision support system?
   - Can stakeholders such as IPMA build upon the developed framework?

The proposed plan also accommodates 12 months to explore innovations in the field (e.g., new data sources), late collaborations and documentation write-up.

4 EXPECTED OUTCOMES

The main output of this project will be a wildfire modeling and simulation pipeline with two applications in mind: i) predict wildfire propagation in the modeled land patches of Portuguese territory, acting as a decision support system for civil protection stakeholders; and, ii) rather than a final product, serve as a platform to be extended by all interested parties.

This is a unique opportunity for the author to consolidate his academic experience (discussed in detail in Section 6) into a project with potentially wide societal and economical benefits. In the first 18 months (task 1 of the project), we plan to publish three journal articles, namely:

1. “The influence of pseudo-random number generators on parallel agent-based simulations”, Computer Physics Communications.
2. “High-performance implementation of a wildfire agent-based model with OpenCL”, IEEE Transactions on Parallel and Distributed Systems.
3. “Data fusion challenges in a multilayered agent-based model of wildfire propagation”, International Journal of Image and Data Fusion.

While it is difficult to predict exact titles of the papers to be written after this period, we can summarize some of the potential contributions for tasks 2 and 3:

1. Automating the process of data fusion in wildfire ABM.
2. Proposing a clear V&V framework for wildfire ABMs in the context of a well defined modeling and simulation pipeline.
3. Effectively use such models for predicting wildfire outcomes under different conditions.

Scientific contributions for the final stages of the project will greatly depend on forged collaborations and stakeholder feedback.

5 SPIN-OFF RESEARCH PROJECTS

This project has considerable potential beyond wildfire forecast in Portugal—i.e., it can potentially be reused for other regions and countries—and indeed beyond wildfire forecast, period. The most obvious benefit, as already highlighted in the previous sections, is the potential for advancing fundamental knowledge in the fields touched by this project, from ML and ABM to parallel computing and multivariate data fusion. A second and less obvious benefit consists of taking the experience gained with wildfire modeling and apply it to other important societal problems such as modeling refugee migration patterns, metabolic cellular networks and pandemic outbreaks—such as the one we are currently experiencing. In our opinion, this is a process of increasing returns.

6 DISCUSSION AND MOTIVATION

The main goal of this project if to develop a complete wildfire modeling and simulation pipeline for the effective and efficient prediction of fire propagation. The framework is expected to be usable as decision support system, for example by the ANPC, and serve as a template to be built upon by project stakeholders. As such, it is directly aligned with goal 15 of the UN's 2030 Agenda for Sustainable Development (United Nations, 2015), namely "Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss".

Concerning the last 10 years, the author has dedicated considerable part this period to basic research in the field of ABM, focusing on high-performance execution and reliable replication (namely, parallel reimplementation) of these simulation models. In particular, in reference (Fachada et al., 2017c) the author proposed a number of ABM parallelization techniques, and statistically compared the output of the parallelized implementations with the original serial one.

Due to the amount of work involved in properly setting up a model comparison experiment, the author also developed and proposed a model-independent framework for comparing the output of simulation models (Fachada et al., 2017b), with concrete implementations for the R (Fachada et al., 2016) and MATLAB/Octave (Fachada and Rosa, 2018) programming environments. This framework vastly improves the cumbersome approaches typically used for replicating and comparing simulation models, namely when the goal is that of parallelizing serial realizations.

During these investigations, we have come to the conclusion that there are several questions that still need answering with respect to faithful and efficient model parallelization. Of these, we highlight two:

1. Determining if the architecture of GPUs is feasible for performing large-scale ABM simulations with highly heterogeneous components. There are some conflicting views on this topic (Coakley et al., 2012), namely the point to which performance gains can be had by developing GPU-focused ABM implementations.
2. Determining if the partition of PRNGs for the purpose of parallelizing ABMs skews results one would obtain with no PRNG partitioning (i.e., with serial model implementations).

In any case, the grounds for this proposal go further back. In the last 15 years the author gained extensive knowledge and experience as a researcher in many of the fields related with the project, namely in complex systems modeling (Fachada et al., 2009; Isidoro et al., 2011; Fachada et al., 2015), parallel/GPU programming (Fachada et al., 2017a), parallel ABM implementations (Fachada et al., 2017c; Fachada and Rosa, 2017), model replication and V&V (Fachada et al., 2017b; Fachada et al., 2016; David et al., 2017), ML (Fachada et al., 2014), stochastic optimization methods (Fernandes et al., 2019; Fernandes et al., 2020; Fernandes et al., 2022) and image processing (Fachada et al., 2012). This proposal offers a unique opportunity to capitalize on this experience by developing a project with potentially vast societal and economical benefits.

7 CONCLUSIONS

We expect that this project will promote activity in a number of scientific areas, while making contributions to current scientific, societal and possibly industrial problems, with positive impacts on society and wealth. Since this is a rapidly evolving field, and we plan to address prominent challenges in the area, it is our hope that this proposal yields a number of high impact publications and captures considerable funding. Given the present-day interest and importance of the topics discussed here, we expect that the proposed research attracts top computer science and engineering students. Importantly, the public and private interest in AI, big data, machine learning and modeling and simulation, not to speak of forest and wildfire management, justifies adequate funding for research and development, while also stimulating innovation and the creation of value in related areas.

ACKNOWLEDGEMENTS

This work is supported by Fundação para a Ciência e a Tecnologia under Grant No.: UIDB/04111/2020 (COPELABS)

REFERENCES

Achtemeier, G. L. (2013). Field validation of a free-agent cellular automata model of fire spread with fire–atmosphere coupling. *International Journal of Wildland Fire*, 22(2):148–156.

Cardil, A., Monedero, S., Ramírez, J., and Silva, C. A. (2019). Assessing and reinitializing wildland fire simulations through satellite active fire data. *Journal of Environmental Management*, 231:996–1003.

Catry, F. X., Rego, F. C., Bação, F. L., and Moreira, F. (2010). Modeling and mapping wildfire ignition risk in portugal. *International Journal of Wildland Fire*, 18(8):921–931.

Coakley, S., Gheorghe, M., Holcombe, M., Chin, S., Worth, D., and Greenough, C. (2012). Exploitation of high performance computing in the FLAME agent-based simulation framework. In *Proc. of the 14th Int. Conf. on High Performance Computing and Communications*, pages 538–545, Liverpool, UK. IEEE.

Collins, R. D., de Neufville, R., Claro, J., Oliveira, T., and Pacheco, A. P. (2013). Forest fire management to avoid unintended consequences: A case study of Portugal using system dynamics. *Journal of Environmental Management*, 130:1–9.

Costa, R. L. C., Miranda, E., Dias, P., and Moreira, J. (2021). Experience: Quality assessment and improvement on a forest fire dataset. *Journal of Data and Information Quality*, 13(1):1–13.

David, N., Fachada, N., and Rosa, A. C. (2017). *Verifying and Validating Simulations*, pages 173–204. Springer International Publishing, Cham.

Denham, M., Wendt, K., Bianchini, G., Cortés, A., and Margalef, T. (2012). Dynamic data-driven genetic algorithm for forest fire spread prediction. *Journal of Computational Science*, 3(5):398–404.

Fachada, N., Figueiredo, M. A. T., Lopes, V. V., Martins, R. C., and Rosa, A. C. (2014). Spectrometric differentiation of yeast strains using minimum volume increase and minimum direction change clustering criteria. *Pattern Recognition Letters*, 45:55–61.

Fachada, N., Lopes, V. V., Martins, R. C., and Rosa, A. C. (2015). Towards a standard model for research in agent-based modeling and simulation. *PeerJ Computer Science*, 1:e36.

Fachada, N., Lopes, V. V., Martins, R. C., and Rosa, A. C. (2017a). cf4ocl: a C framework for OpenCL. *Science of Computer Programming*, 143:9–19.

Fachada, N., Lopes, V. V., Martins, R. C., and Rosa, A. C. (2017b). Model-independent comparison of simulation output. *Simulation Modelling Practice and Theory*, 72:131–149.

Fachada, N., Lopes, V. V., Martins, R. C., and Rosa, A. C. (2017c). Parallelization strategies for spatial agent-based models. *International Journal of Parallel Programming*, 45(3):449–481.

Fachada, N., Lopes, V. V., and Rosa, A. (2009). Simulating antigenic drift and shift in influenza A. In *Proc. of the 2009 ACM symposium on Applied Computing*, SAC '09, pages 2093–2100, New York, NY, USA. ACM.

Fachada, N., Rodrigues, J., Lopes, V. V., Martins, R. C., and Rosa, A. C. (2016). micompr: An R package for multivariate independent comparison of observations. *The R Journal*, 8(2):405–420.

Fachada, N. and Rosa, A. (2018). micompm: A MATLAB/Octave toolbox for multivariate independent comparison of observations. *Journal of Open Source Software*, 3:430.

Fachada, N. and Rosa, A. C. (2017). Assessing the feasibility of OpenCL CPU implementations for agent-based simulations. In *Proceedings of the 5th International Workshop on OpenCL*, IWOCL 2017, pages 4:1–4:10, New York, NY, USA. ACM.

Fachada, V., Fachada, N., Turpeinen, T., Rahkila, P., Rosa, A., and Kainulainen, H. (2012). Topocell – an image analysis tool to study intracellular topography. *The FASEB Journal*, 26(1_supplement):578.2–578.2.

Fernandes, C. M., Fachada, N., Laredo, J. L. J., Merelo, J.-J., and Rosa, A. C. (2020). Population sizing of cellular evolutionary algorithms. *Swarm and Evolutionary Computation*, 58:100721.

Fernandes, C. M., Fachada, N., Merelo, J.-J., and Rosa, A. C. (2019). Steady state particle swarm. *PeerJ Computer Science*, 5:e202.

Fernandes, P. M. A., Inácio, P. M. A., Feliciano, H., and Fachada, N. (2022). SimpAI: Evolutionary heuristics for the ColorShapeLinks board game competition. In Barbedo, I., Barroso, B., Legerén, B., Roque, L., and Sousa, J. P., editors, *Videogame Sciences and Arts, VJ 2020*, volume 1531 of *Communications in Computer and Information Science*, chapter 9, pages 113–126. Springer International Publishing, Cham.

Gómez-González, S., Ojeda, F., and Fernandes, P. M. (2018). Portugal and chile: Longing for sustainable forestry while rising from the ashes. *Environmental Science & Policy*, 81:104–107.

Han, A., Qing, S., Bao, Y., Na, L., Bao, Y., Liu, X., Zhang, J., and Wang, C. (2021). Short-term effects of fire severity on vegetation based on sentinel-2 satellite data. *Sustainability*, 13(1):432.

Hill, D. R. C., Mazel, C., Passerat-Palmbach, J., and Traore, M. K. (2013). Distribution of random streams for simulation practitioners. *Concurrency and Computation: Practice and Experience*, 25(10):1427–1442.

Isidoro, C., Fachada, N., Barata, F., and Rosa, A. (2011). Agent-based model of dengue disease transmission by Aedes aegypti populations. In *Advances in Artificial Life. Darwin Meets von Neumann*, volume 5777 of *Lecture Notes in Computer Science*, pages 345–352. Springer.

Macal, C. (2016). Everything you need to know about agent-based modelling and simulation. *Journal of Simulation*, 10(2):144–156.

Niazi, M. A., Siddique, Q., Hussain, A., and Kolberg, M. (2010). Verification & validation of an agent-based forest fire simulation model. In *Proceedings of the 2010 Spring Simulation Multiconference*, SpringSim '10, pages 1:1–1:8, San Diego, CA, USA. Society for Computer Simulation International.

Ntinas, V. G., Moutafis, B. E., Trunfio, G. A., and Sirakoulis, G. C. (2017). Parallel fuzzy cellular automata for data-driven simulation of wildfire spreading. *Journal of Computational Science*, 21:469–485.

Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B. R., Girvan, M., and Ott, E. (2018). Hybrid forecasting of chaotic processes: using machine learning in conjunction with a knowledge-based model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(4):041101.

Smith, J., Barfed, L., Dasclu, S. M., and Harris, F. C. (2016). Highly parallel implementation of forest fire propagation models on the GPU. In *2016 International Conference on High Performance Computing & Simulation (HPCS)*, pages 917–924. IEEE.

Trucchia, A., D'Andrea, M., Baghino, F., Fiorucci, P., Ferraris, L., Negro, D., Gollini, A., and Severino, M. (2020). Propagator: an operational cellular-automata based wildfire simulator. *Fire*, 3(3):26.

United Nations (2015). Transforming our world: the 2030 agenda for sustainable development. Department of Economic and Social Affairs.

Widyastuti, K., Imron, M. A., Pradopo, S. T., Suryatmojo, H., Sopha, B. M., Spessa, A., and Berger, U. (2020). Peatfire: an agent-based model to simulate fire ignition and spreading in a tropical peatland ecosystem. *International Journal of Wildland Fire*.

Wilensky, U. and Rand, W. (2007). Making models match: replicating an agent-based model. *Journal of Artificial Societies and Social Simulation*, 10(4):2.

Williams, R. A. (2018). Lessons learned on development and application of agent-based models of complex dynamical systems. *Simulation Modelling Practice and Theory*, 83:201–212.

Xiao, J., Andelfinger, P., Eckhoff, D., Cai, W., and Knoll, A. (2019). A survey on agent-based simulation using hardware accelerators. *ACM Computing Surveys*, 51(6):131:1–131:35.

Yassemi, S., Dragićević, S., and Schmidt, M. (2008). Design and implementation of an integrated GIS-based cellular automata model to characterize forest fire behaviour. *Ecological Modelling*, 210(1):71–84.