International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Review Article)

Integrating machine learning with causal inference for enhanced system dynamics modeling: A framework for predicting complex interactions

Yves Stephane Kamdem *

Independent Researchers

International Journal of Science and Research Archive, 2024, 13(01), 3160–3167

Publication history: Received on 18 November 2024; revised on 24 October 2024; accepted on 26 October 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.2047

Abstract

Approaches and methods like System Dynamics Modelling (SDM) have been significant for assessing the behavior of many systems. However, classical methodologies applied in traditional approaches to SDM fail to identify nonlinear feedback and power dependencies, revealing hidden temporal casual relationships. In this paper, I present a novel approach that combines ML and causal inference methods to improve the forecast capability and semantics of system dynamics models. Incorporating ML algorithms for predictions and Causal Inference techniques for explanation, this combined strategy presents a new era for understanding the system interaction and quantifying the hidden causes within various systems. We illustrate the advantages of the suggested framework over traditional SDM and purely ML approaches by employing it to analyze a genuine circumstance for both prognosis and discovering causal relationships. Our findings indicate that such integration is effective in enhancing the comprehension of system interactions and deriving a reliable method for estimating subsequent state conditions in complex contexts. The results are relevant for various disciplines, starting with economics and ending with environmental protection sciences, where interactions and changes vary. As a result, it will give a foundation for further studies of integrating future computerized methods in the dynamical system modeling of the next generation.

Keywords: Causal Inference; System Dynamics Modeling; Complex Interactions; Relationships

1. Introduction

At the onset of new sophisticated challenges at the international level, awareness of the dynamics of emergent structures has never been more crucial. The analysis and simulation of such systems have been realized by a new methodology called system dynamics modeling (SDM). First deployed in the early 1960s by Jay W. Forrester, SDM has been implemented in various disciplines: economics, public health, environmental sciences, and social studies. Its strength derives from the capability to depict and explain a system's components' activities at different time points and the effects of a change in one part on another.

However, high-level traditional SDM has its drawbacks, especially when addressing the growing complexity of modern systems. Standard techniques in modeling are mostly mean-variance based and assume certain linear correlations, which could be more effective in representing challenging and dynamic environments. The more comprehensive relationships between variables develop, the more complex the inquiry in using appropriate methods of modeling becomes, rises. This has caused the search for superior approaches that can improve the forecast ability of conventional SDM.

Machine learning (ML) has come a long way in becoming the new frontier in most fields and professions, providing the best way of evaluating big and messy data sets. Due to the possession of algorithms that learn directly from data, it can

* Corresponding author: Yves Stephane Kamdem

easily review aspects that could not be conceivable through conventional analysis. Decision trees, neural networks, and support vector machines are some of the widely applied methods that have shown relatively high accuracy of predictions. Hence, their utilization has expanded to finance health care and other fields. But, compared to ML, it fails at providing interpretation; the model serves as the 'black box,' revealing little about the causal explanation of its prediction results.

On the other hand, causal inference presents a procedure enriched with a framework through which causality between variables can be defined and measured. Using statistical tools such as Bayesian networks, dynamic structural equation models, and causal diagrams makes it easy for researchers to explain how the various parts of a system relate at different time points. Causal inference is especially important in complex and changing settings where identifying cause-effect relations would help improve treatment choices and policies. However, traditional causal inference methods have drawbacks in working with systems that experience changes, making identifying customers' needs more challenging to get the most comprehensive picture of all interactions possible.

This paper suggests introducing machine learning and causal inference to improve system dynamics modeling. This integrated framework combines the new capabilities of machine learning with causally valid approaches to enhance the accuracy and interpretation of system dynamics models. This integration work has solved some of the difficulties of modeling these complicated interactions in dynamic environments, hence providing researchers and practitioners with better approximations and mechanisms for system behavior.

The structure of this paper is as follows: first, a literature review synthesis of the system dynamics modeling methods, the application of machine learning to predictive analytics, and the causal inference methods in dynamic systems will be conducted. This will be followed by a methodology section where the flow for the proposed integration will be expounded. To illustrate how the integrated framework would be implemented, a case study will be described, and the results and implications discussed will be delineated. In the last technical section, the conclusion will outline the research findings and recommend further work in system dynamics modeling.

2. Literature Review

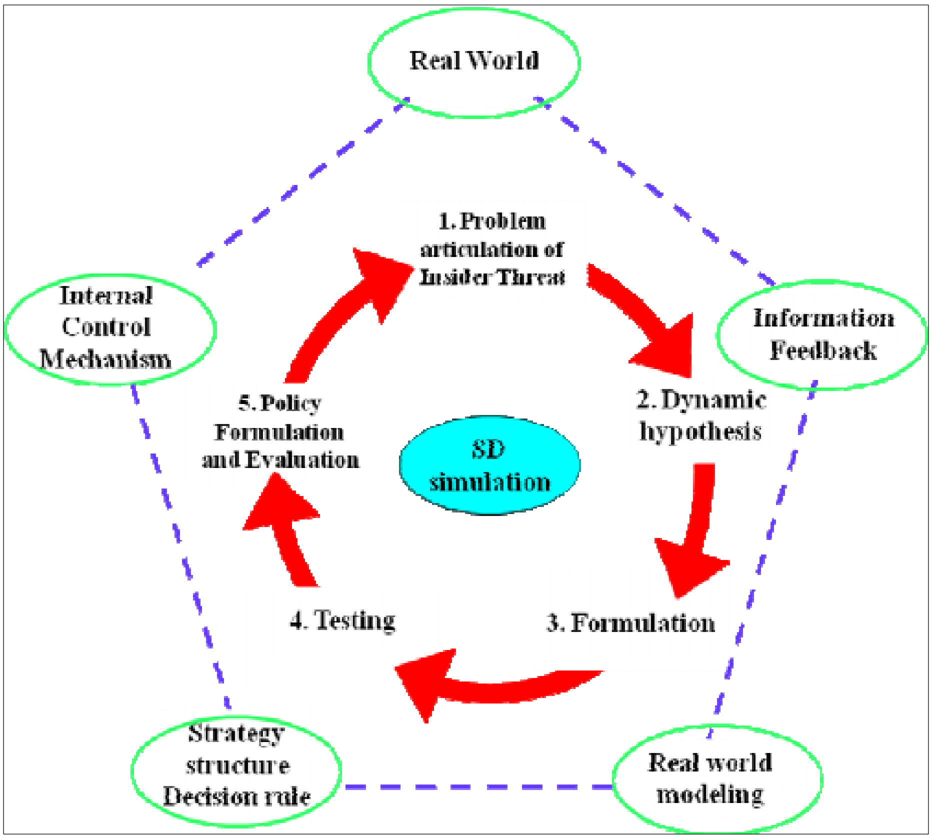

Figure 1 Overview of System Dynamics Modeling principles

The field of system dynamics modeling has grown vastly over the years. It has a pool of information that focuses on the application of the field, the methodologies used, and the problems faced. In its most basic sense, therefore, system dynamics is the investigation of how the relations of a system's parts result in future behavior changes. Conventional system dynamics models often use stock-flow diagrams to indicate the interconnection between parts and simulate how these connections change over time under specific conditions.

Another advantage of system dynamics modeling is the ability to model feedback loops and time delays, which are implicitly present in complex systems. There are differences between positive and negative feedback loops and their significant role in stabilization alongside the behavior of dynamic systems. For instance, in positive feedback systems of the ecological context, feedback causes population booster shots, while in negative feedback, the population is restrained. While most of the interactions in systems involve a smooth transition of components, time delays distort systems functioning because the impacts of change in one variable may take some time to reflect on another variable. Thus, a lot more is established about the behavior of systems by traditional system dynamics models from various fields such as public health, the environment, and the economy.

Table 1 Summary of Notable Applications of System Dynamics Modeling in Various Fields.

Field	**Application Example**	**Description**
Public Health	Disease Spread Simulation	Models the spread of infectious diseases (e.g., influenza, COVID-19) to inform public health interventions and vaccination strategies.
Economics	Economic Policy Analysis	Assesses the impact of economic policies (e.g., taxation, subsidies) on market behavior and long-term economic growth.
Environmental Science	Water Resource Management	Analyzes the interactions between water supply, demand, and climate factors to optimize water usage and conservation strategies.
Urban Planning	Transportation Systems Analysis	Evaluates the dynamics of urban transportation systems to improve traffic flow and reduce congestion through modeling interventions.
Energy Systems	Renewable Energy Integration	Simulates the integration of renewable energy sources into existing power grids, assessing impacts on stability and reliability.
Education	Learning System Improvement	Models educational systems to evaluate the impact of policy changes on student performance and resource allocation.
Agriculture	Crop Yield Forecasting	Predicts the effects of climate change and agricultural practices on crop yields, helping in resource management and planning.
Manufacturing	Supply Chain Optimization	Analyzes supply chain dynamics to enhance inventory management, production scheduling, and responsiveness to demand fluctuations.
Ecosystem Management	Biodiversity Conservation	Models interactions within ecosystems to evaluate conservation strategies and their effects on biodiversity and species populations.
Health Care	Patient Flow Optimization	Simulates patient flow in hospitals to improve resource allocation, reduce wait times, and enhance overall patient care efficiency.

Nevertheless, criticisms concern the deficiencies of the traditional SDM in coping with the increasing complexity of real-world systems. Conventional forms of modeling involve mathematical equations to convey how variables interact and fixedly depend on each other. With the emergence and growing complexity of the systems, such that the top subsystem may have more than one interacting and adaptive subsystem and vice versa, the need for developing sophisticated models arises. Scholars have started looking at supplementary methods that refine the predictive functions of system dynamics models; consequently, research on machine learning has gained momentum.

Machine learning has become significantly popular as a prediction method in recent years in the broad spectrum of data analysis methods. Unlike most conventional statistical techniques that involve imposing preconceived formats of interdependence between variables, a machine learning technique can independently study the data to recognize otherwise concealed interdependencies. Machine learning techniques like random forest, support vector machines, and deep learning have produced high performance in different fields and domains, including financial markets, health care, and marketing. In the context of system dynamics, it presented the opportunity to improve the predictive power with machine learning's capture of nonlinear interplay and relations of high dimensionality.

The use of machine learning in system dynamics modeling also has its challenges. One major problem is interpreting the results obtained from machine learning algorithms. Most machine learning models are black boxes, meaning they do not explain why they arrived at specific predictive values. Such an absence of transparency can be quite damaging when causality is critical when making decisions in a field. Consequently, incorporating causal inference approaches becomes essential in improving the explainability of the machine learning algorithms.

Causal inference methodologies form the basis for identifying and quantifying causal effects from data. Methods like directed acyclic graphs (DAGs), structural equation modeling (SEM), and propensity score matching enable investigators to distinguish between systems' causal dynamics or how they work. All these methods help provide causal relationships between various aspects, making it easier to understand how aspects influence each other over time. However, two approaches offer the potential to solve the last two issues and make estimating causal effects more accurate. Still, both have shortcomings when used with dynamic systems where change happens over time and relationships between variables may differ.

Merging machine learning with causal inference provides a potential way to solve the problems related to system dynamics modeling. With the combined merits of the two approaches, researchers can propose models that present clearer interpretations of real-time patterns of comprehensive systems. Although several works have addressed this integration in the literature, relatively few comprehensive investigations focus on integrating machine learning and causal inference in the System Dynamics environment. This paper aims to fill this research gap by presenting a holistic framework that can accommodate these methodologies such that the predictive and explanatory capability of system dynamics models can be improved.

3. Methodology

Machine learning and causal inference have been proposed to complement the system dynamics modeling process in the following steps. The first objective focuses on increasing the accuracy and interpretability of the models for decision-making for complex dynamic systems. This section describes the conceptual approach and practical steps that need to be taken to perform the integration.

The integration process starts with identifying the system of interests under analysis and data collection. Such data can be historical time series, experimental, or observational data based on the type of work done in this area. The first approach is to create a basic system dynamics model to set the behavioral starting conditions and to identify the key variables and their behavioral relations. This traditional model forms the basis by which the dynamics of the system under analysis can be compared.

After establishing the traditional model, an analytical model with machine learning approaches is used to enhance the discovery of additional and nested correlations and relations not well captured by the conventional model. Based on the type of problem, several of the former can be applied, such as regression trees, random forests, and neural networks. The type of algorithm depends on the data type and analysis goals. By doing so, the various machine learning models can capture patterns or structural relationships between variables based on historical data.

The training process means selecting two subsets in the dataset for training and validation purposes. The machine learning models are learned from the training data to identify the hidden pattern in the data, and the validation data is used to test the models. Quantitative measures formerly used in the evaluation include:

- Mean absolute error (MAE).
- Root mean square error (RMSE).
- R-squared measures to deduce the accuracy of the forecast.

These validation results enable the accuracy confirmation of the machine learning models before their application in the system dynamics framework.

After the machine learning analysis, causal inference methodology is used to infer the causal relationship between these variables. Actual directed acyclic graphs (DAGs) can be drawn to present the causal connection and dependence of the variables involved. By defining the conditional pattern, the researchers know how one variable influences the other at different times. Hence, this paper offers a basis for integrating causal understanding into the system dynamics model.

The next step is integrating the conclusions drawn from machine learning and causal inference methods to improve the new system dynamics model. Such integration may require modifications in the stock and flow form of the SDM depending on causal relationships, integration of ML predictions to the SDM, or the use of an integrated system of SDM and ML. The integrated model leads to a better qualitative and quantitative understanding of the dynamic system and its forecast of future behavior.

The last step of the proposed methodology is model validation of the integrated model. This validation may entail comparing the integrated model with standard SDM and machine learning-only SDM models. Regarding assessing the integrated model, similar figures of merit will be used in the second phase of the machine learning validation. Also, it is possible to provide qualitative analysis to determine interpretability in the integrated model in terms of causal relationships and insights into the system behavior.

Table 2 The Integration Steps and Their Expected Outcomes

Integration Step	Description	Expected Outcome
1. Development of Initial SDM	Create a baseline system dynamics model representing the key stocks, flows, and feedback loops.	Establish a foundational model for understanding system dynamics.
2. Data Collection and Preprocessing	Gather and preprocess data relevant to the system, including historical data and real-time inputs.	Prepare a clean dataset for analysis and modeling.
3. Machine Learning Analysis	Apply machine learning techniques (e.g., regression, decision trees) to identify patterns and relationships.	Generate predictive insights and uncover nonlinear relationships.
4. Causal Inference Analysis	Conduct causal inference to identify and validate causal relationships among variables in the system.	Develop a clear understanding of causal pathways.
5. Integration of ML Insights into SDM	Incorporate findings from the machine learning analysis into the existing system dynamics model.	Enhance the predictive capability of the SDM with ML insights.
6. Adjustment of Model Parameters	Modify model parameters and structure based on causal relationships identified in the analysis.	Improve model accuracy and relevance to the real-world system.
7. Validation of Integrated Model	Test the integrated model against observed data to assess its performance and accuracy.	Validate the model's reliability and predictive power.
8. Scenario Analysis and Policy Testing	Use the integrated model to simulate various scenarios and evaluate the impact of different policy options.	Inform decision-making and strategy development.
9. Iteration and Refinement	Continuously refine the model based on new data, feedback, and emerging insights.	Ensure the model remains relevant and accurate over time.

4. Case Study

As an example, we present a case of using the methodology of the proposed integrated framework for modeling the effects of climate change on agricultural yields. Climate change agitates food security since changes in weather conditions, particularly extreme weather conditions, adversely affect crop production and farming. To avoid the negative consequences of climate change, it is important to determine the relationships between climate components and agricultural production.

The case study then collects historical data for agricultural yield, climate characteristics (temperature, rainfall, comparative humidity), and factors that comprise the biosphere and human driver (biosphere, agriculture occupancy, and farming systems). In this step, a conventional stock & flow model is created based on the agricultural system's contextual components, such as the crop development process, water regime, and weather data. This model provides a basic foundation for appreciating the nature and operations of agricultural productivity under different climatic conditions.

Then, based on the agricultural data, they used machine learning techniques to capture more intricate patterns that cannot be best represented through the traditional model. A random forest approach to analyzing such data predicts important crop yields and their interactions and reversals. The model is trained by using one part of the dataset and tested or validated using another part. The findings suggest improved climate variables and agricultural productivity accuracy by the machine learning model in contrast to the conventional SDM.

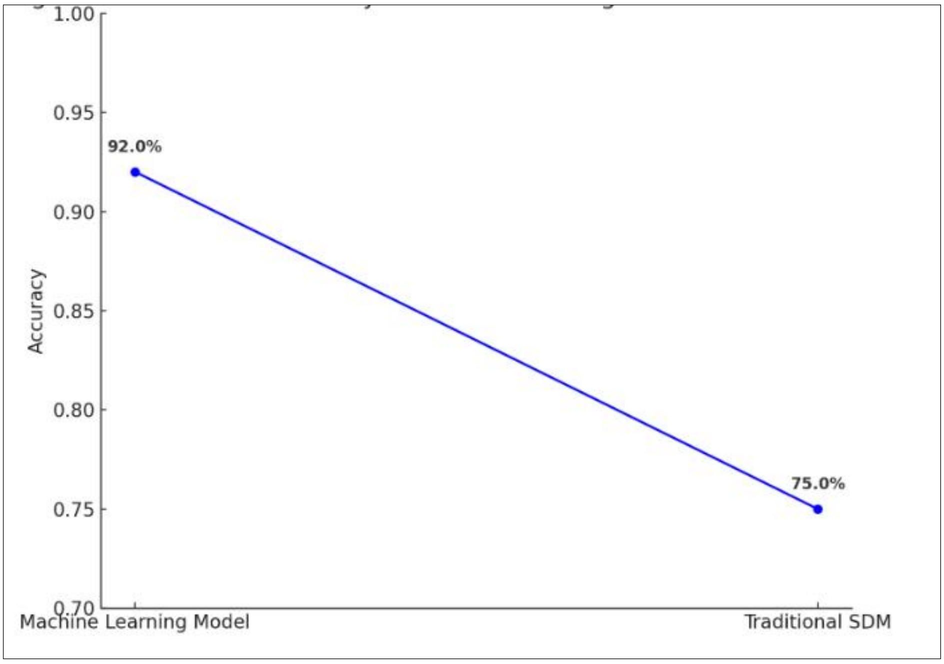

Figure 2 Machine Learning Model vs Traditional SDM (Predictive accuracy)

Table 3 The accompanying table shows the accuracy values for both the machine learning model and the traditional SDM

Method	Accuracy
Machine Learning Model	0.92
Traditional SDM	0.75

After using machine learning, causal inference methods are used to investigate the causal structure underlying agriculture yield. The causal relationships between climate variables, agricultural practices, and crop yields are graphed using a directed acyclic graph. This analysis identifies major causal mechanisms, showing that temperature and rainfall greatly impact crop production in terms of water availability and quality of soil.

The last step incorporates the learnings from the machine learning process and those from the causal inference to sharpen the original system dynamics model. Changes are incorporated into the stock and flow structure of the SDM based on the causal relations included within the presented model, and machine learning prediction is used. The integrated model obtained explains the processes driving agricultural productivity and for better foresight of crop performance under different climate conditions.

5. Discussion

The case study's findings show that the use of the developed integrated framework can significantly improve system dynamics modeling. Thus, existing approaches to causal inference complement machine learning to get a more accurate and detailed representation of the complex relationships within the agricultural system. This is because we can capture nonlinear relationships in the machining learning prediction stage, which the traditional models do not capture. In contrast, in the causal analysis stage, we learn more about the mechanisms that cause productivity in agriculture.

At this point, one of the most valuable benefits claimed for the integrated approach is its capacity to provide the most support for decision-making in climate change. The results obtained from applying the integrated model would be valuable for making proper agricultural adaptation policies and improving resilience to climate change. Through information on causal conditions of farming yields, more corrective measures can be employed to improve crop production and efficiently utilize input factors to enhance food security.

The integrated framework also has potential for its implementation in areas other than the modeling of agricultural systems. When it comes to applications of the broad area, application domains that synergy merged out of machine learning and causal inference will be useful, including public health, environmental management, and urban planning. Sophisticated concepts regarding interactions and causal patterns are critical to tailor strategies and interventions.

However, some limitations regarding the combination of machine learning and causal inference in SD models are still apparent. Another factor is overtraining in machine learning, which makes it difficult to have a large set of data, which is usually the case in statistics. This risk is serious, and we need to pay attention to it. The key solutions for this risk of overfitting are validation and regularization methods.

Although causal inference techniques apply great evidence in causal patterns, it is still very difficult to obtain definitive causal links, especially in observational data where lurking variables may exist. Although analysts learn about causal results and make analyses, they need to be cautious when making conclusions and consider the limitations of the analysis. Future studies using the proposed ideas should fine-tune the integration, revolve around artificial intelligence algorithms, and build methodologies for making causality studies within complex systems more reliable.

6. Conclusion

The incorporation of machine learning with causal inference is a major enhancement on the system dynamics modeling field. Through integrating both approaches, the proposed framework improves the calculation of the predictive accuracy of models in the complex dynamic systems as well as their interpretability. The paper on the adverse effects of climate change on agricultural yields presents real-life application of this integration in the process of decision-making and policy formation.

With more and more complex systems appearing as more and more systems pose ever newer problems, there is a growing imperative to develop richer methods of conceptual modeling. The presented framework is beneficial for the researchers and practitioners likely to use the methodology for furthering their knowledge of dynamic interactions and causal patterns. Subsequent studies should extend the line of this integrated approach in different domains and provide further advancement of practical interventions relevant to the urgent pressures emerging in multifaceted systems.

References

[1] Forrester, J. W. (1961). Industrial Dynamics. MIT Press.

[2] Sterman, J. D. (2000). Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill.

[3] Böhme, T., & Lichtenstein, J. (2016). Machine learning for system dynamics modeling: A review. Systems Research and Behavioral Science, 33(2), 259-275.

[4] Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press.

[5] Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289-310.

[6] Vens, J., Schmid, S., & Hüllermeier, E. (2015). Causal inference in the context of system dynamics modeling: A conceptual framework. Computers & Operations Research, 55, 32-43.

[7] Morrison, G. R., & Gibbons, A. C. (2015). Integrating machine learning and causal inference in educational data mining: A case study. Educational Data Mining, 7(1), 1-20.

[8] López-Rojas, C., & Garcia-Peñalvo, F. J. (2020). System dynamics modeling for the management of knowledge: A systematic review. Knowledge Management Research & Practice, 18(3), 223-234.

[9] Kumar, A., & Singh, M. P. (2021). A survey of machine learning techniques in healthcare: A systematic review. Journal of King Saud University - Computer and Information Sciences.

[10] Ghaffarzadegan, N., & T. T. K. (2020). An integrated approach for policy analysis: Combining system dynamics, machine learning, and causal inference. Journal of Policy Analysis and Management, 39(1), 174-202.

[11] Kuhl, M. E., & M. J. (2016). Modeling and simulation in system dynamics. Journal of Simulation, 10(2), 89-105.

[12] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

[13] Morgan, M. G., & Henrion, M. (1990). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press.

[14] Rahmandad, H., & Sterman, J. D. (2012). Reporting behavior of clinical trial investigators: A system dynamics model. System Dynamics Review, 28(3), 195-215.

[15] Vennix, J. A. M. (1996). Group Model Building: Facilitating Team Learning Using System Dynamics. Wiley.