

OPINION

Moving beyond simulation and learning: Unveiling the potential of complexity data science

Frank Emmert-Streib^{1*}, Hocine Cherifi², Stuart Kauffman³, Olli Yli-Harja^{1,3}

1 Predictive Society and Data Analytics Lab, Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland, **2** Laboratoire Interdisciplinaire Carnot de Bourgogne (ICB), UMR 6303 CNRS, Université de Bourgogne, Dijon, France, **3** Institute for Systems Biology, Seattle, Washington, United States of America

* frank.emmert-streib@tuni.fi



Abstract

Complexity science is a multidisciplinary field that examines various aspects of complex systems. While complexity science places a significant emphasis on simulation, it has a somewhat neglectful treatment of learning. In this paper, we explore a recent example of the potential synergy between simulation and learning, illustrated by the concept of digital twins. We argue that integrating simulation and learning holds significant promise beyond the scope of digital twins alone. In our view, the general amalgamation of complexity science and data science heralds the dawn of a distinct and innovative field in its own right, which we call complexity data science.

OPEN ACCESS

Citation: Emmert-Streib F, Cherifi H, Kauffman S, Yli-Harja O (2024) Moving beyond simulation and learning: Unveiling the potential of complexity data science. *PLOS Complex Syst* 1(2): e0000002. <https://doi.org/10.1371/journal.pcsy.0000002>

Editor: Jin Liu, Shanghai Maritime University, CHINA

Published: October 3, 2024

Copyright: © 2024 Emmert-Streib et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Academy of Finland (352266 and 352263 to FES and OYH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: FES is Section Editor of PCB. HC is Editor-in-Chief of PCB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

Complexity science is a broad multidisciplinary field that explores complex phenomena in various domains, including physics, biology, and economics [1–3]. Its primary focus is on simulation models to unravel the fundamental principles, recurring patterns, and emergent behaviors inherent in complex systems including ecosystems, regulatory networks, multi-scale systems, social structures, and economic markets [4–7].

In our opinion, combining methods from simulation and learning marks a new stage that can lead to profound insights beyond the capabilities of the constituting approaches. In the following, we discuss a prime example of this synthesis: digital twins.

Combining simulation and learning: Digital twins

While the idea of a digital twin (DT) originated in manufacturing, in recent years, the concept has become of great interest for many domains in science [8,9]. Simply put, a digital twin is an ideal concept—a digital representation of a real-world object that closely mirrors its physical counterpart. This digital representation takes the form of a computer-based simulation model, while the real-world object can encompass systems or processes, such as an engine, or an economic process. Notably, a crucial attribute of a DT is its ability to continuously adapt and learn over time [10].

This characteristic closely resembles the adaptive behavior observed in certain complex systems, referred to as complex adaptive systems (CAS), which dynamically respond to environmental changes [11]. These systems have been explored in fields like biology, epidemiology, and economics, where they serve as models for understanding dynamic interactions. However, the concept of a DT distinguishes itself from conventional CAS in several ways. First, a DT should be recognized as a purposefully engineered construct, as opposed to a naturally occurring entity. It provides a certain level of interpretability and replicates key aspects of the underlying system. This makes a digital twin a pragmatic prediction instrument which will be judged based on the quality of its predictions forming the output of the model.

Another significant differentiation lies in the fact that a DT not only generates “simple” predictions but also facilitates the exploration of “What-If” scenarios. This means it enables the examination of virtual interventions. Examples for such interventions are the following:

- Economics: Policy alterations and their effect on the economy
- Epidemiology: Alternative vaccination strategies and their effect on an epidemic spread.

Fig 1 shows a visualization of both concepts behind a digital twin: simulation (Fig 1A) and learning (Fig 1B). The DT possesses a defined structure, e.g., given by a complex network, and exhibits activity patterns providing a dynamic simulation that mirrors a corresponding “physical twin” (PT). By introducing interventions, e.g., to the structural network, it becomes possible to explore consequences of “What-If” scenarios, leading to modified predictions. Employing analytical techniques from machine learning or statistics enables then a comparative analysis of these predictions, ultimately facilitating informed decision-making. The learning aspect of a DT is emphasized in Fig 1B. In contrast to traditional simulations, a DT undergoes a series of updates allowing to adjust its parameters by taking into consideration new data.

Beyond digital twins: Complexity data science

In our opinion, the fusion of simulation and learning, as intrinsic to DT, holds immense promise that extends beyond the realm of digital twins. At its most fundamental level, the amalgamation of complexity science and data science can aptly be termed as “complexity data science,” signifying the emergence of a distinct and innovative field in its own right.

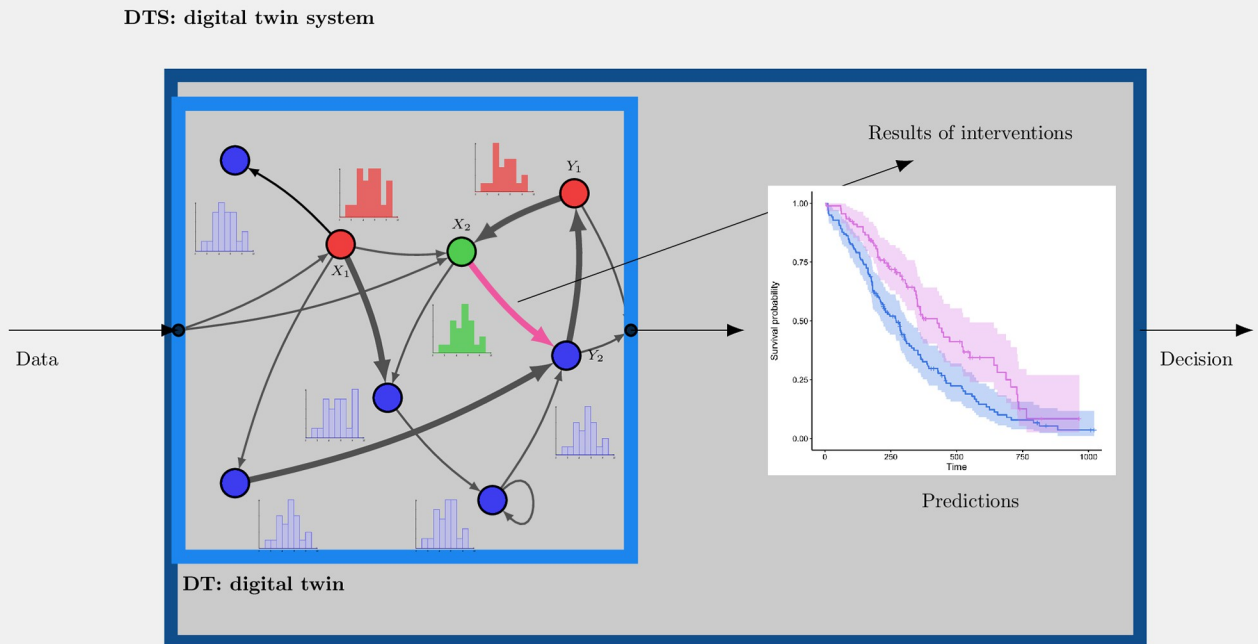
An example of the advantage derived from this integration is the problem of explainable AI [12,13]. As complexity science deals with the study of complex systems, allowing for the simulation of crucial real-world phenomena, such systems not only excel at capturing essential functional elements, but also provide interpretable models. However, achieving accurate predictions with such models requires calibration, akin to the iterative update process observed in digital twins. That means by leveraging interpretable models derived from complexity science, we can move away from opaque black-box models commonly found in data science, such as deep learning networks, towards more explainable solutions.

Another example for the benefit of complexity data science is that it allows for data-driven complex systems. Traditionally, a complex system is either a good or a bad model for a certain phenomenon from the start without the possibility of adjustments. Now we have the possibility to not only adjust parameters of the model but also its structure itself. This enables an increased flexibility to allow for the eventual convergence of the model toward the desired solution.

Conclusion

It is interesting to note that in Vemuri’s classic book, “Modeling of Complex Systems: An Introduction” [14], there is a chapter about forecasting (Chap. 5). This is intriguing because forecasting

A. Simulation



B. Learning

Time evolution of a digital twin system

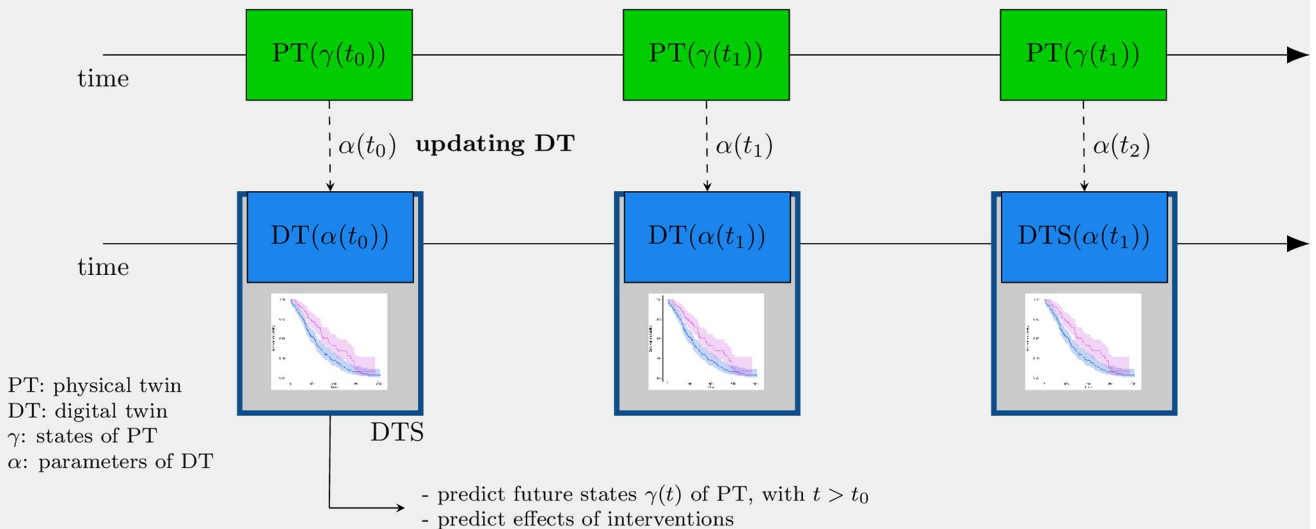


Fig 1. An overview of a DT and the interplay between simulation and learning. (A) The DT can make predictions and enables interventions. (B) The DT learns over time by calibrating its parameters on new data.

<https://doi.org/10.1371/journal.pcsy.0000002.g001>

is not widely emphasized in complex systems but Vemuri, more than 4 decades ago, recognized the importance of connecting both modeling and prediction. In our opinion, this combination holds the potential to provide novel insights and capabilities, pushing the boundaries of what can be achieved in domains that traditionally focus on either simulation or learning.

References

1. Mainzer K. Thinking in complexity: The computational dynamics of matter, mind, and mankind. Springer; 2004.
2. Jensen HJ. Complexity science: the study of emergence. Cambridge University Press; 2022.
3. Anderson PW. More is different. *Science*. 1972; 177(4047):393–396. <https://doi.org/10.1126/science.177.4047.393> PMID: 17796623
4. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol*. 1969; 22:437–467. [https://doi.org/10.1016/0022-5193\(69\)90015-0](https://doi.org/10.1016/0022-5193(69)90015-0) PMID: 5803332
5. Farmer JD, Foley D. The economy needs agent-based modelling. *Nature*. 2009; 460(7256):685–686. <https://doi.org/10.1038/460685a> PMID: 19661896
6. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science*. 2013; 342(6164):1337–1342. <https://doi.org/10.1126/science.1245200> PMID: 24337289
7. Siegenfeld AF, Bar-Yam Y. An introduction to complex systems science and its applications. *Complexity*. 2020; 2020:1–16.
8. Laubenbacher R, Sluka JP, Glazier JA. Using digital twins in viral infection. *Science*. 2021; 371(6534):1105–1106. <https://doi.org/10.1126/science.abf3370> PMID: 33707255
9. Bauer P, Stevens B, Hazeleger W. A digital twin of Earth for the green transition. *Nat Clim Change*. 2021; 11(2):80–83.
10. Emmert-Streib F. Defining a Digital Twin: A Data Science-Based Unification. *Mach Learn Knowl Extr*. 2023; 5(3):1036–1054.
11. Holland JH. Complex adaptive systems. *Daedalus*. 1992; 121(1):17–30.
12. Loyola-Gonzalez O. Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*. 2019; 7:154096–154113.
13. Emmert-Streib F, Yli-Harja O, Dehmer M. Explainable Artificial Intelligence and Machine Learning: A reality rooted perspective. *WIREs Data Min Knowl Discov*. 2020; 10:e1368.
14. Vemuri V. Modeling of complex systems: An introduction. Academic Press; 1978.