Author(s): de Koning, Koen; Broekhuijsen, Jeroen; Kühn, Ingolf; Ovaskainen, Otso; Taubert, Franziska; Endresen, Dag; Schigel, Dmitry; Grimm, Volker

Title: Digital twins : dynamic model-data fusion for ecology

Year: 2023

Version: Published version

Please cite the original version:

de Koning, K., Broekhuijsen, J., Kühn, I., Ovaskainen, O., Taubert, F., Endresen, D., Schigel, D., & Grimm, V. (2023). Digital twins : dynamic model-data fusion for ecology. Trends in Ecology and Evolution, 38(10), 916-926. https://doi.org/10.1016/j.tree.2023.04.010

CellPress OPEN ACCESS
Trends in Ecology & Evolution

Opinion

Digital twins: dynamic model-data fusion for ecology

Koen de Koning,[1,@] Jeroen Broekhuijsen,[2] Ingolf Kühn,[3,4,5] Otso Ovaskainen,[6,7,8] Franziska Taubert,[9] Dag Endresen,[10,*] Dmitry Schigel,[11,@] and Volker Grimm[5,9,12]

Digital twins (DTs) are an emerging phenomenon in the public and private sectors as a new tool to monitor and understand systems and processes. DTs have the potential to change the status quo in ecology as part of its digital transformation. However, it is important to avoid misguided developments by managing expectations about DTs. We stress that DTs are not just big models of everything, containing big data and machine learning. Rather, the strength of DTs is in combining data, models, and domain knowledge, and their continuous alignment with the real world. We suggest that researchers and stakeholders exercise caution in DT development, keeping in mind that many of the strengths and challenges of computational modelling in ecology also apply to DTs.

Highlights

Digital twins (DTs) are rapidly gaining popularity across industries as a digital tool for continuous monitoring of physical phenomena, and the first DTs have now been developed in various environmental science disciplines.

DTs are becoming part of the political sustainability agenda (e.g., in the 'Destination Earth' programme of the European Commission), with the vision of developing DTs for the climate, the ocean, and biodiversity.

Digital transitions are happening across domains (including ecology) and have advanced the use of high-tech sensors for automated data collection and processing.

Technological developments in digital infrastructure have made data storage, automation, large-scale models, and interactive applications cheaper by many orders of magnitude.

These developments clear the way for DT adoption in ecology, but proper guidance is required.

DTs of ecological systems

DTs have become a hot topic in many scientific domains [1–3]. They are described as digital counterparts of a physical object or process that are linked to each other to continuously update and improve realism and reduce uncertainty about the physical object/process [4]. Very recently, also in ecology, references to DTs are being made, in particular triggered by the 'Destination Earth' programme of the European Commission [5], which is based on the vision of developing DTs for the climate [6], the ocean [7], and biodiversity[i]. The idea is to establish statistical and mechanistic models that are continuously updated with data both on drivers and of variables of interest (Box 1). These data streams can be used for continuous recalibration so that the DTs can improve our understanding of ecosystems and support decision-making.

DTs have great potential for ecology. Technological advancements in methods and applications have led to a rapid digitalisation of ecology in the past decades [8–10]. We therefore summarise recent trends which make DTs a timely approach in ecology, but also discuss inherent limitations to the DT concept in ecology to avoid expectations that might misguide stakeholders such as modellers, ecological researchers, policy developers, and natural resource managers.

DTs: why now?

DTs were first implemented for the National Aeronautics and Space Administration (NASA) spacecraft and are widely used in manufacturing, urban planning, healthcare, and the automotive industry [2,3]. DTs have proved to be a useful tool for monitoring and decision-making in industry and engineering because they allow accurate, precise, and real-time monitoring and simulation of processes that are hard to observe in the real world [11]. Moreover, they can be used for early warning signal detection, experiments, and predictive and prescriptive analytics [12].

The increasing use of DTs is driven by advances in monitoring and automated retrievals from databases [13], as well as advances in making models realistic for their purposes. It has become a lot easier and cheaper to collect data, create models and algorithms using tools, platforms, and

[1]Wageningen University and Research, Environmental Systems Analysis Group, P.O. Box 47, 6700, AA, Wageningen, The Netherlands
[2]Nederlandse organisatie voor toegepast natuurwetenschappenlijk onderzoek – TNO, Department of Monitoring & Control Services, Eemsgolaan 3, 9727 DW Groningen, The Netherlands
[3]Helmholtz Centre for Environmental Research – UFZ, Department of Community Ecology, Theodor-Lieser-Strasse, 4, 06120 Halle, Germany
[4]Martin Luther University Halle-Wittenberg, Institute for Biology/ Geobotany & Botanical Garden, Große Steinstraße 79/80, 06108 Halle, Germany

libraries, run models on large-scale infrastructure, compose models and workflows for larger-scale and more complex interactions, and create interactive applications that empower end-users to better understand complex systems and make informed decisions. This transition is happening across domains [3,14–16], including environmental sciences. Yet, the potential use of DTs has only recently been explored in the environmental sciences [17].

Ecology in the digital era

Two important trends make the adoption of ecological DTs timely and relevant. These are global digitalisation, and a call for more evidence-based decision-making in biodiversity conservation [18–21].

New digitalisation technologies are rapidly being adopted in ecology, with drones, high-frequency biologgers, automatic species identification (e.g., through DNA metabarcoding [22], camera-trap sampling [23], or passive acoustic sampling [24]), digitisation of natural history and other scientific collections [25,26], and citizen science [27] through apps [28] and social media platforms [29] as recent examples. Also, many observational and monitoring networks utilising these new methods are in place to provide data flows on biodiversity and ecosystems. This allows us to build DTs on existing ecological research data infrastructures. They encompass, among other things, initiatives and research infrastructures such as Geo-BON[ii], NEON [30], eLTER[iii] [31], SAEON[iv], TERN[v], GBIF[vi], and LifeWatch[vii]. Moreover, there is an increasing vision towards large-scale integration of ecological datasets and generalised models [8,26,31–33], and developing highly accurate simulation models of the earth's systems that guide policy-makers though the 21st century challenges [1].

Second, there is an urgent call for better use of knowledge in biodiversity conservation [34–37], that is, evidence-based conservation [38]. This means that decision-making is ideally based on the latest scientific evidence, sound theoretical knowledge on species and ecosystem functioning, and field measurements that give an accurate real-time representation of a system's state and trends [18,39]. However, this is often not the case in practice [40], due to a lack of incentive,

[5]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstrasse 4, 04103 Leipzig, Germany
[6]Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35 (Survontie 9C), FI-40014 Jyväskylä, Finland
[7]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, P.O. Box 65, Helsinki 00014, Finland
[8]Department of Biology, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim N-7491, Norway
[9]Helmholtz Centre for Environmental Research – UFZ, Department of Ecological Modelling, Permoserstr. 15, 04318 Leipzig, Germany
[10]University of Oslo, Natural History Museum, Sars gate 1, NO-0562 Oslo, Norway
[11]Global Biodiversity Information Facility – GBIF Secreteriat, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark
[12]University of Potsdam, Plant Ecology and Nature Conservation, Am Mühlenberg 3, 14476 Potsdam, Germany

*Correspondence:
dag.endresen@nhm.uio.no
(D. Endresen).
[@]Twitter: @Koendekoning (K. de Koning), @dschigel and @GBIF (D. Schigel).

Box 1. Examples of DTs and their main components

To illustrate what DTs are, two examples are shown in Figure I. Figure IA concerns a DT used in weather forecasting, and Figure IB shows an application in ecology: a DT on bird migration[viii]. What they have in common is that they use a continuous stream of sensor data to keep the DT synchronised with the real world, and models to process the data and make predictions.

Sensor data can come from a wide variety of sources, requiring significant data processing and assimilation as well as adequate cyberinfrastructure (computational power, access to APIs, data storage facilities, etc.). Input data for the DTs can come from direct observations of the phenomenon (such as temperature measurements in weather forecasts, and bird observations in migration forecasts), as well as auxiliary data sources and connections with other DTs (such as wind forecasts affecting migration routes).

The required models can be data-driven statistical models, but ideally they are mechanistic models (e.g., physics-based) that are compatible with real-world processes, in order to make DTs generalisable beyond the data on which they are trained. Hence, domain knowledge is essential for DTs to reliably represent their real-world counterparts. Typically, data-driven DTs are heavier in their data requirement and include more AI components, whereas mechanistic DTs require profound mathematical specifications. Conversely, DTs can also provide fundamental new insights on the systems they represent, which broadens our domain knowledge.

As a decision-support tool, DTs have a strong end-user component, forcing the modeller to think about how output is visualised and interpreted by end-users. Users may also be allowed to interact with the DTs: for example, by providing feedback on the model predictions, or by collecting input data as is the case in the bird migration DT.

Besides updating with real-time data, DTs can also be fed with historical data for calibration. The migration DT, for example, uses historical GPS telemetry and weather data to calibrate the models that predict how cranes respond to wind conditions.

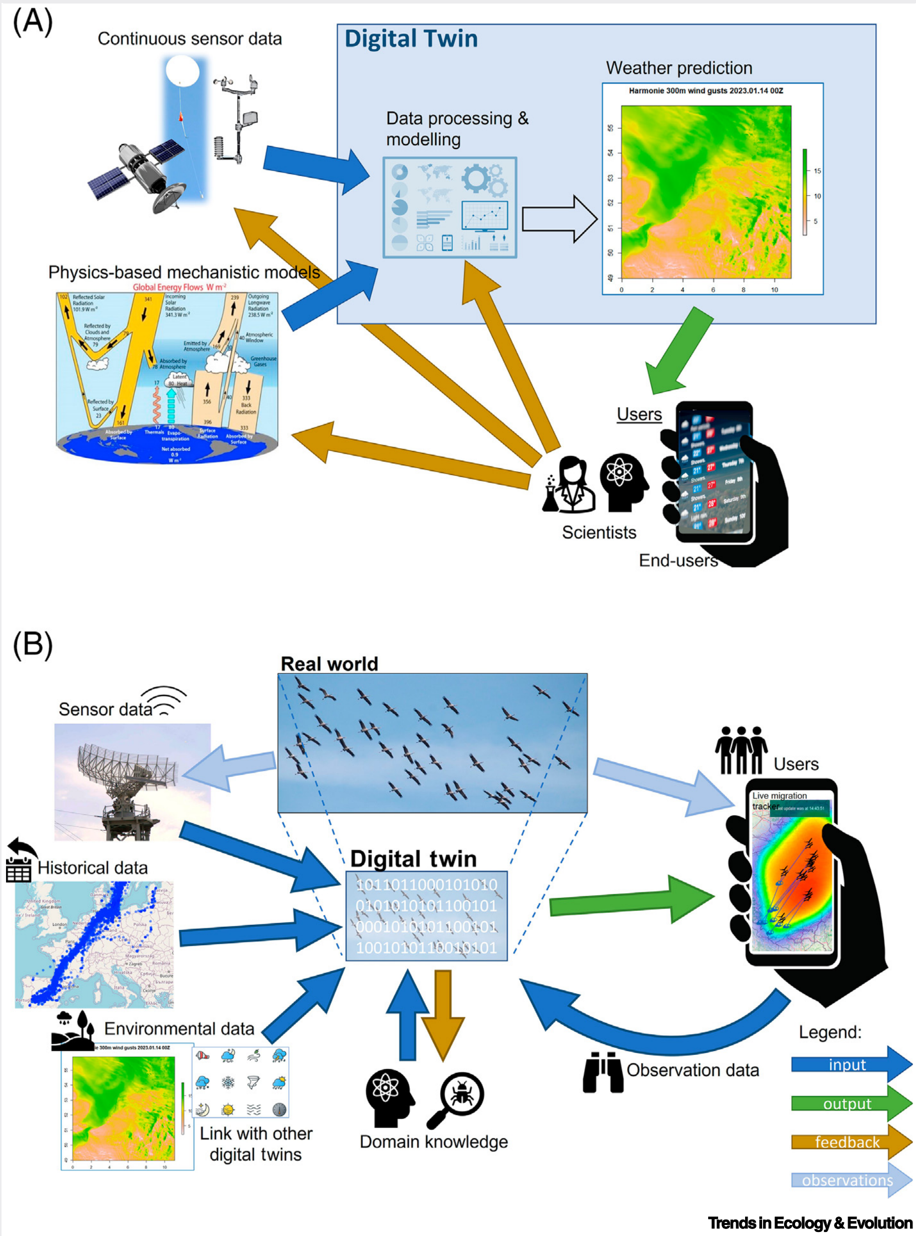


Figure I. Examples of digital twins (DTs). Established physics-based mechanistic models are used for weather forecasting (A), but due to their inherent nonlinearities, their predictions quickly become too uncertain to be useful. Model predictions are therefore continuously updated with high-resolution sensory data. (B) There is no mechanistic model for predicting crane (*Grus grus*) migration. Instead, in an empirical model sensory data from migrating cranes are combined with environmental data such as wind directions. In addition to such data, the output of environmental DTs could be used as well. Again, because predictions of routes taken by cranes quickly become uncertain, citizen science sighting data are used to update the output of the crane DT. In both examples, the design, model, and workflows underlying the DT are driven by user demand for relevant and reliable predictions.

resources, and skills to collect data in a proper way [39], the inability to properly process and interpret data [10,41–43], and lack of time and resources to plough through a vast amount of complex literature [34,40], resulting in an increasing gap between scientific insights and conservation practices [36].

We see a need for more real-time, informative, and iterative ecological forecasting that would make ecology more relevant for decision-making [18–21]. A major question is how we can make conservation efforts more effective. This question operates on scales ranging from the operational level (individual species protection and landscape management [44]) to the global political level addressing global biodiversity decline [45]. DTs may provide the platform to address this question by combining the science and the data, adding context and meaning to ecological data to inform decision-making in biodiversity conservation.

Opportunities for ecology

DTs will likely increase in popularity over the next decades, which is why attention should focus on how to capitalise on the strengths of digital twinning for ecology. We list here a few characteristics of DTs that could benefit the field: (i) DTs could help measure the state of nature with minimal lags or even in real-time [4], thereby adding a dynamic component to commonly static modelling practices such as species distribution maps, vegetation maps and habitat suitability maps, (ii) real-time data flows and continuous updating of models allow conservationists to detect trends and the need for intervention measures at an early stage; environmental drivers are changing so fast that near real-time feedbacks are critical when managing ecosystems [18], (iii) ecosystem trends can be linked with environmental conditions and anthropogenic pressures that are inherently part of the DT inputs; (iv) DTs can help demonstrate the effectiveness of intervention measures, either reactively by keeping track of trends before and after interventions, or proactively by providing a safe virtual environment for simulations and scenario development in which interventions can be tested before they are implemented in the real world, and (v) DTs can help identify uncertainties, information gaps, and knowledge gaps in real ecosystems [46–48], and provide feedback to maximise information uptake in data collection. This is not just relevant for decision-making in global change mitigation, but it also sheds light on scientific knowledge gaps that are still unaddressed [49]. Highlighting information gaps gives insight into where to allocate resources for data collection or fundamental research in ecology.

While most of these aspects are already common practice in ecology, the current methods typically do not address all these aspects simultaneously as do DTs (Figure 1). The strength of DTs is in the combination of domain knowledge and much more diverse data. DTs are based on mechanistic models of a system, rather than just data, machine learning, and high-tech sensors [50]. Their success depends in a large part on the modellers' ability to combine data with sound ecological theories and expert knowledge [10,51–53].

Inherent challenges of DTs in ecology

While the opportunities of DTs are plentiful, there are obviously many challenges and misconceptions in their implementation (Box 2). Some of those challenges are comparable with DTs in engineering and manufacturing, and others are unique to ecology.

First, there are no common methods, standards, or norms about how to build the software and workflows for DTs [54]. A lack of standards and protocols means that researchers need to determine how to build DTs from scratch on their own, which slows down progress and causes frustrations among modellers and stakeholders. Nonetheless, with DTs rapidly gaining popularity, it is likely that those standards will be developed in the near future [55], as is currently the case in other domains such as the ISO 23247 framework for digital twins in manufacturing.

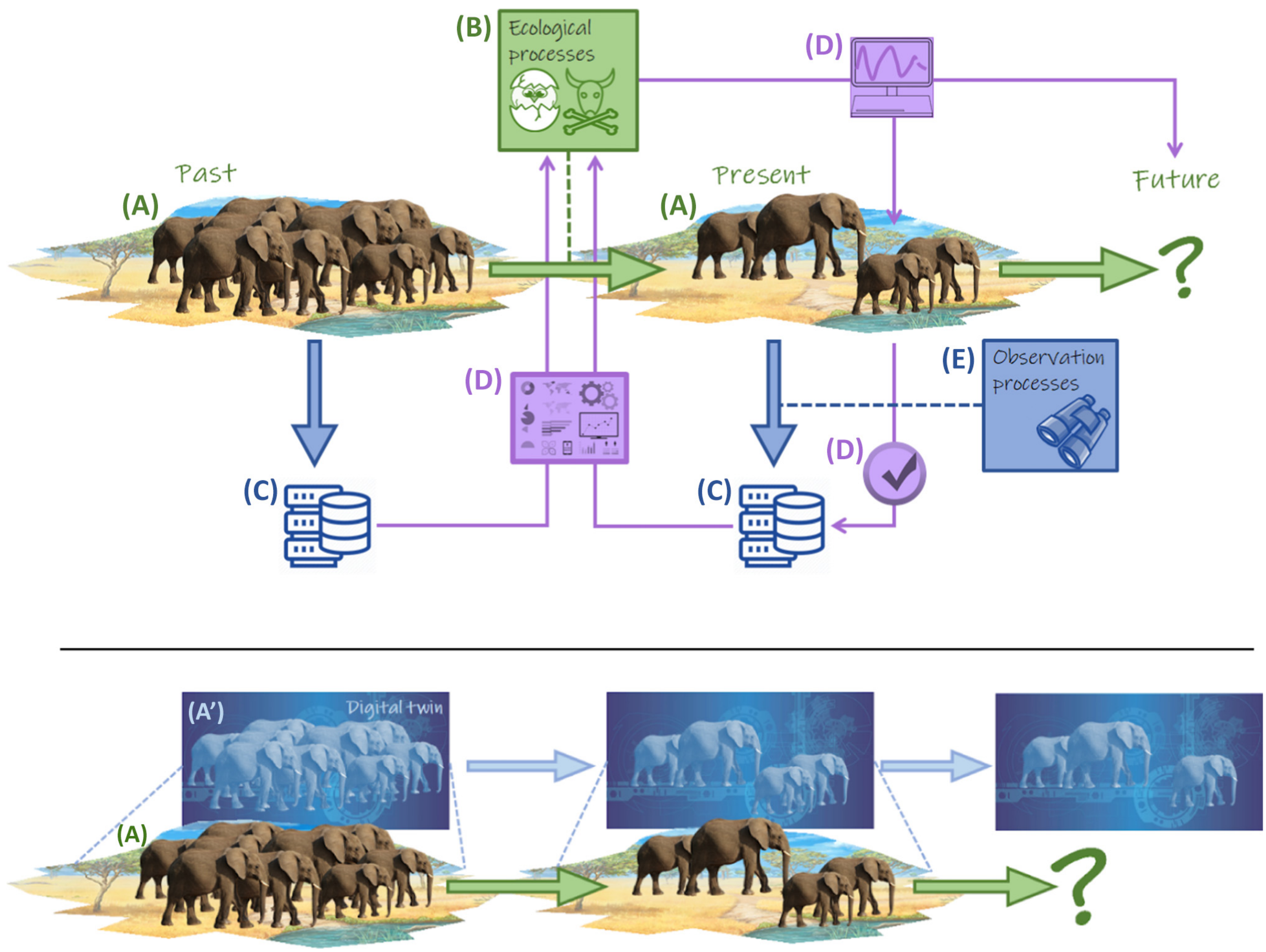


Figure 1. Comparison between ecological research in the status quo (top) and with digital twins (DTs) (bottom). (A) Changes in the state of (part of) an ecosystem (e.g., an elephant population over time), caused by (B) ecological processes such as mortality and birth rate. The blue arrows illustrate the collection and storage of (C) observation data such as aerial surveys, camera traps, or citizen science observations. Typically, these data are processed and analysed a single time on a case-study basis. Common steps in ecological analysis (D) include statistical analysis, models, and validation to understand system dynamics, drivers, and extrapolations, as well as simulations to project desired and undesired future states. Inevitably, representations of ecological processes are simplified in our models, leading to imperfect model predictions. Furthermore, inherent to the sampling strategy, there are observation processes at play (E) that cause biases in the data, such as imperfect sampling, detection errors, and observer-specific biases [58]; these need to be controlled for. The DT (A') functions as a live model that exists parallel to its real-world counterpart. Ideally, DTs aim to integrate all elements (B,C,D,E) to mimic the behaviour of (A) as much as possible by continuously updating models and input data, while at the same time highlighting the uncertainties that are produced through model simplifications, incomplete data and knowledge, and observational biases.

Second, a related challenge concerns the protocols of data collection and storage. Different data collection, harmonisation, and storage protocols lead to different types of sensors, different data, and different formats [54]. Challenges remain to integrate data streams from different research data infrastructures [56], with standards based on demonstrating successes and potential [26]. Furthermore, the required input data are often spread among different owners, with consequential restrictions to data accessibility and use [32].

Third, there is the question of what to include and what not to include in the DT [54], a question that needs to be addressed in all models. DTs can be made on, for example, species

Box 2. Four major misunderstandings about DTs

We summarise four likely misunderstandings about what DTs are, in an attempt to manage expectations and avoid misconceptions.

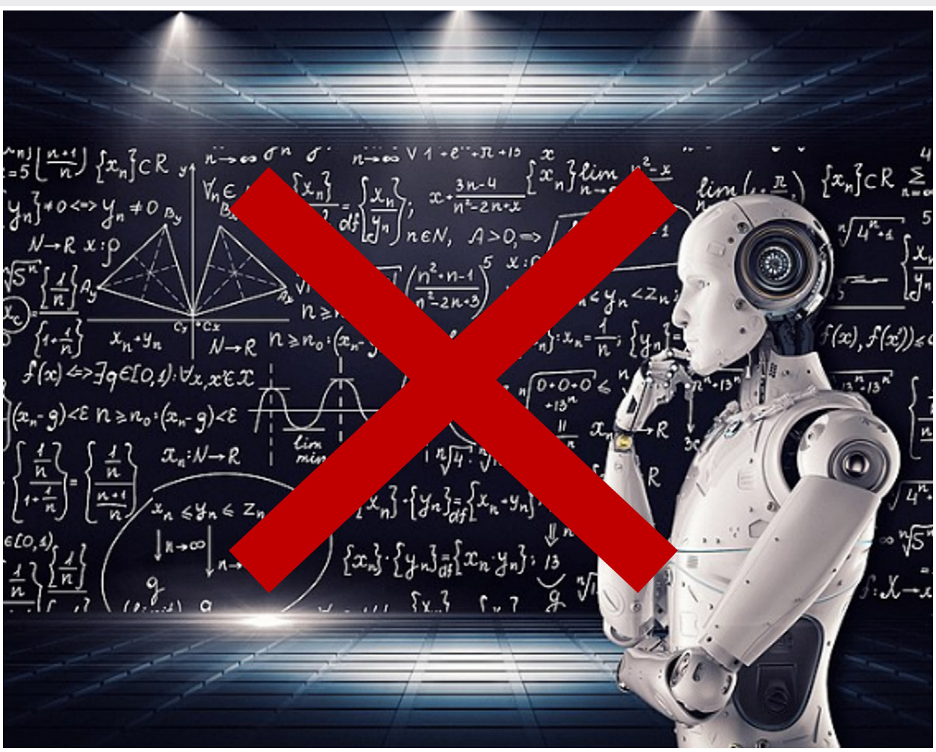
A DT is not just …

... artificial intelligence (AI)/machine learning (Figure I). DTs often contain elements of machine learning and AI to process a continuous stream of data, but they are also based on a mechanistic understanding of the systems they represent. Above all, DTs require a substantial amount of domain knowledge in order for the representations to be meaningful.

… a large database of integrated or connected datasets (Figure II). DTs contain a lot of empirical data, but the key is that DTs add context to data and give data meaning within that context by combining data with modelling and simulations. This trait gives DTs their unique ability to integrate heterogeneous data sources [62,63].

... another word for model (Figure III). Many different types of model exist (static/dynamic/explanatory/descriptive/statistical, etc.), and DTs might include several modelling approaches/techniques addressing different system entities and purposes. What all DTs have in common is that they are by definition dynamic, and they are unique from other forms of modelling in that they are continuously updated with data, thereby representing the state of something in current time. Furthermore, DTs allow users to interactively explore system dynamics and scenarios, and not just tie to the input of data, but also to their use in decision-making.

… a big model of everything (Figure IV). Although in some cases it may be desirable to make DTs interoperable and to develop very large-scale DTs of global systems [64,65], it is not a necessity in order for DTs to be useful for their purpose. As with any other type of model, DTs are simplified representations of a specific part of reality (entity, system, or process) with predefined boundaries depending on their aim and scope.



Trends in Ecology & Evolution

Figure I. A digital twin is not just artificial intelligence (AI)/machine learning. Image source: wikimedia commons.

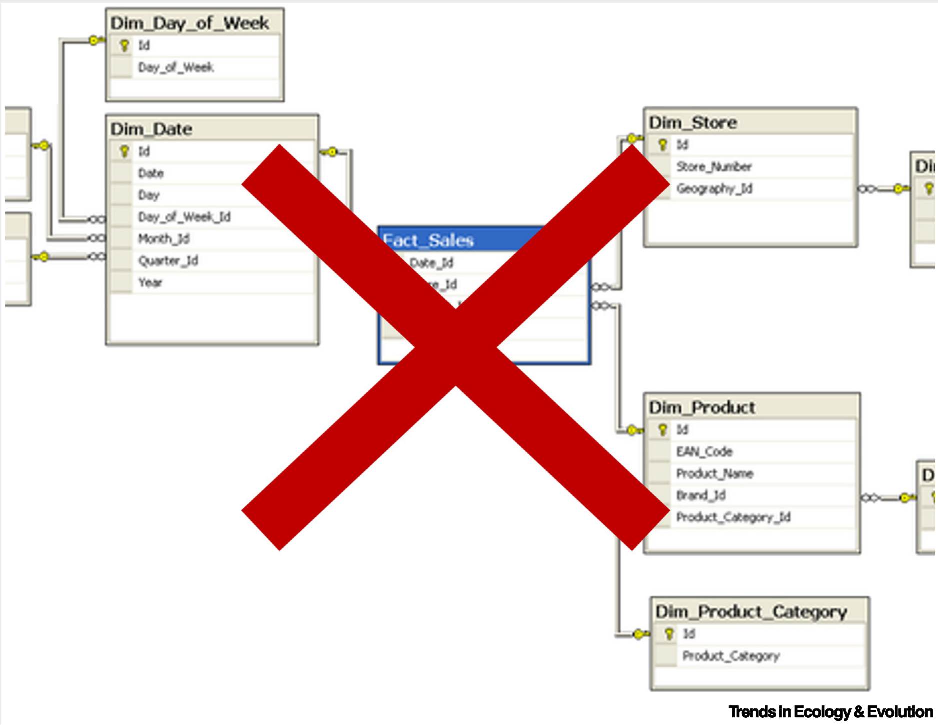


Figure II. A digital twin is not just a large database of integrated or connected datasets. Image source: wikimedia commons.

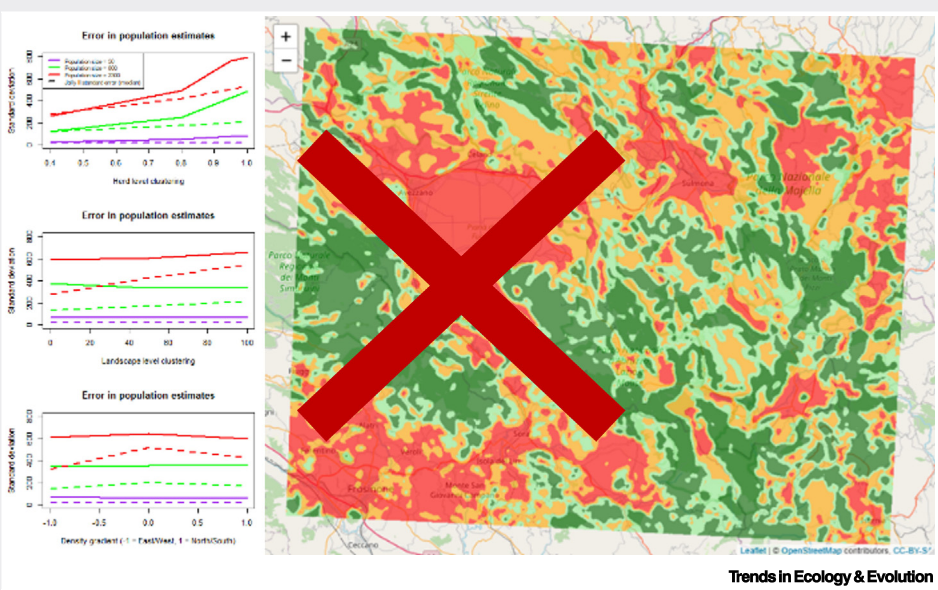


Figure III. A digital twin is not just another word for model.

Figure IV. A digital twin is not just a big model of everything.

distributions, specific behaviours or traits, genetic diversity, ecosystems, communities, and habitats, depending on the underlying research purpose. A temporal resolution has to be chosen that fits with the process being modelled, which again must be done in any dynamic model. The 'real-time' representation needs to reflect the speed of the system under scrutiny. Many engineering DTs are synchronised in the order of milliseconds, whereas many ecological processes are considerably slower. Time intervals are also limited by data collection frequency (including laborious field work and time to digitise data) and uncertainties about processes at finer spatiotemporal scales. The real-time definition of DTs for ecology is more along the lines of 'synchronised at a specific rate suitable for the purpose'. Previous publications may have raised expectations that DTs could create a single 'supermodel' capable of replicating the Earth's entire natural system in great detail [1,50]. This contradicts the basic philosophy of models being a representation of reality with regard to a concrete problem and selecting the degree of complexity and detail on the basis of this problem. Knowing all the details about the functioning of entities across scales is not necessary, and impossible anyway. Hence, as with any model, modellers should consider which level of abstraction, scale, and spatiotemporal resolution fits best with the process of interest and the intended purpose of the DT.

Fourth, the entities of which DTs are made in engineering are human-made, and therefore modellers have by definition a good understanding of these entities (or systems). Modellers know how engineering entities work and, in principle, should be able to predict how the systems will respond to certain environmental conditions and disturbances. By contrast, in ecology we lack such comprehensive understanding. The bigger and more complex the systems get, the harder it is to comprehend the system's structure and functioning. Furthermore, there are issues of heterogeneity and variability in ecological systems that are nontrivial compared to engineering. When you do not know what behaviour to expect, or what drives the behaviour of a system/entity, it is a problem if you want to build a twin that mimics the behaviour of its real-world counterpart. However, trying to capture the system/entity in DTs opens up new avenues for ecological research by shedding light on the gaps in our knowledge about the systems we model.

Fifth, DTs rely on a continuous stream of data for synchronisation with their real-world twin, which inevitably requires collecting data either by automated sensors and classification algorithms or by large numbers of volunteers. Hence, data quality is inherently limited by the capacity of volunteers or sensors and the underlying artificial intelligence (AI) to correctly identify species and taxa and other features of interest, and is dependent on people following the expected data collection protocols [57]. Readily available data (e.g., GBIF[vi]) are frequently biased in space, time, and taxonomical scope in terms of what is measured when and where [58–60]. There is some danger that the existing bias of human field observations towards a few preferred taxa or already well-sampled regions will be further enhanced by the fact that underlying AI can only identify taxa which are already well represented in the reference libraries.

The first two challenges are similar to those in engineering DTs. How these challenges are addressed here can serve as inspiration for ecology. The latter three challenges are familiar in ecology. We would like to stress that addressing those challenges for DTs requires work in the coming years when more DTs will be developed for ecology. And to facilitate a smooth start, it is important to manage false expectations, misunderstandings, and misinterpretations of DTs (Box 2) [61].

Concluding remarks: start building DTs for ecology

There never was a stronger need for DTs in ecology, and we never had better state-of-the-art in monitoring, data science, and modelling to exploit DTs. Unfortunately, we do not yet have any conclusive answers to all the abovementioned challenges. It requires more research and, above all, more experience with building DTs in ecology (see Outstanding questions). We therefore urge modellers and ecologists to start developing DT prototypes for ecology in order to learn from those experiences: have first demos, summarise experiences into recommendations, get realistic estimates of the effort and resources needed to establish DTs, and identify bottlenecks.

Modellers: take your static ecological models and start exploring how to make these models more dynamic; feed them with continuous data flows accessible through application programming interfaces (APIs), or connect with other DTs of the natural environment. Empiricists: join modellers and software engineers and discover the possibilities of DTs to improve ecological monitoring, system understanding, and timely decision-making. End-users: (e.g., policymakers and conservationists) express your support needs for decision-making in terms of what an ecological DT could in the ideal case provide. Ecologists, modellers, and end-users: develop your DTs explicitly with the application in mind and in close consultation with each other. This will allow developers and end-users to interact with DT outputs, provide up-to-date insights on states, trends, and drivers, information gaps, and relevant scenarios that support decision-making [17], thereby engaging stakeholders more closely with science-driven insights in nature, biodiversity, and ecosystems.

Outstanding questions

Which software architectures, protocols, and workflows could be useful in building DTs for ecology? We expect that standards will soon be developed for DTs in other fields, which can be used as guidelines for building DTs in ecology. Testing these standards against the unique challenges for ecology may lead to new protocols specifically for ecology.

What scales, resolutions, and scopes are appropriate for DTs in ecology? Future research on DTs should cover to what extent the real-time aspect is feasible or even necessary to address in ecology. The same accounts for the level of detail and complexity of the system elements that are included in the DTs.

How to control for biases in DT input data, such as sampling biases and observation biases? DTs rely heavily on automated streams of input data for synchronisation, which means that biases need to be controlled for before data are processed by the DTs. Controlling for biases ex-ante adds a new challenge that needs further research; as common in ecological modelling, this can be done ex-post.

How to communicate model outputs in order to give adequate up-to-date insights for conservation decision-making? Complex model outputs and uncertainties need to be visualised understandably in order to make DTs useful as a decision-support tool, which rarely receives the attention it deserves in ecological modelling. Involving end-users is extra-important in DT development, given that many DTs operate on the applied domain by empowering end-users to better understand and monitor complex systems.

Acknowledgments

This study received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057437 (BioDT project, https://doi.org/10.3030/101057437) and No 101060954 (NATURE-FIRST, https://doi.org/10.3030/101060954). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Declaration of interests

No interests are declared.

Resources

[i]https://biodt.eu/

[ii]www.geobon.org

[iii]www.elter-ri.eu

[iv]www.saeon.ac.za

[v]www.tern.org.au

[vi]www.gbif.org

[vii]www.lifewatch.eu

[viii]https://sensingclues.org/craneradar

References

1. Bauer, P. *et al.* (2021) A digital twin of Earth for the green transition. *Nat. Clim. Chang.* 11, 80–83
2. Singh, M. *et al.* (2021) Digital twin: origin to future. *Appl. Syst. Innov.* 4, 36
3. Singh, M. *et al.* (2022) Applications of digital twin across industries: a review. *Appl. Sci.* 12, 5727
4. Jones, D. *et al.* (2020) Characterising the digital twin: a systematic literature review. *CIRP J. Manuf. Sci. Technol.* 29, 36–52
5. Nativi, S. *et al.* (2021) Digital ecosystems for developing digital twins of the earth: The destination earth case. *Remote Sens.* 13, 2119
6. Voosen, P. (2020) Europe builds 'digital twin' of Earth to hone climate forecasts. *Science* 370, 16–17
7. Effrosynidis, D. *et al.* (2022) Exploring climate change on Twitter using seven aspects: stance, sentiment, aggressiveness, temperature, gender, topics, and disasters. *PLoS ONE* 17, 1–19
8. Farley, S.S. *et al.* (2018) Situating ecology as a big-data science: current advances, challenges, and solutions. *Bioscience* 68, 563–576
9. Lahoz-Monfort, J.J. and Magrath, M.J. (2021) A comprehensive overview of technologies for species and habitat monitoring and conservation. *BioScience* 71, 1038–1062
10. Tuia, D. *et al.* (2022) Perspectives in machine learning for wildlife conservation. *Nat. Commun.* 13, 1–15
11. Tao, F. *et al.* (2018) Digital twin in industry: state-of-the-art. *IEEE Trans. Industr. Inform.* 15, 2405–2415
12. Liu, Z. *et al.* (2018) The role of data fusion in predictive maintenance using digital twin. *AIP Conf. Proc.* 1949, 1–6
13. Kays, R. *et al.* (2020) Born-digital biodiversity data: millions and billions. *Divers. Distrib.* 26, 644–648
14. Kamel Boulos, M.N. and Zhang, P. (2021) Digital twins: from personalised medicine to precision public health. *J. Pers. Med.* 11, 745
15. Neethirajan, S. and Kemp, B. (2021) Digital twins in livestock farming. *Animals* 11, 1–14
16. Pylianidis, C. *et al.* (2021) Introducing digital twins to agriculture. *Comput. Electron. Agric.* 184, 105942
17. Bauer, P. *et al.* (2021) The digital revolution of Earth-system science. *Nat. Comput. Sci.* 1, 104–113
18. Dietze, M.C. *et al.* (2018) Iterative near-term ecological forecasting: needs, opportunities, and challenges. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1424–1432
19. McIntire, E.J. *et al.* (2022) PERFICT: a re-imagined foundation for predictive ecology. *Ecol. Lett.* 25, 1345–1351
20. Mouquet, N. (2015) Predictive ecology in a changing world. *J. Appl. Ecol.* 52, 1293–1310
21. Petchey, O.L. *et al.* (2015) The ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.* 18, 597–611
22. Nilsson, R.H. *et al.* (2022) Introducing guidelines for publishing DNA-derived occurrence data through biodiversity data platforms. *Metabarcoding Metagenom.* 6, 239–244
23. Whytock, R.C. *et al.* (2021) Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods Ecol. Evol.* 12, 1080–1092
24. Gibb, R. *et al.* (2018) Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* 10, 169–185
25. Hardisty, A. *et al.* (2022) The specimen data refinery: a canonical workflow framework and FAIR digital object approach to speeding up digital mobilisation of natural history collections. *Data Intell.* 4, 320–341
26. Heberling, J.M. *et al.* (2021) Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2018093118
27. Chandler, M. *et al.* (2017) Contribution of citizen science towards international biodiversity monitoring. *Biol. Conserv.* 213, 280–294
28. Mäder, P. *et al.* (2021) The Flora Incognita app – interactive plant species identification. *Methods Ecol. Evol.* 12, 1335–1342
29. Fink, C. (2021) Digital conservation: novel methods and online data to address the biodiversity crisis. Department of Geosciences and Geography A. http://urn.fi/URN:ISBN:978-951-51-6583-1
30. Balch, J.K. *et al.* (2020) NEON is seeding the next revolution in ecology. *Front. Ecol. Environ.* 18, 3
31. Mirtl, M. *et al.* (2018) Genesis, goals and achievements of long-term ecological research at the global scale: a critical review of ILTER and future directions. *Sci. Total Environ.* 626, 1439–1462
32. Ahumada, J.A. *et al.* (2020) Wildlife insights: a platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. *Environ. Conserv.* 47, 1–6
33. Fer, I. *et al.* (2021) Beyond ecosystem modeling: a roadmap to community cyberinfrastructure for ecological data-model integration. *Glob. Chang. Biol.* 27, 13–26
34. Fabian, Y. *et al.* (2019) How to close the science–practice gap in nature conservation? Information sources used by practitioners. *Biol. Conserv.* 235, 93–101
35. Junker, J. *et al.* (2020) A severe lack of evidence limits effective conservation of the world's primates. *BioScience* 70, 794–803
36. Kadykalo, A.N. *et al.* (2021) Bridging research and practice in conservation. *Conserv. Biol.* 35, 1725–1737
37. Salafsky, N. *et al.* (2019) Defining and using evidence in conservation practice. *Conserv. Sci. Pract.* 1, 1–15
38. Sutherland, W. *et al.* (2004) The need for evidence-based conservation. *Trends Ecol. Evol.* 19, 305–308
39. Fox, H.E. *et al.* (2017) Generating actionable data for evidence-based conservation: the global center of marine biodiversity as a case study. *Biol. Conserv.* 210, 299–309

40. Downey, H. *et al.* (2022) Principles for the production of evidence-based guidance for conservation actions. *Conserv. Sci. Pract.* 4, 1–9
41. Beardsley, T.M. *et al.* (2018) Addressing biological informatics workforce needs: a report from the AIBS Council. *BioScience* 68, 847–853
42. Hamel, S. *et al.* (2013) Towards good practice guidance in using camera-traps in ecology: influence of sampling design on validity of ecological inferences. *Methods Ecol. Evol.* 4, 105–113
43. Ellwood, E.R. *et al.* (2020) Biodiversity science and the twenty-first century workforce. *BioScience* 70, 119–121
44. Sutherland, W.J. *et al.* (2009) One hundred questions of importance to the conservation of global biological diversity. *Conserv. Biol.* 23, 557–567
45. Lynch, K.E. and Blumstein, D.T. (2020) Effective conservation. *Trends Ecol. Evol.* 35, 857–860
46. Medlyn, B.E. *et al.* (2016) Using models to guide field experiments: a priori predictions for the CO_2 response of a nutrient- and water-limited native Eucalypt woodland. *Glob. Chang. Biol.* 22, 2834–2851
47. Norby, R.J. *et al.* (2016) Model–data synthesis for the next generation of forest free-air CO_2 enrichment (FACE) experiments. *New Phytol.* 209, 17–28
48. Rocchini, D. *et al.* (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Prog. Phys. Geogr.* 35, 211–226
49. Christie, A.P. *et al.* (2021) The challenge of biased evidence in conservation. *Conserv. Biol.* 35, 249–262
50. Purves, D. *et al.* (2013) Time to model all life on Earth. *Nature* 493, 295–297
51. Tao, F. and Qi, Q. (2019) Make more digital twins. *Nature* 573, 490–491
52. Gurumurthy, S. *et al.* (2018) Exploiting data and human knowledge for predicting wildlife poaching. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 1–8
53. Karpatne, A. *et al.* (2017) Theory-guided data science: a new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.* 29, 2318–2331
54. Reichstein, M. *et al.* (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204
55. Qamsane, Y. *et al.* (2021) A methodology to develop and implement digital twin solutions for manufacturing systems. *IEEE Access* 9, 44247–44265
56. Jeffery, K. *et al.* (2021) Not ready for convergence in data infrastructures. *Data Intell.* 3, 116–135
57. Brown, E.D. and Williams, B.K. (2019) The potential for citizen science to produce reliable and useful information in ecology. *Conserv. Biol.* 33, 561–569
58. Meyer, C. *et al.* (2015) Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* 6, 8
59. Meyer, C. *et al.* (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* 19, 992–1006
60. Mollenhauer, H. *et al.* (2018) Long-term environmental monitoring infrastructures in Europe: observations, measurements, scales, and socio-ecological representativeness. *Sci. Total Environ.* 624, 968–978
61. Wright, L. and Davidson, S. (2020) How to tell the difference between a model and a digital twin. *Adv. Model. Simul. Eng. Sci.* 7, 1–13
62. Durão, L.F. *et al.* (2018) Digital twin requirements in the context of industry 4.0. In *IFIP International Conference on Product Lifecycle Management*, pp. 204–214, Springer
63. Uhlemann, T.H.J. *et al.* (2017) The digital twin: demonstrating the potential of real time data acquisition in production systems. *Procedia. Manuf.* 9, 113–120
64. Kuruppuarachchi, P. *et al.* (2022) An architecture for composite digital twin enabling collaborative digital ecosystems. In *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 980–985, IEEE
65. Platenius-Mohr, M. *et al.* (2020) File- and API-based interoperability of digital twins by model transformation: an IIoT case study using asset administration shell. *Future Gener. Comp. Sy.* 113, 94–105