TwinEco: A Unified Framework for Dynamic Data-Driven Digital Twins in Ecology

Authors: Taimur Khan[1], Koen de Koning[2], Dag Endresen[3], Desalegn Chala[3], Erik Kusch[3]

[1] Community Ecology, Helmholtz Centre for Environmental Research - UFZ, Theodor-Lieser-Straße 4, 06120 Halle (Saale), Germany – taimur.khan@ufz.de

[2] Wageningen University and Research, Earth Systems and Global Change Group, Droevendaalsesteeg 4, 6708 PB Wageningen, Netherlands – koen.dekoning@wur.nl

[3] Natural History Museum, University of Oslo, Sars' gate 1, 0562 Oslo, Norway – dag.endresen@nhm.uio.no, d.c.gelete@nhm.uio.no, erik.kusch@nhm.uio.no

Keywords: Digital Twin, Ecological Modelling, Research Infrastructure, Framework, TwinEco, Data, Modelling, DDDAS, Design.

Corresponding Author: Taimur Khan, Mail @ taimur.khan@ufz.de, Telephone @ +49 341 235 5532

High-Resolution Figures:
https://drive.google.com/drive/folders/1xY2AIT_jRshqsZrvBBxJl3SAVUlgSoOS?usp=sharing

1. Abstract

The burgeoning interest in digital twin (DT) technology presents a transformative potential for ecological modelling, offering new ways to model the complex dynamics of ecosystems. This paper introduces the TwinEco framework, designed to mitigate fragmentation in the development and deployment of DT applications in ecology. Compared to traditional modelling frameworks, TwinEco emphasises modularity and flexibility by introducing “layers” and “components” for DTs, accommodating diverse ecological applications without necessitating the deployment of all components. This modular approach ensures adaptability and scalability, promoting interoperability and integration with broader initiatives like Destination Earth.

Digital twins in ecology offer significant advancements over traditional approaches by explicitly modelling changing processes and states over time, integrating extensive data, and enabling real-time feedback loops by actuating events or policies in the “real-world”. The framework's capacity to adjust to changing environmental conditions enhances its predictive accuracy and responsiveness. This paper highlights the necessity for a unified framework to prevent divergent interpretations and ensure the interoperability of DT applications across ecological domains. Future recommendations include expanding case studies to demonstrate the framework's applicability, assessing the potential of Dynamic Data-Driven Application Systems (DDDAS) paradigm within ecological DTs, and exploring interactions between components to optimise performance. Emphasising model-data fusion and fostering a shared terminology within the ecological community are crucial for the framework's success. TwinEco aims to provide a robust foundation for ecological digital twins, enabling timely, data-driven decision-making to address global environmental challenges.

2. Introduction

In the rapidly evolving realm of environmental change and technology, just-in-time modelling and timely mitigational or adaptive actions become increasingly relevant. Ecological systems are dynamic, complex, and highly responsive to environmental factors, making them challenging to model and manage effectively (Carpenter and Brock, 2004; Donohue et al., 2016; Vermeiren et al., 2020). Understanding, predicting, effectively managing, and aligning human activities with these intricate systems are paramount for mitigating the profound impacts of human activities on the environment (Brown and Williams, 2015; Ruckelhaus et al., 2020; Newton, 2016). In response, ecological research has witnessed a remarkable evolution over the years, driven by technological advancements and the growing awareness of environmental issues. Most traditional approaches often rely on models incorporating observational data at a given point in time without subsequent re-examination as additional data becomes available thus producing static outputs (Zurell et al., 2021). While often informative, these traditionally static quantitative approaches fall short of capturing the dynamic, interconnected, and responsive nature of ecosystems (Damgaard, 2019). In response to these challenges, a transformative concept offering a promising paradigm shift to enact more informative quantification of ecosystem dynamics has emerged — the "digital twin”.

Digital twin systems, rooted in the realm of cyber-physical systems, are virtual replicas of physical entities or processes that capture real-time data and behaviour of system components, enabling dynamic simulation, monitoring, and analysis (Segovia and Garcia-Alfaro, 2022). Initially, these systems were designed for applications such as manufacturing, where they facilitate predictive maintenance, quality control, and process optimization (Wu and Li, 2021; Onaji et al., 2022). However, the potential for digital twins extends well beyond industry boundaries and is rapidly finding its place in the study and conservation of ecological systems (de Koning et al., 2023). This adaptation of twins from industry cannot be technically verbatim as ecological modelling presents its own unique challenges. Whereas traditional ecological modelling approaches often face limitations in their ability to update and adapt to represent the intricacies of natural ecosystems, respond to (near) real-time data, and adapt to changing conditions, digital twins are designed to iteratively update the process knowledge they create as well as the data products they incorporate and produce. Consequently, there is a growing recognition within the ecological research community that a paradigm shift is required to bridge the gap between observation, understanding, and action. Digital twin systems offer a promising avenue to achieve this (Sharef et al., 2022).

The adaptation of digital twin systems to ecology is a burgeoning field that presents both opportunities and complexities (Trantas et al., 2023; de Koning et al., 2023). This process has already been explored in several applications of ecologically relevant DTs[1][2], but no unified framework has been established governing structure and behaviour of digital twins for ecological research and management (Golivets et al., 2024). Consequently, while the current implementation of DT frameworks in ecology is still in its infancy, concepts and understandings of how a digital twin ought to be established in this study domain have already become fragmented. To harmonise existing variation among ecological DT initiatives and rein in the potential for increasing confusion over the application of DTs to ecological applications, a unified well-structured design framework for ecological DTs is essential. This framework should encompass not only the technical aspects of data integration and modelling but also ecological expertise, ensuring that the resulting digital twins represent the intricacies of natural systems.

Conceptualisation of such a design framework encompassing the enormous variety of potential applications for DTs in ecological research/management is non-trivial as a host of research questions and considerations must be addressed. These include: Firstly, what should be the fundamental components of a design framework for dynamic data-driven digital twins in ecology? Given different scales and resulting management pathways for ecosystems, a design framework for ecological DTs must be malleable to account for the breadth of potential applications and resulting management and observation processes. Secondly, how can digital twins be structured in ecology? For example, the real-time data ingestion characterising industrial applications of DTs are usually not feasible for ecological research. Thirdly, how can ecological data from diverse sources be integrated into this framework seamlessly? Data products and types supporting ecological research are as

[1] https://sensingclues.org/craneradar
[2] https://biodt.eu/use-cases

diverse as the systems and processes studied by contemporary quantitative ecology. Interoperability of such disparate data poses a serious challenge to the establishment of digital twins at scale. Resolving these research questions will unlock hitherto unrealised potential and opportunities of digital twinning for ecological research, management, and governance.

In addressing the aforementioned challenges and opportunities we propose a comprehensive design framework, called TwinEco, for dynamic data-driven digital twins in ecology. To do so, we (1) develop a structured framework that outlines the essential layers and components involved in the creation of dynamic data-driven digital twins for ecological systems, (2) investigate data integration and data/model fusion techniques that can accommodate diverse and dynamic data sources, including remote sensing, sensor networks, manually processed data and ecological monitoring data, (3) assess the potential of Dynamic Data-Driven Application Systems (DDDAS) as a system design approach to ecological digital twins to automate system understanding, decision-making, mitigation, and management strategies, and (4) make future recommendations on what is needed to steer the process of digital twinning in ecological use cases. The resulting framework has the potential to leapfrog the way ecological systems are studied, monitored, and managed. The TwinEco framework aims to empower researchers and practitioners to effectively model and understand ecosystems. Ultimately, this knowledge can guide conservation efforts, inform sustainable land use decisions, and contribute to the preservation of biodiversity and ecological balance.

3. Methodology

1. TwinEco Layers

In order to define concrete components that drive a digital twin, we first have to define the dynamics of the twin with its physical counterpart. The dynamics between a digital twin and its physical counterpart involve a complex interplay of data exchange, feedback loops, and synchronisation. All these aspects can be categorised as the “layers” of a digital twin. Each layer has its own individual mechanics, but there also exists mechanics of interaction between the layers. We will be discussing these mechanics as well. Similar methods of layering have already been adapted in existing digital twin approaches that do not necessarily follow any common framework, but share commonalities in the design patterns of the underlying systems (Buonocore et al., 2022; Fissore et al., 2023; Sougioultzogloua and Cook, 2023; Morlot et al., 2024).

Physical State, S

In the context of digital twins, the concept of state spaces refers to the representation of the current or past states of the Physical and Digital Twins. The Physical State (*S*) space encompasses all the relevant variables and parameters that describe the parameterized state of the physical asset at a given time in the Physical Twin (Malik et al., 2020) (Figure 1). These variables may include physical properties, environmental conditions, operational parameters, and any other factors that

influence the behaviour of the physical asset. The Physical State changes over time (t) due to natural or unnatural events. At each time-step (t), the Physical State changes as well.

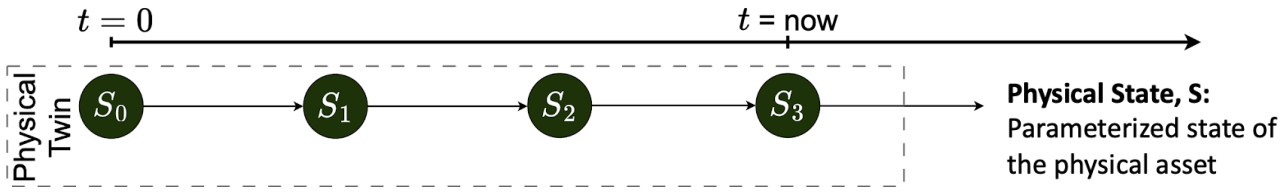

Figure 1: Parameterized state of the physical asset represented by the digital twin is defined as "Physical State" or S. S changes over time, starting at S_0.

The Physical State space is characterised by its complexity and dynamism, as it encompasses the ever-changing conditions and interactions within the physical environment. This complexity poses challenges for accurately capturing and representing the full extent of the Physical State space in the digital twin. However, advancements in sensor technology, data analytics, and modelling techniques enable increasingly detailed and accurate representations of the physical asset (Onjaji et al., 2020).

Digital State, *D*

The Digital State (*D*) space refers to the virtual representation of the physical twin within the digital twin (Malik et al., 2020). It comprises a set of variables and parameters that mirror those in the Physical State space but are represented computationally. Like the Physical State, the Digital State also moves forward in time (*t*) and should be temporally in sync with the Physical State (Figure 2). The variables in the Digital State are typically captured through sensors, measurements, models or simulations and are used to simulate the behaviour of the physical system in the digital domain.

In contrast to the Physical State, the Digital State space offers advantages such as controllability, reproducibility, and scalability (Wu and Li, 2021). Since it exists within the computational domain, the Digital State space can be manipulated, analysed, and experimented with in ways that may be impractical or impossible in the physical realm. This flexibility allows researchers to explore different scenarios, optimise system performance, and predict future states of the physical twin (Wu and Li, 2022).

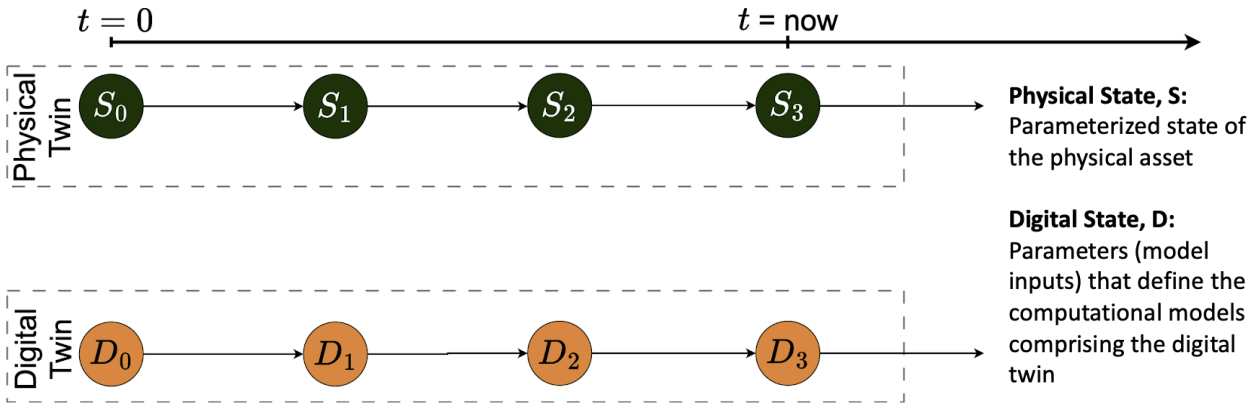

Figure 2: The Physical and Digital States represented in parallel, where the Digital State is a model or simulation that mimics the Physical State with a defined set of state variables, or "State Space".

Observational Data, O

Observational Data (O) forms the link between *S* and *D* States (Figure 3). The linkage between physical and digital states through observational data is fundamental to the operation of digital twins, serving as the conduit through which real-world phenomena are mirrored and simulated in the digital realm (Malik et al., 2020). It is also important to note that O only captures a certain representation of *S* at a certain frequency and fidelity, therefore O should be seen as a simplified abstraction of *S*. Observational data, gathered from a variety of sources including field surveys, remote sensing, and monitoring stations, provide insights into the current state of the environment, encompassing factors such as biodiversity, habitat conditions, and ecosystem processes. These data serve as the backbone for digital twins, acting as the primary means by which the digital representation of the ecosystem is updated and refined.

Moreover, observational data play a crucial role in validating the accuracy of ecological models embedded within the digital twin, ensuring that they capture the complexity and variability of real-world ecosystems. Through sophisticated data integration and analysis techniques, the digital twin synchronises its virtual state space with the ever-changing dynamics of the physical ecosystem. This synchronisation enables researchers to model and simulate ecological processes, predict future trends, and assess the potential impacts of human activities or environmental disturbances.

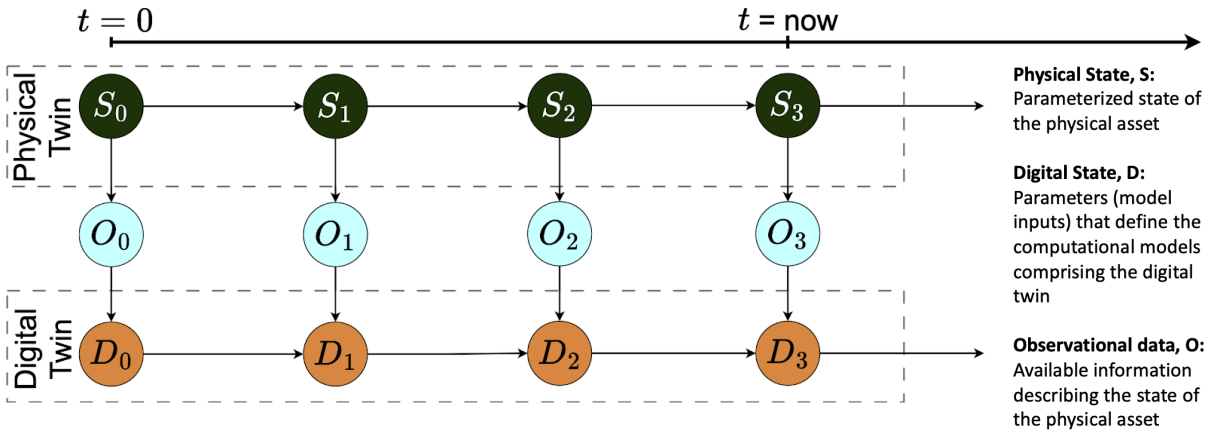

Figure 3: The Physical (S) and Digital (D) States are synchronised by the Observational Data (O), however the direction from S to D is not always linear as a temporal lag in O can shift the synchronisation. As S changes from S_0 to S_n, O_0 and D_0 also change to O_n and D_n.

Adjusting Temporal Delay

There is a delay between when any changes happen in the Physical State, when any observation is made of the changes, and when the observation is added to the Digital State. To compensate for this delay, we introduce the concept of Temporal Delay.

Temporal Delay refers to the time difference between the timestamps of the Digital State (D_n) and Physical State (S_n). This can be represented as:

$$Temporal\ Delay\ (T)\ =\ Time\ Physical\ State\ (Ts)\ -\ Time\ Digital\ State\ (Td)$$

It can be seen that T_d is completely dependent on when an observation is made (T_o) and when the Physical State actually changes (S_n) (Figure 4). For example, actual fish population size over time would be the physical state time series, incomplete and imprecise counts of fish sampled in a survey, or caught in a fishery, would be the observation time series, and the modelled fish count would be the Digital State time series. Such hierarchical time series structures have been previously described in State-Space and Non-Linear Dynamics modelling approaches, where the Temporal Delay has been referred to as the “hidden state” (Auger-Méthé et al., 2021).

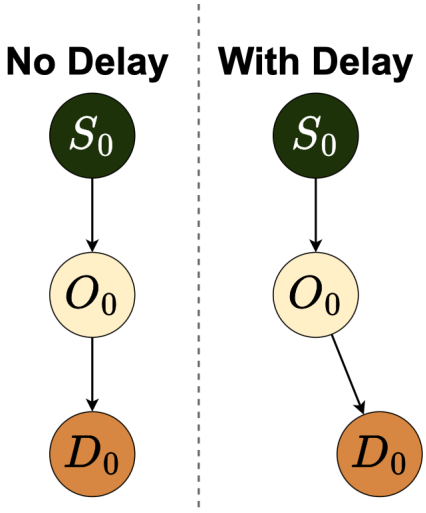

Figure 4: A visual representation of the Temporal Delay between when Physical Twin updates and when the Digital Twin updates. In many ecological cases, the Observational data representing a system is updated rather slowly, and usually considerable time has passed by the time models consume the data. Thus, the models represent a historic version of the Physical State.

Many aspects of ecology show stability over considerable spans of time. Hence, the temporal lag between Digital and Physical States may not matter in numerous cases. In contrast to engineering DTs, another differentiating feature of ecological DTs (and ecological data in particular) is that there are still many manual processes involved, both on providing the Observation Data as well as on the Control Inputs (managing the ecosystems based on insights from the DTs). Species range shift, time of flowering and other phenological aspects may not change overnight for example. Therefore, what matters perhaps is the acceptable length of temporal delay. The Temporal Delay is inherent in all digital twins, though the degree of the lag varies. In most cases, an observation (*O*) is made of something (*S*) and its state is simulated (*D*). *O* and *D* cannot happen at the same time. The lag is actually between the observed (*O*) and predicted (*D*) states, not between the actual (*S*) and predicated (*D*) states.

Control Inputs, *U*

Control Inputs (*U*) in the context of digital twins represent actions or interventions that can be applied to the physical system based on insights or predictions generated from its digital counterpart (Malik et al., 2020) (Figure 5). These inputs serve as a mechanism to influence and modify the Physical State of the asset in response to changes detected or anticipated in the Digital State. Control Inputs are a crucial component of digital twins, enabling stakeholders to actively manage and optimise the behaviour of complex systems across various domains (Buonocore et al., 2022).

In ecology, Control Inputs can either be of direct *interventional* type or of *policy* type. Such Control Inputs can encompass a wide range of actions that can be applied to the physical asset, including adjustments to operational policy parameters, deployment and management of resources, and implementation of biodiversity control strategies. These actions are typically informed by analyses

conducted within the digital twin, which may include optimization algorithms, predictive models, or decision-support systems. By leveraging the insights derived from the digital twin, stakeholders can identify optimal control strategies that maximise performance, efficiency, or other desired objectives while minimising risks or undesirable outcomes (Malik et al., 2020).

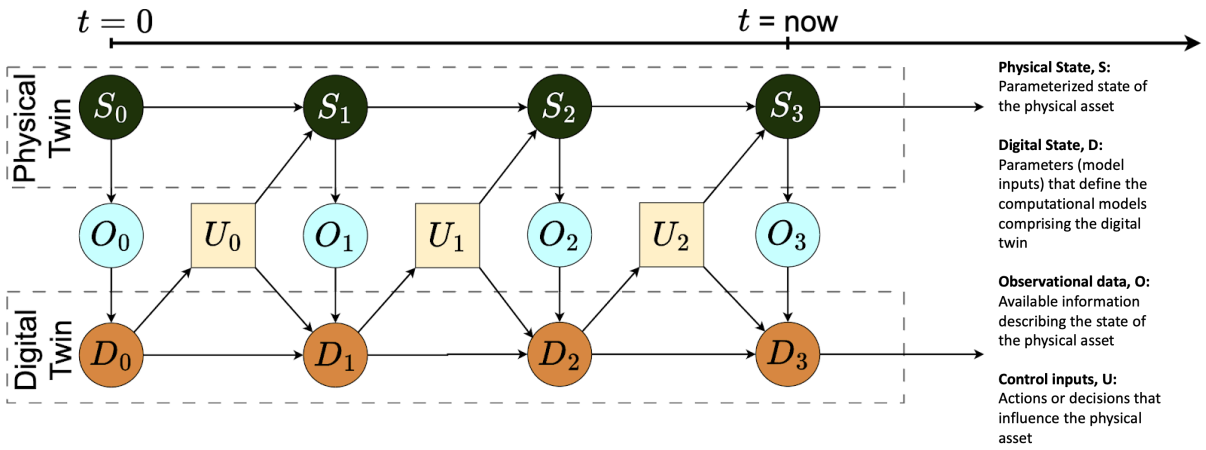

Figure 5: As the Digital State (D) updates at each time-step (t) based on new observational data, the computational modelled/simulated outputs can be used to drive a set of Control Input (U) at each step that changes the Physical State (S) of the asset.

Here are examples of Interventional and Policy Control Input types:

1. Interventional Control Inputs:

In the management of a freshwater ecosystem such as a lake or reservoir using digital twins, ecological Control Inputs could involve the manipulation of water flow rates (Qiu et al., 2023). By adjusting the flow rates of water into or out of the ecosystem, resource managers can influence factors such as water temperature, nutrient levels, and habitat availability. For instance, during periods of high nutrient runoff from surrounding agricultural areas, managers may increase the outflow rate to prevent eutrophication and maintain water quality. Conversely, during dry periods, managers may decrease the outflow rate to conserve water and maintain habitat integrity for aquatic species. These Control Inputs are informed by insights from the digital twin, which simulates the ecological dynamics of the system and predicts the impacts of different flow management strategies on ecosystem health and resilience (Qiu et al., 2023).

2. Policy Control Inputs:

A policy control input in ecology comprises suggested actions by the digital twin that help in implementing regulations or management strategies aimed at conserving endangered species or protecting critical habitats (Sharef et al., 2022). Such suggested actions from the digital twin, which may include predictive models of species distribution and population dynamics, can inform the development of policies that mitigate threats to biodiversity and ecosystem health (Scheibmeir and Malaiya, 2022). For instance, the digital twin could suggest establishing protected areas, habitat corridors, or zoning regulations based on projections of habitat loss due to urban development or

climate change, helping policymakers safeguard key habitats and mitigate fragmentation. These policy interventions aim to address broader societal and environmental goals by integrating ecological knowledge with regulatory frameworks and stakeholder engagement, ultimately shaping land-use decisions and promoting sustainable development practices. In such cases, the policy could even be dynamic, and adjust with the changing digital twin outputs. It is worth mentioning that Policy Control Inputs add more time to the Temporal Delay, as implementations and subsequent changes take time to be implemented and show results.

Evaluation, $\widehat{O}$:

Evaluation ($\widehat{O}$) of the Digital State based on Observational Data is a critical step in ensuring alignment between the virtual representation and the real-world conditions (Malik et al., 2020)(Figure 6). Discrepancies between the two can reveal areas where the digital twin may be inaccurately representing the physical twin or where Control Inputs may need adjustment. For example, if the digital twin predicts a decrease in species occurrence following the implementation of a control strategy but observational data show no corresponding reduction, it may indicate a need to refine the model or reassess the effectiveness of the control inputs.

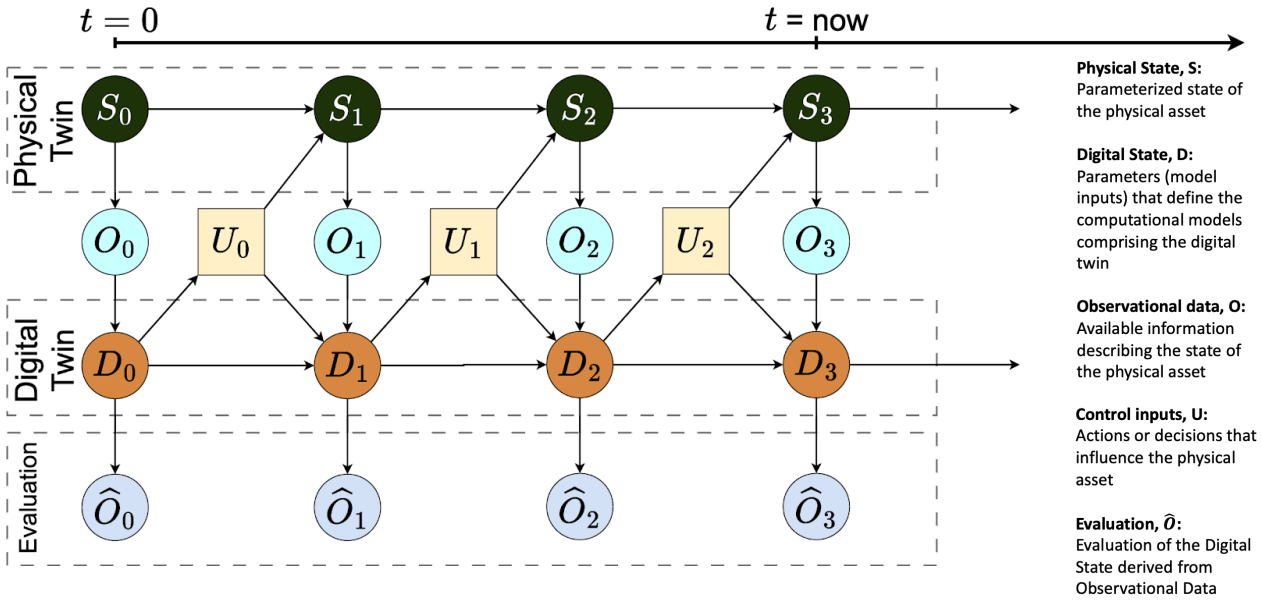

Figure 6: After the output of Digital State is updated after a model/simulation run, it is imperative to validate the results when compared to the Observational Data (O) in order to keep the Digital State (D) aligned with Physical State (S) at a desired fidelity and frequency.

Moreover, the evaluation of the Digital State serves as a feedback mechanism for validating the accuracy and reliability of the digital twin. By comparing simulated outputs to observed data, stakeholders can assess the fidelity of the digital twin in capturing the complex interactions and dynamics of the physical system. This validation process is essential for building trust and confidence in the digital twin as a decision-support tool for managing real-world systems (Trantas et al., 2023).

Rewards, *R*:

Rewards (R) serve as evaluative metrics or objectives that guide the optimization of system behaviour towards an ideal state (Malik et al., 2020) (Figure 7). Rewards are typically defined based on the goals, objectives, or performance criteria established for the system being modelled. By quantifying the desirability or effectiveness of different Digital States or trajectories, rewards provide a feedback mechanism that informs the selection and application of Control Inputs to steer the system towards desired outcomes. Rewards can also be used to adjust the Temporal Delay if the Digital State is not at a desired fidelity or frequency.

Rewards are often used in conjunction with reinforcement learning algorithms, a type of machine learning approach that enables autonomous decision-making and control based on feedback from the environment where Q-learning is an algorithm that finds an optimal action-selection policy for any finite Markov decision process (MDP) (Dewey, 2014). In this context, rewards act as signals that reinforce or discourage specific actions taken by the digital twin in response to changes in the Physical or Digital States (Zekri et al., 2022). By associating positive rewards with actions that lead to desirable outcomes and negative rewards with actions that result in undesirable outcomes, digital twin algorithms can iteratively learn and optimise control strategies to achieve desired objectives.

For example, in the management of a water distribution network, rewards could be defined based on criteria such as system efficiency, water quality, and customer satisfaction (Zekri et al., 2022). By assigning higher rewards to Control Inputs that reduce water losses, improve water quality, and meet demand while minimising energy consumption, the digital twin can learn to optimise pump schedules, valve settings, and flow rates to achieve these objectives. Over time, the digital twin algorithm adjusts control strategies based on feedback from the environment, gradually steering the system towards an ideal state characterised by efficient water delivery and minimal waste.

In ecological applications, Rewards could be defined based on conservation goals such as biodiversity conservation, habitat restoration, or ecosystem resilience (Buonocore et al., 2022). For instance, in the management of a protected area, Rewards could be assigned to Control Inputs that enhance habitat quality, support endangered species, and mitigate threats such as invasive species or habitat degradation. By incentivizing actions that promote ecological health and resilience, Rewards provide a framework for prioritising management interventions and guiding decision-making in complex and dynamic ecosystems.

Overall, Rewards in digital twins serve as a mechanism for aligning system behaviour with predefined objectives, enabling autonomous decision-making and ability to steer the physical asset towards desired states. By quantifying the desirability of different outcomes and providing feedback to learning algorithms, Rewards empower digital twins to learn and adapt to changing conditions, ultimately facilitating the optimization of system performance and the achievement of strategic goals (Malik et al., 2020).

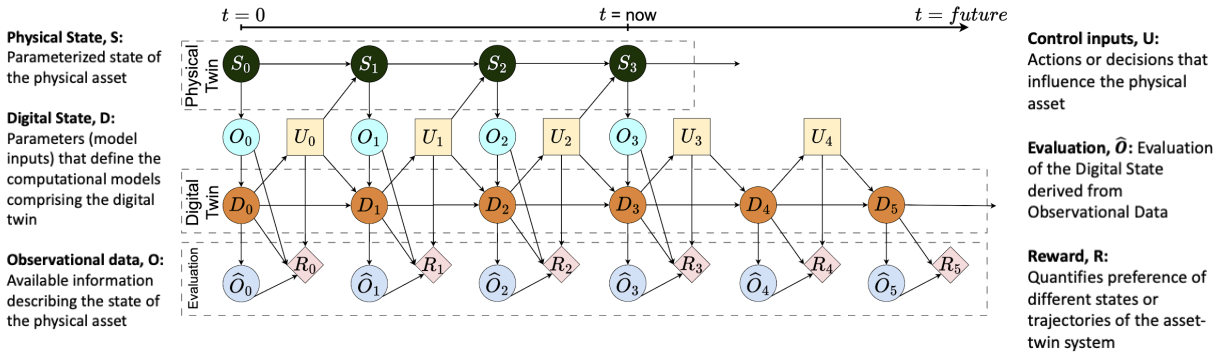

Figure 7: The complete overview of the layers of digital twins in the TwinEco framework. Rewards (R) are linked to the Digital State (D), Observational Data (O), Evaluation ($\widehat{O}$) and Control Inputs (U).

2. TwinEco Components

After defining the layers of digital twins, the next step is to define the interaction between these layers. For this, certain components are suggested here. The components define the mechanics of the dynamics between the layers, meaning they realise the movement from *S* to O to *D* to *U* and so on. Inside a digital twin software environment, the components can be implemented in various software forms, allowing developers to tailor the system to their specific use cases and preferences. These components can be designed using Object-Oriented Programming (OOP) principles, where each component is represented as a class with its own attributes and methods, facilitating modularity and reuse. Alternatively, developers might employ functional programming techniques, encapsulating each component's functionality within discrete functions that can be composed and reused across the system. In other scenarios, simple scripts might suffice, particularly for straightforward tasks or data processing pipelines. This flexibility of TwinEco in implementation ensures that digital twin environments can be customised to meet the diverse requirements of different ecological applications, whether they involve complex simulations, real-time data integration, or iterative model adjustments. By accommodating various programming paradigms and structures, digital twin systems can leverage the strengths of each approach, promoting efficient development, ease of maintenance, and scalability.

The Dynamic Data-Driven Application Systems (DDDAS) paradigm prioritises the integration of real-time data streams within computational models, facilitating adaptive decision-making and system control (Darema, 2004). Within the realm of digital twins, DDDAS further extends this principle to develop dynamic, data-driven simulations that interact in real-time with evolving conditions within the physical system. By integrating feedback loops and state management mechanisms, this approach elevates the accuracy and adaptability of digital twins (Malik et al., 2020). Consequently, within the TwinEco framework, individual components are constructed in the digital twin application to fulfil the functions of feedback loops and state management, effectively extending the capabilities of DDDAS (Table 1).

<u>Feedback Loop:</u>

In a DDDAS-enabled digital twin system, feedback loops play a crucial role in closing the loop between the physical and digital realms. These feedback loops enable real-time interactions between the digital twin and the physical system, facilitating the exchange of information, analysis, and control actions (Malik et al., 2020).

State Management:

State management is another critical aspect of DDDAS-enabled digital twins, ensuring that the digital representation accurately reflects the current state of the physical system. This involves maintaining a dynamic and up-to-date model of the system's state, incorporating real-time data streams and adjusting model parameters as conditions evolve (Malik et al., 2020). State management mechanisms facilitate the integration and fusion of heterogeneous data sources, harmonising disparate data streams to create a cohesive and consistent view of the system's behaviour.

Component	Type	Purpose
State Space	State Management	The definition of input and output parameters of the Digital Twin.
vSensor	Feedback Loop	Detect changes in data sources and ensure new data availability before triggering other DT components, preventing data duplication.
Intaker	Feedback Loop	Pull new data into the system.
Processor	Feedback Loop	Clean and validate source data to fit desired rules and structures, ensuring data quality and consistency for the DT system.
Assimilator	Feedback Loop	Assimilate the new data into the existing data and model, ensuring that the DT remains up-to-date with the latest information from the physical system.
Simulator	State Management	Run ecological models and simulate system dynamics based on data inputs and predefined parameters.
Actuator	Feedback Loop	Update the physical state based on the digital state, including ecological interventions and policy recommendations (Control Inputs).
Logger	Feedback Loop	Log events inside the DT to check whether all components function.
Indexer	State Management	Record data, modelling, and log files associated with each timestep, enabling stakeholders to track changes over time and examine different states of the DT.
Diff-Checker	State Management	Track changes in observational data and the digital

		state for each DT at each time-step, providing metadata for version control and system monitoring.
Evaluator	State Management	Validate models against empirical data to ensure accuracy and reliability, comparing simulated outputs with observed real-world conditions.

Table 1: This table encapsulates the essential components of the TwinEco framework, outlining their roles and interactions in creating a robust and flexible digital twin system for ecological applications. The type of components have been colour coded as red for State Management and green for Feedback loop components.

State Space

Ecological systems exhibit complexity across broad temporal and spatial scales. Despite this complexity, ecological modelling approaches seek to distil the intricate dynamics of these systems into a set of manageable input and output variables. These variables collectively define the state of the system they represent and are collectively referred to as the State Space.

The State Space constitutes a fixed set of parameters consumed and outputted by the digital twin application. These parameters, once defined, evolve through time, shaping the dynamics of the digital twin. Moreover, the State Space establishes guidelines for the characteristics of the fixed parameters within the software, such as their data types (e.g., integer, float, string, date, etc.).

Ecological data is heterogeneous, stemming from diverse methodologies and standards employed for data collection in the field, alongside a lack of standardisation practices across studies (Niu et al., 2014). Consequently, a predefined State Space serves a crucial role in enabling the DT to manage this heterogeneity in its state data effectively. By imposing a structure, the State Space ensures that data adheres to desired formats, thereby facilitating the integration and harmonisation of disparate data sources to yield desired outcomes within the DT. In essence, the State Space functions akin to a schema or data model for the DT, guiding the organisation and interpretation of ecological data for effective modelling and analysis.

For instance, consider a digital twin modelling the dynamics of a forest ecosystem. One of the state variables within the State Space could represent the population density of a species like the red squirrel within a specific area of the forest. This state variable would capture the current population size of red squirrels in that area, providing valuable information about the health and dynamics of the ecosystem. As the digital twin evolves, this state variable would change dynamically in response to factors such as predation, habitat loss, or resource availability, reflecting the complex interactions within the ecosystem.

vSensor

Virtual Sensors (vSensors) serve as the initial point of interaction within the digital twin, detecting changes and updates in data sources across virtual spaces such as Application Program

Interfaces (APIs), data files, and databases. They operate irrespective of whether these sources are directly connected to the digital twin, or obtained from third-party providers like research infrastructures. This ensures that other DT components are only triggered when new data is available for further processing, minimising data redundancy within the DT.

Change detection involves the DT actively monitoring the current state of data sources and comparing them for any alterations. For instance, a researcher utilising APIs of a research infrastructure to access species presence data can regularly query these APIs to ascertain the availability of new data on the servers. Virtual Sensors are tailored to each data source, accommodating unique formats, structures, and access protocols. Moreover, many data sources lack standardised metadata protocols, further complicating data integration processes.

In some cases, the digital twin first has to retrieve data by Intakers and subsequently check for changes by Virtual Sensors, necessitating periodic downloads and assessments to ensure that the data is up-to-date and accurate. Therefore, the order of which component comes first (vSensor or Intaker) depends on the source of the data and the available tools.

Intaker

Intakers are responsible for extracting raw observational data from diverse data sources. These sources encompass a broad spectrum of possibilities, including APIs, databases, file storage systems, repositories, and more. In addition to retrieving data, Intakers play a crucial role in validating the integrity and format of the data to ensure it aligns with the expectations of the DT.

For ecological applications, Intakers could be employed to download sensor data from various environmental monitoring stations scattered across a region. For instance, Inktakers might retrieve data from weather stations measuring temperature, humidity, and precipitation, or from soil moisture sensors deployed in agricultural fields. In another scenario, Intakers could pull data from satellite imagery repositories to track changes in land cover and vegetation density over time. These examples demonstrate the versatility of Intakers in sourcing data from disparate sources to fuel ecological modelling and analysis within the DT framework.

In many ecological use cases, Intakers must communicate with diverse research infrastructure APIs and file systems, complicating the data intake process. They handle various data formats and protocols, manage authentication and authorization, and integrate and harmonise data to fit the Digital Twin's predefined state space. Additionally, they must ensure real-time data collection, implement robust error handling, and manage metadata for data integrity and traceability. For example, Intakers might download sensor data from cloud storage or interact with APIs, requiring them to navigate authentication, parsing, and data transformation challenges. These complexities underscore the need for sophisticated Intaker mechanisms to ensure reliable and accurate data ingestion into the Digital Twin system.

Processor

Processors play a critical role in the refinement and validation of source data within a Digital Twin system, ensuring adherence to desired rules and structures. This encompasses a spectrum of tasks, from addressing geospatial requirements such as transforming projections and Coordinate Reference Systems (CRSs), to harmonising taxonomic classifications. Furthermore, Processors are responsible for organising the data into specified structures, such as matrices, lists, or dataframes.

Processors also enforce data quality standards and ensure compliance with established data protocols within the DT. Given the automated nature of DT software, this step is important in preventing disruptions during data processing and analysis.

Data cleaning in ecological contexts presents a multifaceted challenge. Ecological datasets are often characterised by complexity and variability, requiring extensive filtering, cleaning, transforming, and validation procedures. Indeed, ecological models commonly find that the time invested in data cleaning far exceeds that allocated to subsequent analysis and modelling tasks. Automating this process poses significant challenges. Consequently, Processors emerge as a complex yet essential component of ecological DTs, serving as a linchpin for ensuring the integrity and reliability of data-driven analyses within ecological systems.

Assimilator

Assimilators assimilate the processed data into the existing data structure within the DT from previous DT runs. At this stage, the assimilator also handles tasks like versioning of the Observational Data, creating relations within the datasets (assigning unique identifiers), and storing the newly generated datasets into the storage system of the DT in the needed data format. The assimilators can also handle tasks related to cleaning the storage system by deleting the previous datasets or appending the new data to the previous dataset.

Simulator

The Simulator Component is the foundational element of a digital twin system, responsible for replicating the behaviour and dynamics of the physical system within the digital domain. It serves as the computational engine that drives the simulation of the physical system's behaviour, enabling stakeholders to explore scenarios, predict outcomes, and optimise performance in a virtual environment. Simulators take the processed input datasets described in the State Space, and output the Digital State of the DT. A single DT could have a single or multiple Simulators.

At its core, the Simulator Component comprises mathematical or statistical models, algorithms, and computational techniques that capture the essential characteristics and interactions of the physical system. These models may range from simple mathematical approximations or statistical models to complex, process-based simulations, depending on the complexity and fidelity required for the specific ecological use case.

The Simulators take input data from various sources, including Observational Data, Control Inputs, and variables from the State Space, and use this information to simulate the behaviour of the physical system over time. By iteratively updating the state of the simulation based on these inputs, the Simulator Component generates a digital representation of the physical system's behaviour, which can be visualised, analysed, and manipulated by stakeholders.

Actuator

In ecological modelling, direct intervention of models in the habitat or ecological process is not desired. In many cases, in fact, the interventions are in the shape of policy recommendations and updates. However, within the context of a Digital Twin system, predefined Control Inputs can guide the system towards an optimal state, driven by desired Rewards. Actuators serve as the components responsible for implementing these Control Inputs. With this in mind, Actuators can have two different types:

1. Interventional Actuators

Interventional Actuators intervene directly in the Physical State to cause change. For instance, during periods of high nutrient runoff from surrounding agricultural areas, the Interventional Actuators may trigger an increase of the outflow rate to prevent eutrophication and maintain water quality.

2. Policy Actuators

Policy Actuators, on the other hand, are action recommendations that steer the implementation of regulations or management strategies aimed at addressing ecological issues. These actuators facilitate the execution of policies designed to conserve endangered species or protect critical habitats. For instance, a predictive digital twin can inform Policy Actuators by providing insights into species distribution and population dynamics, thus guiding the development of effective policy parameters that safeguard biodiversity and ecosystem health.

A policy control input in ecology could involve implementing regulations or management strategies aimed at conserving endangered species or protecting critical habitats (Sharef et al., 2022). Insights from the digital twin, which may include predictive models of species distribution and population dynamics, can inform the development of policies that mitigate threats to biodiversity and ecosystem health (Scheibmeir and Malaiya, 2022).

Logger

Loggers serve as a vital tool for monitoring and recording events occurring within the digital environment. Its primary function is to track the behaviour and interactions of various components within the DT to ensure that they are functioning as intended. By logging events and activities, the Loggers provide valuable insights into the performance, reliability, and overall functioning of the DT system.

The Logger Component operates by capturing and storing information about key events, such as the execution of Control Inputs, updates to the Digital State, and responses from Actuators. This information is typically recorded in a structured format, including timestamps, event descriptions, and relevant contextual metadata. By maintaining a comprehensive log of events, the Logger Component enables stakeholders to trace the sequence of code executions and identify any logical errors in the algorithms or bugs in the software.

Loggers play a crucial role in system diagnostics and troubleshooting. In the event of errors or malfunctions within the DT, the log data can be analysed to pinpoint the root cause of the issue and facilitate corrective actions. By correlating events and identifying patterns in the log data, engineers and operators can gain valuable insights into the underlying dynamics of the DT system and implement improvements to enhance its performance and reliability.

Indexer

Indexers are a temporal indexing mechanism implemented within the TwinEco Digital Twin framework that serve as a crucial tool for maintaining a comprehensive record of data, Digital State, and Physical State evolution over time. By tracking timestamps and associating them with relevant data and files, this metadata infrastructure facilitates version control and enables stakeholders to navigate through different temporal states of the DT.

For instance, as the DT progresses from its initial state at t=0 to subsequent time-steps, the indexers diligently record all associated data, modelling outputs, and log files. This temporal labelling allows users to seamlessly traverse through various time-steps of the DT, whether it's for retrospective analysis, model evaluation, or performance assessment.

The temporal indexing system finds applications in meta-modelling endeavours and evaluating model performance. By indexing single updates of the Digital State and correlating them with the timestamps of corresponding Observational Data recordings in the physical system, stakeholders gain insights into the temporal alignment between simulated and observed states. Three distinct types of Indexers fulfil this function: Observational Data indexers, Digital State indexers, and Physical State indexers. These Indexers provide answers to critical questions such as when the data was recorded, when the prediction was produced or the DT was updated, and what specific time and date the DT predictions represent. This comprehensive temporal indexing infrastructure enhances the utility and transparency of DT systems, empowering stakeholders to make informed decisions and derive valuable insights from temporal data dynamics.

Diff-checker

Diff-checkers serve a pivotal function within the TwinEco framework by monitoring alterations in both Observational Data and the Digital State during each DT update. These mechanisms, essentially metadata, provide an invaluable record of what changes occur over time.

The utility of Diff-checkers extends across various domains, including metamodelling, debugging, model evaluation, and decision-making processes. Diff-checkers offer comprehensive insights into system dynamics, enabling stakeholders to track changes not only from a data perspective but also from a model viewpoint, thereby furnishing a holistic understanding of system behaviour. In essence, Diff-checkers play a vital role in monitoring system dynamics, facilitating informed decision-making and enhancing the value proposition of DTs over traditional approaches.

For instance, consider a digital twin designed to model the distribution of a particular species within a forest ecosystem. The Diff-checker would continuously compare updates in Observational Data, such as species occurrence records from field surveys or remote sensing data, with the predictions generated by the digital twin, capturing what exactly changes from one time-step to the next. If the digital twin predicts an increase in the population density of the species within a certain area, but the Observational Data indicate a decline or no change, the Diff-checker could flag this discrepancy. This could prompt further investigation into potential factors influencing the discrepancy, such as habitat degradation, climate change effects, or inaccuracies in the model assumptions. Overall, the Diff-checker serves as a valuable tool for ensuring the accuracy and reliability of the State Space and the Simulators within digital twin systems.

Evaluator

The role of Evaluators within the development and validation of Digital Twin (DT) systems is paramount. The Evaluator Component assumes a pivotal role in the pre-deployment phase, serving to validate the accuracy and reliability of the models against empirical data. This validation process entails rigorous comparison of simulated outputs with observed behaviour in real-world conditions. Through iterative refinement, the models are honed until they faithfully replicate the behaviour of the physical system at the desired frequency and fidelity.

The nature of evaluation undertaken is contingent upon the type of Simulator utilised within the DT framework.

Mapping components to layers

As shown previously, the TwinEco framework is organised into distinct layers, each comprising specific components that collectively enable comprehensive ecological modelling. The components introduced above can be mapped to the layers of the TwinEco framework to create the interactions between layers (Figure 8).

As the digital twin transitions from the Physical State to the Digital State at each time-step, it incorporates components such as Virtual Sensors and Intakers, which are responsible for gathering and integrating raw data from various sources. This layer ensures that the digital twin has access to up-to-date and relevant Observational Data.

In the Observational Data layer, components like Processors and Diff-Checkers play a crucial role in cleaning, validating, and tracking changes in the data. This layer transforms raw data into

structured formats suitable for modelling and analysis. The Digital State layer includes the Simulator, which runs the ecological models, and Actuators, which implement control inputs based on model predictions. This layer is vital for generating simulations that reflect real-world ecological dynamics.

The Evaluation layer comprises Evaluators and Indexers. Evaluators validate the accuracy and reliability of the models by comparing simulated outputs with empirical data, while Indexers manage metadata and ensure version control, allowing for the tracking of changes over time. Together, these layers and components create a robust and adaptable framework for developing digital twins in ecology. Figure 8 illustrates the mapping of these layers to their respective components as single time-steps, providing a visual representation of the TwinEco framework.

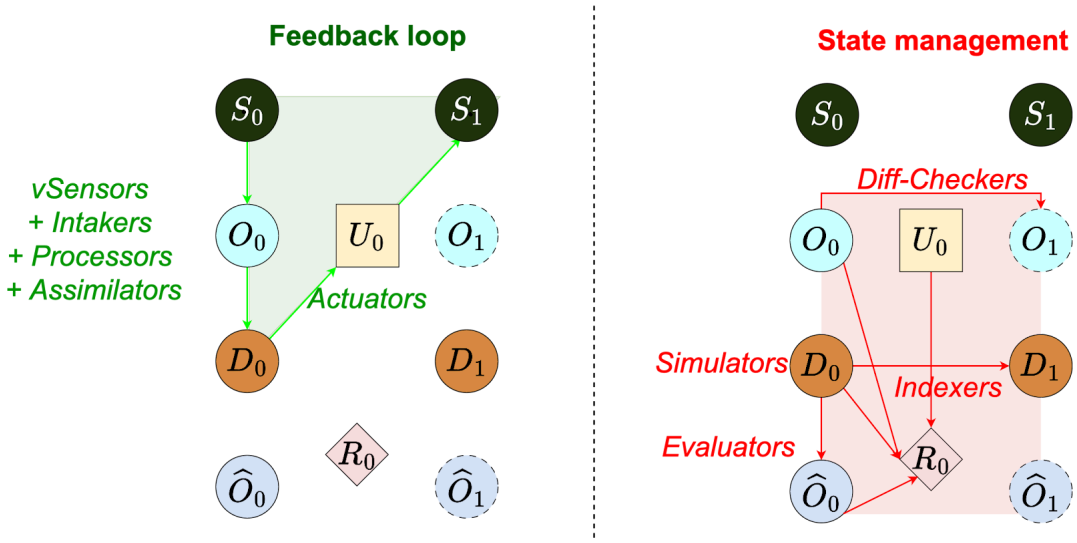

Figure 8: Mapping of the components to the layers of the TwinEco framework, where feedback loop (green) and state management (red) components are colour coded.

4. Discussion

Digital twins herald a significant shift in ecological modelling, prompting rapid adoption for scientific research. With burgeoning interest among modellers, however, there arises a heightened risk of fragmentation in defining and structuring Digital Twin applications. This fragmentation could lead to divergent interpretations across communities, resulting in relevance confined within specific domains but diminishing outside reach. This challenge becomes particularly pronounced with initiatives like Destination Earth, aiming to construct comprehensive digital twins of Earth and its myriad systems (Nativi et al., 2021). Consequently, the imperative for a unified digital twin framework, such as TwinEco, becomes more pressing than ever, ensuring coherence amidst the proliferation of diverse use cases.

Recommendation 1: Researchers building new digital twins should be mindful of this risk of fragmentation of practices and aim to follow shared principles including frameworks such as TwinEco.

A digital twin can serve numerous purposes in ecology and life sciences. For example, satellite image based digital twins can detect forest disturbances, map and monitor crop stresses, such as drought and disease, as well as crop maturity in near real time, providing alerts for necessary actions. Furthermore, digital twins can access molecular data of medically important viruses to identify genetic changes, evaluate the potential for vaccine resistance, and inform measures to prevent the spread to other regions (Baaden, 2022). They can also detect changes in data availability to build and update habitat suitability models for species in real time, and enable real-time detection of invasive species spread.

In general, digital twins in ecology are used to access and integrate data from different sources to calibrate models and produce the necessary knowledge in real time. Thus, in most cases, they are more about informing the state of knowledge regarding certain aspects of the model targets (the physical states) than about the state of the physical twin itself.

Depending on the specific aspect of an ecosystem or the taxa being modelled, digital twins offer flexibility in the components and layers that could be used. For example, we may not need the evaluation layer/component to detect changes in molecular makeup of a virus. One can use DTs to update data on certain aspects of an ecosystem/taxa and inform the gap in sampling to advise where to sample next. In this specific case, for example, the “Diff-checker” component of the DT may not be that important. Additionally, functionality of multiple components could also be combined into certain use cases.

However, adhering to a common framework is crucial for making digital twins interoperable or capable of communicating with one another. This common framework allows ecological digital twins to integrate easily into initiatives like Destination Earth (Le Moigne, 2022; Ossing et al., 2023; Le Moigne 2024). It ensures they can use explanatory data from the same sources of the destination earth data lake and guarantees their sustainable existence and functionality. The components of digital twins could be shared as “common” components by other twins, specially the components that pull and process data from common research infrastructure in ecology.

Recommendation 2: Digital twins in ecology should aim to follow common design patterns to increase interoperability of the twins across domains and for the twins to be able to reuse parts of each other to create newer, more novel digital twins.

Unlike many traditional ecological modelling frameworks, establishment of an operational and useful analysis workflow using the TwinEco framework does not necessitate deployment of all components we have introduced here. This particular aspect of flexibility/modularity inherent to TwinEco presents two distinct opportunities as well as one marked challenge in adopting its principles into ecological research applications. Firstly, not requiring deployment of all structural components to achieve functionality makes TwinEco adaptable to many ecological applications. For example, many, if not most, ecological disciplines focus on the study of systems which are difficult to influence via actuator-driven feedback loops and thus will require study via TwinEco DTs without the actuator component. Secondly, clear communication of implementation of TwinEco components in ecological analyses pipelines can serve as a benchmark for quantification of degree to which a fully integrated digital twin has been established. This may serve as a guiding principle when designing studies as well as augment past research efforts by adding classifications to their workflows. Lastly, however, this modularity also begs the question "which components are required for a workflow to be called a digital twin?". This is a non-trivial topic with potentially far-reaching consequences that ought to be discussed with the wider ecological research community (Ossing et al., 2023). TwinEco lays the foundation for such consensus-finding on what constitutes an ecological digital twin creating room for direct discussion of components, layers, and time-sensitivity of digital twinning applications.

Recommendation 3: Adjust frameworks like TwinEco to specific use cases, as not all components of the framework will be applicable to every digital twin.

Digital twins require dynamic real time input data while common data collection protocols in ecology result in a time lag and batch updated data sources. Recent experiments with real-time data collecting sensors point towards the need for new data streams more fit for use by digital twins. However, automatic realtime data fusion across distributed networks of sensors add new even stricter requirements for machine-actionable and machine-interoperable data formats and semantics. On the other side, solving these operational challenges facilitates the up-scaling of monitoring of ecological systems.

Recommendation 4: Researchers building new monitoring systems should be mindful of enabling true machine interoperability and machine data fusion to support emerging digital twin systems.

Building digital twins in ecology requires stable data streams facilitated by permanent research infrastructures. Current research infrastructures for biodiversity and ecology data in Europe were designed and implemented before the emerging implementation of digital twins in ecology. Without the strong demand for real-time sensor data streams the current research infrastructures have been implemented with a time lag in data delivery (Li et al., 2023).

Recommendation 5: Research infrastructures for biodiversity and ecology data should aim to position themselves to become more fit for use by the emerging digital twins and establish functionality and new data stream services for more real-time data delivery, and to work together with downstream data sources to mobilise such real-time data streams.

Digital twinning is new to ecologists and there is a need to establish a shared terminology and understanding of digital twin components and concepts (Korenhof et al., 2021). This paper introduces terminology to promote a shared community definition of components needed for building digital twins in ecology. The interTwin project[3] is working on establishing a shared vocabulary of terminology for key concepts of digital twinning.

Recommendation 6: Digital twinning engineers in ecology should contribute to and follow the emerging shared terminologies, conceptualisations, and component frameworks (such as the TwinEco proposed here).

Using the TwinEco framework, ecological dynamics can be captured and studied with greater precision compared to traditional ecological modelling approaches. The framework explicitly models knowledge and processes/states over time, allowing for a more dynamic and realistic representation of ecological systems that updates as the ecological systems change. By integrating vast amounts of diverse data from various sources, the framework ensures a comprehensive and up-to-date understanding of ecological changes.

5. Conclusion

The development and implementation of digital twins in ecology mark a transformative advancement in ecological modelling. By integrating the Dynamic Data-Driven Application Systems (DDDAS) paradigm, the TwinEco framework offers a unified approach to constructing DT models that are adaptive, responsive, and capable of real-time data integration. This framework addresses the challenges of fragmentation in ecological DTs by providing an overview of the important layers and components of DTs, thereby enhancing clarity on the concept for ecological modellers and ensuring consistency in the basic architecture of ecological DTs (Boyes and Watson, 2022). Consequently, the TwinEco framework facilitates the creation of robust and reliable ecological models that can inform decisions or actions with unprecedented accuracy and timeliness.

A key strength of the TwinEco framework lies in its flexibility and modularity. This allows researchers to tailor DT components to specific ecological contexts, whether it involves tracking forest disturbances, monitoring crop health, or assessing the spread of invasive species. By enabling

[3] https://github.com/interTwin-eu/dtc-glossary

dynamic data processing and continuous model refinement, TwinEco helps ecological modellers to conceptualise complex ecological dynamics and respond to environmental changes swiftly. Moreover, the framework's emphasis on interoperability ensures that DTs can seamlessly integrate with larger initiatives like Destination Earth, leveraging shared data sources and contributing to a comprehensive understanding of global ecological systems (Rao et al., 2023).

Future research and development in ecological Digital Twins should focus on enhancing real-time data collection and integration capabilities of the Research Infrastructures, particularly through focusing on data streaming technologies and data standardisation. Efforts should be made to improve the interoperability of data sources, ensuring seamless communication and data exchange between different DT systems and across various ecological domains. Additionally, it is essential to demonstrate the framework's utility through additional case studies. Documenting diverse applications will illustrate the TwinEco framework's versatility in modelling various ecological systems and addressing a wide range of ecological challenges. These case studies will serve as practical examples showcasing the framework's strengths and areas for improvement. Collaboration among ecologists, data scientists, and technologists will be crucial in driving these innovations forward. Establishing shared terminologies, best practices, and guidelines will also be essential to foster a cohesive community approach and to maximise the impact and utility of Digital Twins in ecological research and management (Korenhof et al., 2021).

In conclusion, the TwinEco framework represents a significant step forward in understanding DTs in ecological modelling, providing a unified yet flexible approach to developing Digital Twins. It addresses the critical need for data update integration and model adaptability, offering a powerful tool for advancing ecological research and management. As the field of ecological modelling continues to uptake cutting-edge technology, the adoption of frameworks like TwinEco will be essential in ensuring that DTs remain relevant, reliable, and capable of addressing the complex challenges facing our ecosystems today and in the future. By fostering collaboration and standardisation, TwinEco paves the way for more effective and coordinated efforts in ecological conservation and sustainability.

6. Acknowledgements

This study has partially 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). Views and opinions expressed are those of the authors 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. We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LUMI, hosted by CSC[4] (Finland) and the LUMI consortium through a EuroHPC Development Access call.

[4] https://www.csc.fi

7. Author Contributor Roles

Table 2 shows the contribution roles that have been ascertained using CRediT[5] (Contributor Roles Taxonomy).

Author	Roles
Taimur Khan	Conceptualization, Investigation, Methodology, Project administration, Supervision, Conceptualization, Writing – original draft
Koen de Koning	Conceptualization, Writing – review & editing, Methodology, Validation
Dag Endresen	Conceptualization, Writing – review & editing, Validation
Desalegn Chala	Conceptualization, Writing – review & editing, Validation
Erik Kusch	Conceptualization, Writing – original draft, Writing – review & editing, Methodology, Validation

Table 2: The CRediT roles for the authors of this manuscript.

8. Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve readability and grammar for sentences that were too long or awkward, as none of the authors are native English speakers. No aspect of this manuscript used ChatGPT for content, logic or reasoning. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

9. References

Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A.A., Leos-Barajas, V., Mills Flemming, J., Nielsen, A., Petris, G., others, 2021. A guide to state–space modelling of ecological time series. Ecological Monographs 91, e01470. https://doi.org/10.1002/ecm.1470

Baaden, M., 2022. Deep inside molecules-digital twins at the nanoscale. Virtual Reality & Intelligent Hardware 4, 324–341. https://doi.org/10.1016/j.vrih.2022.03.001

[5] https://credit.niso.org/

Boyes, H., Watson, T., 2022. Digital twins: An analysis framework and open issues. Computers in Industry 143, 103763. https://doi.org/10.1016/j.compind.2022.103763

Brown, E.D., Williams, B.K., 2015. Resilience and resource management. Environmental management 56, 1416–1427. https://doi.org/10.1007/s00267-015-0582-1

Buonocore, L., Yates, J., Valentini, R., 2022. A proposal for a forest digital twin framework and its perspectives. Forests 13, 498. https://doi.org/10.3390/f13040498

Carpenter, S.R., Brock, W.A., 2004. Spatial complexity, resilience, and policy diversity: fishing on lake-rich landscapes. Ecology and Society 9. https://doi.org/10.5751/ES-00622-090108

Damgaard, C., 2019. A critique of the space-for-time substitution practice in community ecology. Trends in ecology & evolution 34, 416–421. https://doi.org/10.1016/j.tree.2019.01.013

Darema, F., 2004. Dynamic data driven applications systems: A new paradigm for application simulations and measurements, in: International Conference on Computational Science. Springer, pp. 662–669. https://doi.org/10.1007/978-3-540-24688-6_86

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 & Evolution. https://doi.org/10.1016/j.tree.2023.04.010

Dewey, D., 2014. Reinforcement learning and the reward engineering principle, in: 2014 AAAI Spring Symposium Series.

Donohue, I., Hillebrand, H., Montoya, J.M., Petchey, O.L., Pimm, S.L., Fowler, M.S., Healy, K., Jackson, A.L., Lurgi, M., McClean, D., others, 2016. Navigating the complexity of ecological stability. Ecology letters 19, 1172–1185. https://doi.org/10.1111/ele.12648

Fissore, V., Bovio, L., Perotti, L., Boccardo, P., Borgogno-Mondino, E., 2023. Towards a Digital Twin Prototype of Alpine Glaciers: Proposal for a Possible Theoretical Framework. Remote sensing 15, 2844. https://doi.org/10.3390/rs15112844

Golivets, M.,Sharif, I., Wohner, C., Grimm, V., Schigel, D., 2024. Building Biodiversity Digital Twins. Research Ideas and Outcomes. https://doi.org/10.3897/rio.coll.240

Korenhof, P., Blok, V., Kloppenburg, S., 2021. Steering representations—towards a critical understanding of digital twins. Philosophy & technology 34, 1751–1773. https://doi.org/10.1007/s13347-021-00484-1

Le Moigne, J., 2024. Towards Earth System Digital Twins (ESDT), in: La Journée de l'Innovation Du CNES.

Le Moigne, J., 2022. Earth System Digital Twins (ESDT) Technology for NASA Earth Science, in: Living Planet Symposium.

Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., Shen, H., Xiao, Q., Su, J., Yuan, S., others, 2023. Big Data in Earth system science and progress towards a digital twin. Nature Reviews Earth & Environment 4, 319–332. https://doi.org/10.1038/s43017-023-00409-w

Malik, S., Rouf, R., Mazur, K., Kontsos, A., 2020. A dynamic data driven applications systems (DDDAS)-based digital twin IoT framework, in: Dynamic Data Driven Applications Systems: Third International Conference, DDDAS 2020, Boston, MA, USA, October 2-4, 2020, Proceedings 3. Springer International Publishing, pp. 29–36. https://doi.org/10.1007/978-3-030-61725-7_6

Morlot, M., Rigon, R., Formetta, G., 2024. Hydrological digital twin model of a large anthropized italian alpine catchment: The Adige river basin. Journal of Hydrology 629, 130587. https://doi.org/10.1016/j.jhydrol.2023.130587

Nativi, S., Mazzetti, P., Craglia, M., 2021. Digital ecosystems for developing digital twins of the earth: The destination earth case. Remote Sensing 13, 2119. https://doi.org/10.3390/rs13112119

Newton, A.C., 2016. Biodiversity risks of adopting resilience as a policy goal. Conservation Letters 9, 369–376. https://doi.org/10.1111/conl.12227

Niu, S., Luo, Y., Dietze, M.C., Keenan, T.F., Shi, Z., Li, J., III, F.S.C., 2014. The role of data assimilation in predictive ecology. Ecosphere 5, 1–16. https://doi.org/10.1890/ES13-00273.1

Onaji, I., Tiwari, D., Soulatiantork, P., Song, B., Tiwari, A., 2022. Digital twin in manufacturing: conceptual framework and case studies. International journal of computer integrated manufacturing 35, 831–858. https://doi.org/10.1080/0951192X.2022.2027014

Ossing, F., Attinger, S., Jung, T., Visbeck, M., Brune, S., Cotton, F., Teichmann, C., 2023. Synthesis paper Digital Twins of Planet Earth: First Draft for the General Assembly.

Qiu, Y., Liu, H., Liu, J., Li, D., Liu, C., Liu, W., Wang, J., Jiao, Y., 2023. A Digital Twin Lake Framework for Monitoring and Management of Harmful Algal Blooms. Toxins 15, 665. https://doi.org/10.3390/toxins15110665

Rao, Y., Redmon, R., Dale, K., Haupt, S.E., Hopkinson, A., Bostrom, A., Boukabara, S., Geenen, T., Hall, D.M., Smith, B.D., others, 2023. Developing Digital Twins for Earth Systems: Purpose, Requisites, and Benefits. arXiv preprint arXiv:2306.11175. https://doi.org/10.48550/arXiv.2306.11175

Ruckelshaus, M.H., Jackson, S.T., Mooney, H.A., Jacobs, K.L., Kassam, K.-A.S., Arroyo, M.T., Báldi, A., Bartuska, A.M., Boyd, J., Joppa, L.N., others, 2020. The IPBES global assessment: pathways to action. Trends in Ecology & Evolution 35, 407–414. https://doi.org/10.1016/j.tree.2020.01.009

Scheibmeir, J., Malaiya, Y., 2022. A Social Media-driven Digital Twin of an Invasive Species, in: Proceedings of the 10th International Workshop on Simulation for Energy, Sustainable Development & Environment, SESDE, Rome, Italy. pp. 19–21. https://doi.org/10.46354/i3m.2022.sesde.002

Segovia, M., Garcia-Alfaro, J., 2022. Design, modeling and implementation of digital twins. Sensors 22, 5396. https://doi.org/10.3390/s22145396

Sharef, N.M., Nasharuddin, N.A., Mohamed, R., Zamani, N.W., Osman, M.H., Yaakob, R., 2022. Applications of data analytics and machine learning for digital twin-based precision biodiversity: a review, in: 2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS). IEEE, pp. 1–7. https://doi.org/10.1109/ICACNIS57039.2022.10055149

Sougioultzogloua, F., Cook, E., 2023. Digital Twins, Planes, and Drones: Bridging the Gap in Arctic Polar Altimetry Data.

Trantas, A., Plug, R., Pileggi, P., Lazovik, E., 2023. Digital twin challenges in biodiversity modelling. Ecological Informatics 102357. https://doi.org/10.1016/j.ecoinf.2023.102357

Vermeiren, P., Reichert, P., Schuwirth, N., 2020. Integrating uncertain prior knowledge regarding ecological preferences into multi-species distribution models: Effects of model complexity on predictive performance. Ecological modelling 420, 108956. https://doi.org/10.1016/j.ecolmodel.2020.108956

Wu, Z., Li, J., 2021. A framework of dynamic data driven digital twin for complex engineering products: the example of aircraft engine health management. Procedia manufacturing 55, 139–146. https://doi.org/10.1016/j.promfg.2021.10.020

Zekri, S., Jabeur, N., Gharrad, H., 2022. Smart water management using intelligent digital twins. Computing and Informatics 41, 135–153. https://doi.org/10.31577/cai_2022_1_135

Zurell, D., König, C., Malchow, A.-K., Kapitza, S., Bocedi, G., Travis, J., Fandos, G., 2022. Spatially explicit models for decision-making in animal conservation and restoration. Ecography 2022. https://doi.org/10.1111/ecog.05787