A digital twin approach for the identification and update of ecological infrastructure

Lambiel, Audrey; Giuliani, Gregory; Kuelling, Nathan; Lehmann, Anthony

How to cite

LAMBIEL, Audrey et al. A digital twin approach for the identification and update of ecological infrastructure. In: Big earth data, 2026, p. 1–19. doi: 10.1080/20964471.2026.2615504

This publication URL: https://archive-ouverte.unige.ch/unige:190880

Publication DOI: 10.1080/20964471.2026.2615504

BIG EARTH DATA
https://doi.org/10.1080/20964471.2026.2615504

TECHNICAL NOTE

OPEN ACCESS Check for updates

A digital twin approach for the identification and update of ecological infrastructure

Audrey Lambiel [a], Gregory Giuliani [a,b], Nathan Külling [a] and Anthony Lehmann [b,c]

[a]Institute for Environmental Sciences, enviroSPACE Lab, University of Geneva, Geneva, Switzerland; [b]Institute for Environmental Sciences, GRID-Geneva, University of Geneva, Geneva, Switzerland; [c]Department of F.-A. Forel of Environment and Aquatic Sciences, University of Geneva, Geneva, Switzerland

ABSTRACT

Addressing the global environmental crisis necessitates coordinated efforts, supported by open and reproducible research practices. Such practices aim to enhance the reliability, efficiency, and credibility of scientific outputs. Innovative tools are necessary for systematic conservation planning. This technical note presents a reproducible and automated approach for supporting land management and planning by identifying and updating ecological infrastructure (EI). Grounded in open science Findable, Accessible, Interoperable, and Reusable principles data management, and digital twin (DT) concepts, the method focuses on the Canton of Geneva, Switzerland. It integrates species distribution modelling, ecosystem service assessments, and spatial prioritization within a shared JupyterLab environment. The infrastructure centralizes data, automates indicator calculations, and ensures transparency, traceability, and reproducibility through version control and metadata generation. Ecological tools like Zonation enable the identification of high-priority conservation areas aligned with international targets. The system facilitates collaborative workflows and indicator updates. Its architecture allows scalability to broader regions and scenario modelling, laying the foundation for a DT of Geneva's environment. Despite challenges in harmonizing workflows across institutional partners, this solution enhances efficiency and replicability in EI planning. The methodology is transferable to other regions and adaptable to various environmental modelling domains, offering a robust base for sustainable territorial management.

ARTICLE HISTORY
Received 12 August 2025
Accepted 9 December 2025

Keywords
Ecological infrastructure; species distribution modeling; ecosystem services; systematic conservation planning; prioritization; digital twin

1. Introduction

Addressing the triple planetary crisis (i.e. climate change, biodiversity loss, and pollution) (Larsen & Tararas, 2024) requires coordinated global efforts. Initiatives such as the Sustainable Development Goals (SDGs) (UN, 2015), the Paris Agreement

CONTACT Audrey Lambiel Audrey.Lambiel@unige.ch Institute for Environmental Sciences, enviroSPACE Lab, University of Geneva, Bd Carl-Vogt 66, Geneva 1205, Switzerland

(UNFCCC, 2015), and the Kunming-Montreal Global Biodiversity Framework (GBF) have been established to drive action and foster international collaboration towards a more sustainable and resilient environment. In this context, environmental monitoring is recognized as a critical tool for assessing progress and informing policy decisions (Campbell et al., 2020). However, many existing monitoring systems operate in isolation, designed for specific mandates, and thus provide only a fragmented view of the current state of the environment (Otsu & Maso, 2024).

The loss of biodiversity represents a significant threat to human well-being, due to its role in undermining ecosystem services essential for sustaining life (Díaz et al., 2019). Global initiatives, such as the Kunming-Montreal GBF, aim to address this crisis, with ambitious goals like the Convention on Biological Diversity (CBD) 30×30 goal, i.e. protecting 30% of the planet's land and oceans by 2030 (target 3), as well as restoring another 30% of degraded ecosystems (target 2) (CBD, 2022). These efforts reflect a shift in conservation strategies, from strictly protecting nature for itself to integrating it into human landscapes, balancing ecosystem preservation with societal needs in socio-ecological systems (Mace, 2014).

At the core of this approach is the concept of ecological infrastructure (EI), defined as natural or semi-natural landscape features crucial for delivering ecosystem services (IPBES, 2018). Identifying and then conserving EI is therefore a central and very powerful tool for achieving global conservation goals. The spatial prioritization approach, as tested on a regional (Honeck, Moilanen, et al., 2020; Sanguet et al., 2023) and national scale (Külling 2025) in Switzerland, and in many other parts of the world (Moilanen et al., 2009) offers an effective and systematic method for identifying and prioritizing areas of high conservation value, thereby ensuring an efficient allocation of resources to maximize the protection of biodiversity and the provision of ecosystem services. Safeguarding EI is therefore an essential step towards achieving sustainable conservation results.

Assessing and designing an EI generally relies on a large and complex set of information layers that describe the different environmental dimensions that are intended to protect. Systematic Conservation Planning (Margules & Pressey, 2000) has been developed to guide this complex process from data collection, to setting goals with decision makers, to identify optimal network of protection, and finally to monitor its implementation and effectiveness. It is therefore a long-term scientific, technical, social, political, and economic process that often needs to be approached iteratively. As it involves several stakeholders, it is also important to provide reproducible and transparent information about data and methods used to implement the proposed solutions, and to keep track of different versions.

The scientific community has increasingly emphasized the need for open and reproducible research practices to improve the reliability, efficiency, and credibility of scientific outputs (Munafò et al., 2017). Nevertheless, the adoption of standardized data-sharing guidelines remains limited across many scientific disciplines (Stall et al., 2019). To promote the broad reusability of scientific data, the Findable, Accessible, Interoperable, and Reusable (FAIR) principles, a set of principles for scientific data management and stewardship, have been proposed as a reference framework (Wilkinson et al., 2016). However, adherence to the FAIR principles alone does not guarantee open science, particularly within the field of environmental research, where openness has not yet been systematically integrated (Giuliani et al., 2021). While FAIR compliance facilitates data discovery and reuse, open data policies are also required to fully realize open science objectives (Ramachandran et al., 2021).

Over the past decade, DTs have gained significant attention across various scientific and engineering disciplines. Initially developed for industrial applications, DTs have progressively expanded into sectors such as healthcare, energy, and more recently, environmental sciences (Otsu & Maso, 2024). Thanks to advancements in digital technologies, DTs now offer a transformative approach to modelling and understanding complex natural and social phenomena (Annoni et al., 2023). By integrating vast amounts of environmental data, derived from remote sensing, in situ observations, and artificial intelligence, DTs enable the simulation and prediction of ecological and climatic processes with unprecedented precision (Nativi et al., 2021).

The conceptual view and definition of DTs vary across industrial, scientific, and standardization communities (DiMarzo et al., 2022). According to El Saddik (2018), a DT is *"a digital replica of a living or non-living physical entity"*. This concept involves bridging the physical and virtual worlds by enabling seamless data transmission, thereby ensuring that the virtual counterpart dynamically reflects real-world changes. By decoupling digital systems from their physical entities, DTs facilitate modifications and optimizations without directly altering the physical counterpart. Moreover, they enable advanced data-driven modelling approaches that extend beyond traditional observation-based methods (Nativi et al., 2020).

To date, DTs have been explored for improving simulation and predictive capabilities across multiple environmental domains, including agriculture (Purcell & Neubauer, 2023; Tagarakis et al., 2024), ecology (De Koning et al., 2023), climate (Nativi et al., 2021), land cover and land use (Kussul et al., 2025) and Earth system science (Guo et al., 2020; Li et al., 2023). Importantly, DTs are not merely large-scale models that aggregate vast datasets and machine learning algorithms. Their strength lies in the integration of data, computational models, and domain expertise, ensuring continuous alignment with real-world dynamics. By leveraging this synergy, DTs offer a novel paradigm for environmental monitoring and decision-making, ultimately contributing to more effective strategies for addressing global environmental challenges (Saltelli et al., 2024, 2025).

Based on the above considerations, the aim of this technical note is to present a reproducible approach that follows the FAIR principles and DT concepts, with the intention of facilitating the production and updating of EI. The benefits and limitations of this approach will be exemplified through the EI of the Canton of Geneva in Switzerland. Geneva represents a particularly relevant case for EI planning due to its highly urbanized and fragmented landscape, which poses significant challenges for biodiversity conservation. The region has also been the focus of previous methodological developments in EI identification (Honeck, Moilanen, et al., 2020; Sanguet et al., 2023). These efforts have resulted in a rich set of spatial and ecological data as well as a network of engaged stakeholders, making Geneva an ideal pilot region for testing and implementing a reproducible and automated DT approach.

2. Background and objectives

This work builds upon the EI identification methodology developed by a local network of experts groups (Honeck et al., 2021), with the primary objective of automating and operationalizing it within a reproducible DT framework.

The methodology to identify the Geneva EI is based on an assessment of the ecological quality of the territory, hereafter named biodiversity diagnosis (Lambiel et al., 2024). The general framework of this method is illustrated in Figure 1. Biodiversity diagnosis relies on four pillars, each representing a set of indicators that collectively aim to characterize biodiversity from a specific perspective. To facilitate the integration of these various dimensions of biodiversity, the spatial prioritization tool Zonation (Moilanen et al., 2022) is employed to rank each spatial unit. Using an iterative algorithm, each cell across the entire territory is assigned a unique value reflecting its relative importance compared to others. The resulting product is what we will refer to as a biodiversity diagnosis' from now on. Using the Zonation software, this approach identifies areas that optimally cover a representative and complementary set of biodiversity dimensions, such as species diversity (Sanguet et al., 2022), ecosystem services (Honeck, Moilanen, et al., 2020), ecological structure and connectivity (Urbina et al., 2023, 2024). The algorithm ensures that selected areas complement one another, maximizing coverage across all dimensions while minimizing redundancy. According to target 3 of the CBD (2022), the international

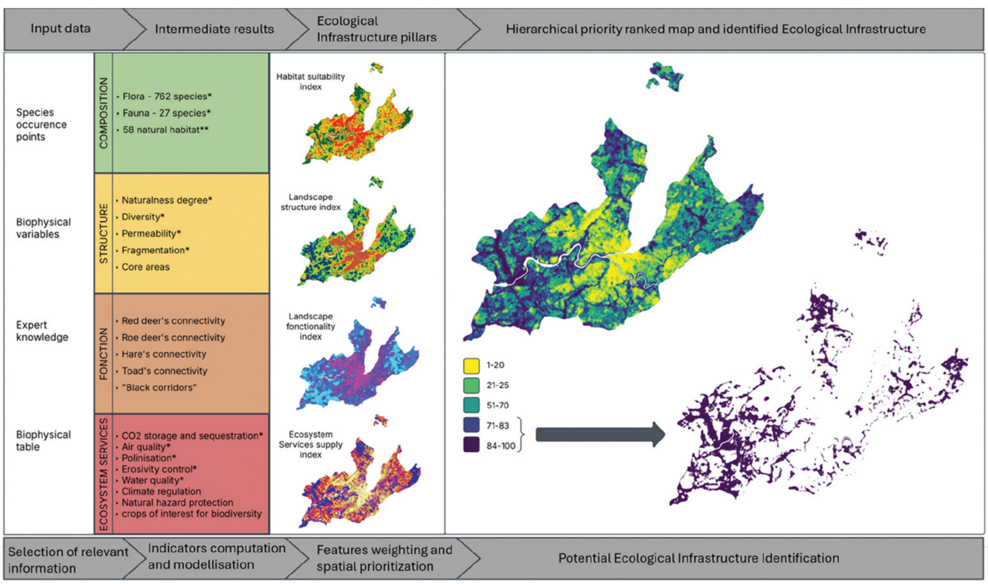

Figure 1. General framework for establishing the ecological infrastructure in the Canton of Geneva based on four pillars of biodiversity dimensions, following the methodology of Honeck, Moilanen, et al. (2020), Honeck, Sanguet, et al. (2020), Sanguet et al. (2023) and Urbina et al. (2024). After hierarchically ranking all the pixels between 1 and 100, only the best 30% are retained and then identified as the ecological infrastructure. The presented work focuses on automating this existing method. *indicator is currently selected and fully implemented in the automation process. **indicator is currently selected and partially implemented (i.e. automation has been done for all natural habitats, but not for isolated trees layer) in the automation process.

treaty that promotes biodiversity conservation and sustainable use of natural resources, the Geneva EI can be determined by identifying the 30% of the territory with the highest priority score from the biodiversity diagnosis. This represents an area where the most representative proportion of the selected dimensions is covered by biodiversity, as considered in systematic conservation planning (Honeck, Moilanen, et al., 2020).

This method defines numerous indicators that enable a comprehensive analysis of biodiversity from complementary perspectives. However, the construction of these pillars requires working with large volumes of often varied data (tables, raster data, GPS coordinates, etc.) sourced from different institutions (cantonal or federal offices, universities, associations, etc.). When these datasets are updated, the heterogeneity of data, coupled with the variety of methods used to calculate indicators, makes it both complex and inefficient to update the EI manually. This hinders the reproducibility of the process and makes comparability over time almost impossible. Nevertheless, regular updates are essential to ensure the EI remains relevant and coherent.

To address this challenge, a key objective is to automate the calculation of indicators, with the aim of facilitating the updating of the EI for the canton of Geneva or the Greater Geneva region. Additionally, the complexity of managing core datasets, particularly in terms of storage, versioning, and updates, can affect the efficiency of the workflow and, indirectly, the accuracy of the generated indicators. Inconsistent or outdated data inputs, as well as errors generated during manual updates, may lead to differences in the final data and ultimately impact the overall quality of the identified EI. The main limitations identified in this regard include accessibility to base data and generated indicators, the availability of accompanying metadata (e.g., data sources, update dates, and applied processing methods), and transparency of computational workflows (including input data, software, algorithms, and parameter settings). These aspects are crucial for ensuring the traceability of different versions of indicators and the EI itself.

While the methodology for identifying EI has been previously developed (Honeck et al., 2021; Honeck, Moilanen, et al., 2020; Sanguet et al., 2023), the present work has as primary objective to automate and strengthen the existing methodology by embedding it within a reproducible and scalable DT framework. This includes centralizing data, streamlining indicator calculations, and enabling transparent versioning and collaborative workflows. By doing so, we aim to enhance the efficiency, consistency, and long-term sustainability of EI planning.

In response to these challenges, it is considered essential to implement a shared working environment via a server, combined with the automation of indicator calculations used in EI identification. Beyond improving data management and facilitating the traceability of produced EI, this approach also enhances the reproducibility and transferability of computational processes. This enables the potential application of this method to other regions around the world.

3. Implementation

To achieve these objectives, we have established a shared working environment using JupyterLab (Perkel, 2018), an interactive and flexible web-based computational development tool that provides significant processing power. Initially, efforts were focused on structuring the working environment to centralize all the essential base data required for calculating indicators, as well as the different versions of these indicators. Subsequently, we automated the calculation of selected indicators relevant to identifying the cantonal EI. These indicators are particularly suited to automation due to the repetitive and time-consuming nature of their calculation processes, as well as their high resource requirements.

The automation has been implemented using scripts written in the R programming language (R Core Team, 2022). Our objective was to minimize user intervention in the computation process by requiring only the input of data file paths, while ensuring maximum transparency. To this end, all scripts are fully annotated. Additionally, metadata for each calculated indicator is automatically generated to improve version traceability and monitor computational performance. Figure 2 below illustrates the overall structure envisaged for this working environment.

Implementing this on the server first requires careful consideration of the shared environment architecture. This environment can be accessed on the JupyterLab server as a shared folder. It consists of directories containing the data necessary for calculating EI indicators, as well as scripts that allow users to initiate the automated calculation of these indicators or the prioritization of the territory to identify the EI itself. This can be done simply by executing dedicated scripts. More specifically, the architecture of the shared environment is structured as follows (Figure 2):

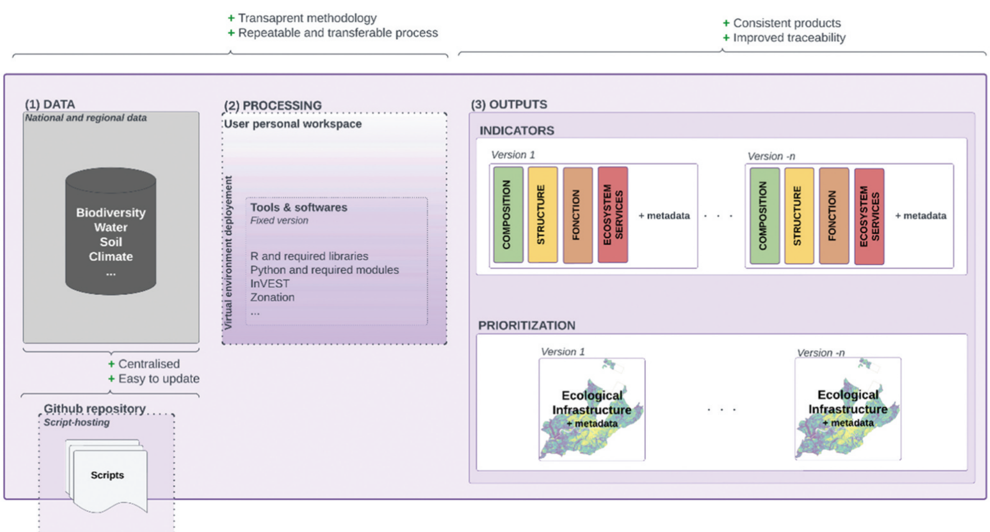

Figure 2. Architecture of the shared JupyterLab environment and the automation process for EI calculation and its indicators. The shared environment, represented by purple boxes, contains several folders, represented by solid rectangles, including (1) DATA, a shared database containing all necessary data for indicator computation, and (3) OUTPUTS, an output folder to store each version of indicator sets and output of prioritization. To allow (2) PROCESSING, each user can access shared resources in the shared environment through their personal workspace and deploy a virtual environment (dashed rectangles), which gives them access to the required tools and software. All the scripts needed for the process are stored in a GitHub repository (https://github.com/ALambiel/envirospace_IE) and can be accessed through the JupyterLab interface.

(1) DATA: A centralized database has been established to facilitate data accessibility and retrieval. This structure simplifies the distribution and use of new or updated datasets. In our work, calculating indicators requires the use of spatial data, such as raster files (e.g., habitat maps and environmental variables), vector layers (e.g., species occurrence points), or data tables (e.g., biophysical tables). These datasets originate from various sources and in different formats. Bringing them together in one place not only makes it easier to update the data when a new version is available but also ensures that all users are working with the same version. Furthermore, data centralization prevents each user from maintaining a separate copy, thereby saving storage space. In this work, the datasets used for indicator computation were primarily collected manually from various institutional sources, including cantonal and federal offices and universities. Dataset specific sources are described in related scientific papers (Honeck, Moilanen, et al., 2020; Sanguet et al., 2023; Urbina et al., 2024). For our work, manual collection was necessary due to the heterogeneity of the data sources and their access protocols. Once collected, the datasets were pre-processed. This included harmonizing coordinate reference systems and resolution for rasters as well as cleaning biophysical tables. All datasets were then structured in a centralized database organized by thematic folders (e.g., species, habitats, services). Ideally, data should be named using clear naming conventions and versioning protocols to facilitate traceability and reproducibility.

(2) PROCESSING: The calculation processes are automated using R scripts. These scripts form the core components of the automation process, enabling the selection of relevant datasets, mobilization of necessary software tools, and the storage of results in the appropriate directories. The scripts are fully accessible in a GitHub repository and extensively annotated to maximize understanding of the methodology while minimizing the need for users to make significant modifications. We leveraged the computing capabilities of the JupyterLab environment to optimize performance by parallelizing scripts, notably using the doParallel[1] package. The JupyterLab environment allows for the installation and use of a wide range of tools and software (Table 1). While the environment was deployed on a Linux system for convenience and performance, the architecture and all components are fully compatible with other operating systems such as Windows and macOS, ensuring broad applicability. In our study, we used MaxEnt[2] (implemented in the R package dismo[3]) for the species distribution layers in the "Composition" pillar, InVEST[4] (implemented in the Python module natcap) for several ecosystem services in the corresponding pillar, and Zonation[5], a software tool used for prioritizing indicators across the study area and ultimately identifying the EI. A key step in implementation was therefore setting up environments containing the necessary languages, packages, and modules for the applied methodologies. We used Conda[6] to create these environments and install R and Python[7], and the required modules. These environments were then saved as.yml files, also accessible from the GitHub repository, enabling new users to easily deploy them in their workspace and execute the scripts. An important advantage of this approach is that it standardized software and package versions, preventing compatibility issues that might arise due to updates. Calculations will always be executed in strictly identical environments, regardless of whether they are run in five months, one year, or later. Deployment is

Table 1. Overview of fully implemented indicators, grouped by thematic pillars. For each indicator, the computational tool and the validation method used or proposed by the reference study are listed. The final output of our approach, i.e. the ecological infrastructure, is evaluated thanks to performance curves.

Pillar	Indicator	Tools	Validation method	Reference study
Composition	Fauna and flora species distribution modelling	MaxEnt	Cross-validation and Area Under the Curve (AUC)	Sanguet et al. (2023)
Ecosystem Services	Pollination	InVEST Crop Pollination model	Expert knowledge and literature review	Honeck, Moilanen, et al. (2020)
	Water quality	InVEST Nutrient Delivery Ratio InVEST	Expert knowledge and literature review	
	Erosivity control	InVEST Sediment Delivery Retention model	Expert knowledge and literature review	
	CO_2 storage and sequestration	InVEST Carbon Storage and Sequestration model	Expert knowledge and literature review	
	Air quality	NDVI + empirical formula (Sanguet et al., 2023)		
Structure	Permeability	Binary reclassification	Expert knowledge and literature review	Honeck, Moilanen, et al. (2020); Sanguet et al. (2023)
	Fragmentation	Based on Fragstat MESH tool		
	Naturalness	Focal statistic	Expert knowledge	
	Diversity	Based on Fragstat SHDI tool		
Biodiversity diagnostic/ Ecological Infrastructure		Zonation	Performance curves (e.g., cumulative indicator coverage)	Lehtomäki and Moilanen (2013); Moilanen et al. (2022)

made as simple as possible thanks to the instructions provided in the README file. Once the environments have been set up and the scripts are ready to use, executing them allows the EI indicators to be calculated or the EI itself to be identified. During the calculation process, two files are automatically generated to track progress and retrieve any error messages, if necessary.

(3) OUTPUTS: The resulting outputs are then stored in a folder named after the specified version. Each new version is saved in a dedicated subdirectory, to ensure version of traceability and the proper archiving of previous iterations. Additionally, metadata files are generated to document the outputs of the automation process (input data, parameters used, computation time, and output data), thereby ensuring transparency and traceability.

In its current implementation, twelve selected indicators (Figure 1; Table 1), based on their suitability for automation and the availability of well-defined and shared methodologies, are automated (1 only partially implemented). Although the process allows indicators to be calculated (currently those that have been automated), it is also possible to manually push the use of an indicator that has been produced independently.

This significantly facilitates updates to these indicators and to the Geneva EI. For example, species distribution maps for 706 species (fauna and flora) were generated for the canton of Geneva at a resolution of 25 meters in less than 3h30. Users can easily produce new versions of these outputs by simply specifying the file paths to the core data. They can also access previous versions of these indicators. Each new version is stored in a separate directory, and generation of metadata (including date, username, input and output data, algorithms and parameters used, performance metrics, etc.) is fully automated.

The automation process has been designed to enable intuitive navigation within the workspace while streamlining data management and EI updates. Although a user manual is provided in the GitHub repository, we consider training users on server operations to be essential in order to ensure the tool is used efficiently and consistently.

We embedded a standalone Zonation 5 (Moilanen et al., 2022) AppImage within the JupyterLab environment to enable seamless execution of the prioritization from JupyterLab. Zonation is a spatial conservation planning software that ranks the entire study area by iteratively removing grid cells with the lowest marginal loss of conservation value, computed based on the input indicators. It produces a continuous priority map, assigning values from 0 (lowest) to 1 (highest), based on the relative contribution of each pixel to the conservation objective (e.g., maximizing indicator coverage or favoring areas of high occurrence) (Lehtomäki & Moilanen, 2013).

The script automatically retrieves all spatial indicators and generates the input parameter files required by Zonation, based on user-defined settings such as the choice of prioritization algorithm and indicator weights, before executing the embedded Zonation AppImage. The resulting outputs include a spatially ranked map of EI, performance curves, and summary statistics. This modular and scalable setup can be easily adapted to other geographic scales (e.g., municipalities, cantons, or national boundaries) with minimal script adjustments, assuming adequate data availability.

4. Discussion

4.1. Comparison with existing tools and approaches

Honeck, Sanguet, et al. (2020) have conducted a review on common practices in identifying what authors call "green infrastructure". Traditional approaches often rely on standalone tools such as MaxEnt for species distribution modelling, InVEST for ecosystem services assessment, Fragstat8 for landscape metrics, CircuitScape9 for connectivity. While these tools are robust and widely adopted in ecological research, using them individually typically involves manual configuration, iterative adjustments, and fragmented data workflows. In addition, many researchers rely on desktop GIS platforms such as ArcGIS10, QGIS11, or FME12 to manage and process spatial data. This fragmentation across interfaces and programming languages can be time-consuming, error-prone, and difficult to reproduce, particularly when the methodology implies computation of numerous indicators or updating results across regions. This may lead to the integration of fewer dimensions of the ecological infrastructure, despite the recognised need to include for example diversity, ecosystem services and landscape connectivity (Honeck, Sanguet, et al., 2020). Our approach addresses these limitations by embedding these tools within

Table 2. Comparison between common methods used to identifying ecological infrastructure (Honeck, Sanguet, et al., 2020), and the proposed approach, based on JupyterLab environment and using spatial prioritization with Zonation.

	Common methods	This approach
Number of dimensions	Often limited to 1 or 2	3 partially implemented and 1 more will be implemented in future
Automation and reproducibility	Manual Fragmented workflows	Automated Streamlined workflows
Tool interoperability	Multiple platforms Multiple software versions	Unified environment Fixed software version
Scalability and transferability	Case specific	Designed for scalability across regions
Transparency and documentation	Varies	Open-source code User manual on GitHub

a unified, automated, and reproducible structure (Table 2). By using a single programming language (R) within fixed and stable environments with defined software versions, the entire process, i.e. from indicator calculation to prioritization, can be executed in a streamlined and simplified way.

There is a growing shift in ecological DT development, toward building modular systems that can be easily extended, ensuring that different ecological models and datasets can work together without issue, and supporting ecosystem monitoring and management, from data collection to scenario analysis and decision-making, as emphasized in recent environmental applications (De Koning et al., 2023; Durden, 2025). Recent frameworks such as TwinEco (Khan et al., 2025) emphasize modularity and interoperability in ecological DT. Similarly, De Koning et al. (2023) highlight the importance of continuous model-data fusion and alignment with real-world dynamics. Our work tries to align with these principles by enabling the integration of new indicators and tools, supporting future extensions (e.g., connectivity modelling with CircuitScape), and facilitating automated data ingestion via APIs. Compared to real-time DT focused on sensor-based monitoring, our framework prioritizes reproducibility and collaborative workflows.

4.2. Added value and limitations of the digital twin approach

We have implemented the automated calculation of 12 selected indicators (11 fully implemented, 1 only partially) (Figure 1), which has significantly facilitated updates to these indicators as well as to Geneva EI. The results we obtained for identifying Geneva EI support the usefulness of implementing such an approach in this type of project. JupyterLab's infrastructure provides a collaborative space with an intuitive and flexible user interface, centralized storage, and significant computing power.

Our approach stands out due to its ability to reinforce coherence, centralize information, and thus improve overall project management. JupyterLab's unique features, combined with the way the shared environment has been designed, enable better management of the data used and produced, while ensuring a rigorous methodology. Thanks to this shared environment, data can be easily updated, and the latest version remains accessible to all collaborators. Nevertheless, this implies the implementation of a database management system, whether carried out by a member of staff or automated.

In addition, shared access to annotated scripts ensures transparency of the methodology and enhances traceability, promoting the reproducibility of analyses. This makes it easy to generate new results from updated data while applying the same method systematically. In the context of temporal monitoring, for instance, this guarantees the comparability of results over time. Furthermore, once a new user has been trained in how the environment works, they can easily take over certain tasks.

Finally, the scripts have been optimized to speed up calculations and obtain results more quickly. This means that users can start processing various indicators while continuing with other tasks, without having to repeat each step manually, thereby reducing the risk of human error.

This approach increases efficiency and consistency by making it easier to update and access data and by guaranteeing a transparent, reproducible, and transferable methodology. It also ensures the long-term sustainability of the project.

Although the proposed approach has several advantages, it is facing some resistance from external partners who contribute to the definition of the EI but use different technological solutions for their geospatial analyses. In order to take into account the diversity of tools and workflows used by external partners, and in the event that the user manual fails to convince them that the environment is easy to use, we are considering a flexible configuration that allows for initial data processing or indicator development using the solution of their choice (within JupyterLab or not). However, final integration and production of the EI should be conducted within the proposed architecture, with its advantages in terms of processing, sharing, versioning, transparency and reproducibility. Despite the benefits of automation and reproducibility, quantifying uncertainty across the pipeline remains challenging. Each tool offers its own validation metrics (Table 2), and the final output is evaluated using Zonation's performance curves, which summarize cumulative indicator coverage but not quantify the error.

4.3. Towards a digital twin

The proposed JupyterHub infrastructure is a central component of a larger system, incorporating selected workflows that transform raw data into final knowledge communicated to specific end users (Figure 3). This proposed infrastructure represents a digital twin of Geneva environment that can be expanded in future to include other environmental models. This approach can also be used to explore different climate or land-use scenarios, as well as potential policy decisions and unforeseen events (e.g., Covid crisis), by modifying the data taken as input by the different scripts.

Such DT relies on accessible input data to monitor the state of the environment. As advocated in the GEOEssential project (Lehmann et al., 2020), we strongly advise defining such input data as Essential Variables to fully describe the socio-ecological system under study. This includes essential climate, biodiversity, water and geological variables on the Earth System side, as well as essential urban, agricultural, health, infrastructure and energy variables in the socio-economical side (Lehmann et al., 2022). The current system could be expanded by integrating real-time environmental data, socio-economic indicators, and other Essential Variables. These additions would allow the DT to more comprehensively represent the socio-ecological system and support scenario modeling and policy evaluation.

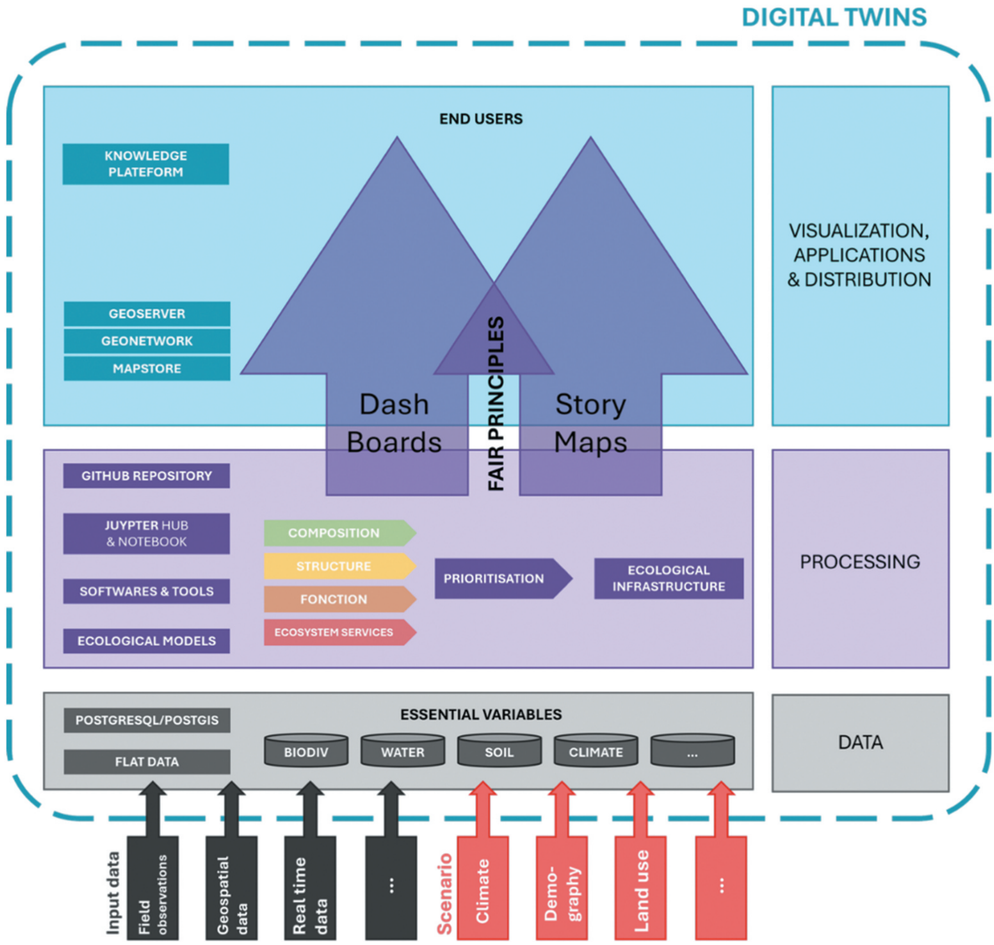

Figure 3. Integration of the JupyterHub solution into the general objective of developing a digital twin for the environment in Geneva. The gray and purple components represent the parts of the system currently supported by our automated infrastructure. To advance toward a higher-level digital twin, as described by DiMarzo et al. (2022), further developments are needed, particularly the integration of a spatial data infrastructure (SDI) component to enable visualization, dissemination, and interaction with outputs via dashboards and web services. The system follows Findable, Accessible, Interoperable and Reusable (FAIR) principles, which guide data management and stewardship. Essential variables include biodiversity (BIODIV), water, soil and climate, complemented by socio-economic factors such as demography and land use. Connecting the system to geospatial data platform via an API would allow for near real-time data updates, reducing manual intervention and supporting the evolution toward a more dynamic and responsive DT.

Although a dedicated user interface has not yet been developed, the system is designed to be interoperable through its integration with the Spatial Data Infrastructure of the Canton of Geneva. The SDI, known as the *Système d'Information du Territoire à Genève* (SITG)13, is a collaborative platform that centralizes and shares geospatial data across public institutions, researchers, and the public. This interoperability allows outputs from our framework to be published as standardized web services, facilitating access and reuse by a wide range of users and projects. As a future development, integrating an API connection to the SITG could enable automated data ingestion into our system, reducing manual intervention. Such automation would support the transition toward a more dynamic DT, in line with the higher levels of integration described by DiMarzo et al. (2022), where real-time data flows between physical and digital systems are essential.

Scaling up the DT will require robust server-based deployment with sufficient computing resources to support multi-user access, large-scale data processing, and integration with external systems. The development of user-friendly interfaces (e.g., dashboard and story map) will also be essential to facilitate stakeholder engagement and decision-making. Several challenges must be addressed, including data heterogeneity across institutions, interoperability of tools and formats, and resistance to adopting new workflows. Ensuring methodological consistency and reproducibility across regions and timeframes will be critical for broader implementation. These developments will not only enhance the technical capabilities of the DT but also strengthen its role as a decision-support tool for ecological planning and environmental governance.

4.4. Spatial data infrastructure and code sharing

In addition to generating model outputs, the DT encompasses a spatial data infrastructure (SDI) that shares these outputs as web services (Giuliani et al., 2017). This SDI also serves as a basis for creating communication products, such as dashboards and story maps. Examples of such story maps can be found in ValPar.CH Project's implementation of the EI at the Swiss scale (https://valpar.unige.ch/mapstore/#/). Furthermore, all the codes used to implement the EI are shared on GitHub to promote greater transparency and reproducibility.

5. Conclusions and perspectives

Implementing the JupyterLab-based infrastructure has demonstrated its ability to support complex workflows efficiently and streamline EI analysis. It provides a structured, automated approach to handling large datasets, performing indicator calculations, and ensuring the traceability and reproducibility of results.

This architecture could be used for the multi-user implementation of complex modeling projects, improving shareability. By centralizing and automating calculations, this approach facilitates collaboration on large-scale modeling projects, enhances shareability, and ensures that all contributors work with standardized datasets and methodologies. However, difficulties remain convincing all partners to adopt this new approach.

The developed system reinforces methodological rigor while improving project management through enhanced version control, transparency, and data consistency. The flexibility of this infrastructure makes it a powerful tool. By enhancing forecasting capabilities, expanding its application to new regions and disciplines, and reinforcing database management and training, this approach holds significant potential for improving the efficiency, consistency, and long-term sustainability of ecological and environmental modeling projects. The foreseen next steps of development are:

(1) Forecasting capacities under different scenarios: the infrastructure can be used for prediction by simply changing the input data, which could then represent the same elements but adapted according to future climate scenarios, for example. This would enable decision-makers to assess the potential impacts of various conservation and land-use policies.
(2) Reproducibility and replicability of a complex process in different regions (other cantons or countries): By maintaining a standardized and automated methodology, local reproducibility and broader replicability can be ensured.
(3) Generalization of the approach to other complex workflows (e.g., hydrological modeling): the structured, automated, and transparent approach demonstrated in this study can be generalized to other complex workflows beyond EI analysis. Potential applications (hydrological modeling for instance) where handling large datasets and ensuring reproducibility are critical.

Notes

1. https://cran.r-project.org/package=doParallel
2. https://biodiversityinformatics.amnh.org/open_source/maxent/
3. https://CRAN.R-project.org/package=dismo
4. https://naturalcapitalproject.stanford.edu/software/invest
5. https://www.helsinki.fi/en/researchgroups/conservation-biology/zonation
6. https://www.anaconda.com
7. https://www.python.org
8. https://www.fragstats.org
9. https://circuitscape.org
10. https://www.esri.com/en-us/arcgis/geospatial-platform/overview
11. https://qgis.org
12. https://fme.safe.com
13. https://sitg.ge.ch

Acknowledgements

The authors would like to thank the Swiss Federal Office of the Environment (FOEN) for their financial support and express their gratitude to the RPT4 pilot committee. The authors gratefully acknowledge the members of the GE21 expert group from the University of Geneva (UNIGE), the Geneva School of Engineering, Architecture and Landscape (HEPIA), the Conservatory and Botanical Garden of Geneva (CJB) and the Cantonal Office for Agriculture and Nature (OCAN) for their previous work on establishing indicators and, more broadly, on developing the method for identifying ecological infrastructure, from which the present work is a continuation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work received funding from Swiss Federal Office of the Environment (Grant RPT "Nature and Landscape 2020" with the Canton of Geneva).

Notes on contributors

Audrey Lambiel is a doctoral candidate affiliated with the Institute of Environmental Sciences, University of Geneva, where she conducts research focusing on land degradation and ecosystem services. She initially specialized in spatial analysis applied to biodiversity and ecosystem services, notably by contributing to the Swiss ValPar.CH project (2020–2024). She was also involved in the work carried out by the GE-21 expert group on Ecological Infrastructure.

Dr. Gregory Giuliani serves as Head of the Digital Earth Unit and Project Leader of the Swiss Data Cube at GRID-Geneva, United Nations Environment Programme (UNEP), and holds the position of Senior Lecturer at the University of Geneva's Institute for Environmental Sciences. He is a geologist and environmental scientist who specializes in Remote Sensing, Geographical Information Systems (GIS) and Spatial Data Infrastructures (SDI). He also works at GRID-Geneva of the United Nations Environment Programme (UNEP) since 2001, where he was previously the focal point for Spatial Data Infrastructure (SDI) and is currently the Head of the Digital Earth Unit. Dr. Giuliani's research focuses on Land Change Science and how Earth observations can be used to monitor and assess environmental changes and support sustainable development initiatives.

Dr. Nathan Külling is a biologist holding an MSc in Behaviour, Evolution and Conservation and a PhD in Life Sciences. He specializes in Geographical Information Systems (GIS), spatial analysis, and spatially explicit models for ecosystem services and species distribution. He currently works to leverage spatial conservation prioritization research to inform and support landscape management.

Prof. Anthony Lehmann is a specialist in Species Distribution Modeling (SDM). More recently, he has focused on leveraging hydrological modeling to inform decision-making processes. He served as coordinator of the FP7 enviroGRIDS project, a four-year endeavour involving over 100 scientists aimed at bridging the gap between scientific data and decision-making in the Black Sea catchment area. Prof. Lehmann's recent research emphasizes spatially explicit assessments of ecosystem services. He coordinated the H2020 ERA-PLANET /GEOEssential project (2017–2021), which developed geoprocessing workflows to link Earth Observation data with environmental policy indicators using Essential Variables. He participated in the ValPar.CH project (2020–2024), exploring ecosystem services, biodiversity, and ecological infrastructures in and around Switzerland's regional parks.

ORCID

Audrey Lambiel http://orcid.org/0000-0002-1512-3623

Gregory Giuliani http://orcid.org/0000-0002-1825-8865
Nathan Külling http://orcid.org/0009-0006-2942-3009
Anthony Lehmann http://orcid.org/0000-0002-8279-8567

AI disclosure statement

The authors report that no AI tools were used in the preparation of this work.

Data availability statement

Most data that support this study are openly available at: https://sitg.ge.ch. The remaining datasets (i.e. not under open data licenses) are available upon request from the authors.

User manual and codes

All scripts needed for the process are stored and described in a GitHub repository: https://github.com/ALambiel/envirospace_IE.
User manual in two versions (English and French) is also available on the repository.

References

Annoni, A., Nativi, S., Çöltekin, A., Desha, C., Eremchenko, E., Gevaert, C. M., Giuliani, G., Chen, M., Perez-Mora, L., Strobl, J., & Tumampos, S. (2023). Digital Earth: Yesterday, today, and tomorrow. *International Journal of Digital Earth*, *16*(1), 1022–1072. https://doi.org/10.1080/17538947.2023.2187467

Campbell, J., Neuner, J., See, L., Fritz, S., Fraisl, D., Espey, J., & Kim, A. (2020). The role of combining national official statistics with global monitoring to close the data gaps in the environmental SDGs. *Statistical Journal of the IAOS*, *36*(2), 443–453. https://doi.org/10.3233/sji-200648

CBD. (2022). *Kunming-Montreal global biodiversity framework. Decision adopted by the conference of the parties to the convention on biological diversity*. https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-04-en.pdf

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*, *38*(10), 916–926. https://doi.org/10.1016/j.tree.2023.04.010

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., … Willis, K. J. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science*, *366*(6471), eaax3100. https://doi.org/10.1126/science.aax3100

DiMarzo, S. G., Cutting-Decelle, A. F., Guise, L., Cormenier, T., Khan, I., & Hossenlopp, L. (2022). Digital twins: From conceptual views to industrial applications in the electrical domain. *Computer*, *55*(9), 16–25. https://doi.org/10.1109/MC.2022.3156847

Durden, J. M. (2025). Environmental management using a digital twin. *Environmental Science & Policy*, *164*, 104018. https://doi.org/10.1016/j.envsci.2025.104018

El Saddik, A. (2018). Digital twins: The convergence of multimedia technologies. *IEEE Multimedia*, *25*(2), 87–92. https://doi.org/10.1109/MMUL.2018.023121167

Giuliani, G., Cazeaux, H., Burgi, P.-Y., Poussin, C., Richard, J.-P., & Chatenoux, B. (2021). SwissEnvEO: A FAIR national environmental data repository for Earth observation open science. *Data Science Journal*, *20*. https://doi.org/10.5334/dsj-2021-022

Giuliani, G., Lacroix, P., Guigoz, Y., Roncella, R., Bigagli, L., Santoro, M., Mazzetti, P., Nativi, S., Ray, N., & Lehmann, A. (2017). Bringing GEOSS services into practice: A capacity building resource on spatial data infrastructures (SDI). *Transactions in GIS*, *21*(4), 811–824. https://doi.org/10.1111/tgis.12209

Guo, H., Nativi, S., Liang, D., Craglia, M., Wang, L., Schade, S., Corban, C., He, G., Pesaresi, M., Li, J., Shirazi, Z., Liu, J., & Annoni, A. (2020). Big Earth data science: An information framework for a sustainable planet. *International Journal of Digital Earth*, *13*(7), 743–767. https://doi.org/10.1080/17538947.2020.1743785

Honeck, E., Gallagher, L., von Arx, B., Lehmann, A., Wyler, N., Villarrubia, O., Guinaudeau, B., & Schlaepfer, M. A. (2021). Integrating ecosystem services into policymaking—a case study on the use of boundary organizations. *Ecosystem Services*, *49*, 101286. https://doi.org/10.1016/j.ecoser.2021.101286

Honeck, E., Moilanen, A., Guinaudeau, B., Wyler, N., Schlaepfer, M. A., Martin, P., Sanguet, A., Urbina, L., von Arx, B., Massy, J., Fischer, C., & Lehmann, A. (2020). Implementing green infrastructure for the spatial planning of peri-urban areas in Geneva, Switzerland. *Sustainability*, *12*(4), 1387. https://doi.org/10.3390/su12041387

Honeck, E., Sanguet, A., Schlaepfer, M. A., Wyler, N., & Lehmann, A. (2020). Methods for identifying green infrastructure. *SN Applied Sciences*, *2*(11), 1916. https://doi.org/10.1007/s42452-020-03575-4

IPBES. (2018). *The IPBES regional assessment report on biodiversity and ecosystem services for Europe and Central Asia*. https://files.ipbes.net/ipbes-web-prod-public-files/2018_eca_full_report_book_v5_pages_0.pdf

Khan, T., de Koning, K., Endresen, D., Chala, D., & Kusch, E. (2025). TwinEco: A unified framework for dynamic data-driven digital twins in ecology. *Ecological Informatics*, *91*, 103407. https://doi.org/10.1016/j.ecoinf.2025.103407

Külling, N. (2025). Translating global commitments into national actions: A framework for identifying the ecological infrastructure in Switzerland [Doctoral thesis]. University of Geneva]. University of Geneva]. pp. 1–148. https://doi.org/10.13097/archive-ouverte/unige:183825

Kussul, N., Giuliani, G., Shelestov, A., Cherniatevych, A., Drozd, S., Kolotii, A., Salii, Y., Yavorskyi, O., Malyniak, V., Lavrenyuk, A., & Poussin, C. (2025). AI-powered digital twin framework for land use change in disaster affected regions IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.18. In (pp. 27473–27492 2151-1535. https://doi.org/10.1109/JSTARS.2025.3623870 .

Lambiel, A., Waller, N., Wyler, N., & Lehmann, A. (2024). *Evolution et consolidation de l'infrastructure écologique-analyse comparative des méthodologies d'identification de l'infrastructure écologique sur le canton de Genève*. University of Geneva. https://archive-ouverte.unige.ch/unige:182161

Larsen, P. B., & Tararas, K. (2024). Editorial: Enhancing the right to science: The triple planetary crisis and the need for comprehensive approaches [editorial]. *Frontiers in Sociology*, *9*, 1406640. https://doi.org/10.3389/fsoc.2024.1406640

Lehmann, A., Mazzetti, P., Santoro, M., Nativi, S., Masò, J., Serral, I., Spengler, D., Niamir, A., Lacroix, P., Ambrosone, M., McCallum, I., Kussul, N., Patias, P., Rodila, D., Ray, N., & Giuliani, G. (2022). Essential earth observation variables for high-level multi-scale indicators and policies. *Environmental Science & Policy*, *131*, 105–117. https://doi.org/10.1016/j.envsci.2021.12.024

Lehmann, A., Nativi, S., Mazzetti, P., Maso, J., Serral, I., Spengler, D., Niamir, A., McCallum, I., Lacroix, P., Patias, P., Rodila, D., Ray, N., & Giuliani, G. (2020). Geoessential-mainstreaming workflows from data sources to environment policy indicators with essential variables. *International Journal of Digital Earth*, *13*(2), 322–338. https://doi.org/10.1080/17538947.2019.1585977

Lehtomäki, J., & Moilanen, A. (2013). Methods and workflow for spatial conservation prioritization using Zonation. *Environmental Modelling & Software*, *47*, 128–137. https://doi.org/10.1016/j.envsoft.2013.05.001

Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., Shen, H., Xiao, Q., Su, J., Yuan, S., & Guo, H. (2023). Big data in Earth system science and progress towards a digital twin. *Nature Reviews Earth & Environment*, *4*(5), 319–332. https://doi.org/10.1038/s43017-023-00409-w

Mace, G. M. (2014). Whose conservation? *Science*, *345*(6204), 1558–1560. https://doi.org/10.1126/science.1254704

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*(6783), 243–253. https://doi.org/10.1038/35012251

Moilanen, A., Lehtinen, P., Kohonen, I., Jalkanen, J., Virtanen, E. A., & Kujala, H. (2022). Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance. *Methods in Ecology and Evolution*, *13*(5), 1062–1072. https://doi.org/10.1111/2041-210X.13819

Moilanen, A., Wilson, K. A., & Possingham, H. P. (2009). *Spatial conservation prioritization: Quantitative methods and computational tools*. Oxford University Press. https://doi.org/10.1093/oso/9780199547760.001.0001

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Nativi, S., Delipetrev, B., & Craglia, M. (2020). *Destination earth - survey on "digital twins" technologies and activities, in the green deal area, EUR 30438 EN*. Publications Office of the European Union. https://doi.org/10.2760/430025

Nativi, S., Mazzetti, P., & Craglia, M. (2021). Digital ecosystems for developing digital twins of the Earth: The Destination Earth case. *Remote Sensing*, *13*(11), 2119. https://doi.org/10.3390/rs13112119

Otsu, K., & Maso, J. (2024). Digital twins for research and innovation in support of the European Green Deal data space: A systematic review. *Remote Sensing*, *16*(19), 3672. https://doi.org/10.3390/rs16193672

Perkel, J. M. (2018). Why Jupyter is data scientists' computational notebook of choice. *Nature*, *563*(7729), 145–146. https://doi.org/10.1038/d41586-018-07196-1

Purcell, W., & Neubauer, T. (2023). Digital twins in agriculture: A state-of-the-art review. *Smart Agricultural Technology*, *3*, 100094. https://doi.org/10.1016/j.atech.2022.100094

Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From open data to open science. *Earth and Space Science*, *8*(5). https://doi.org/10.1029/2020EA001562

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Saltelli, A., Gigerenzer, G., Hulme, M., Katsikopoulos, K. V., Melsen, L. A., Peters, G. P., Pielke, R., Jr., Robertson, S., Stirling, A., Tavoni, M., & Puy, A. (2024). Bring digital twins back to Earth. *WIREs Climate Change*, *15*(6), e915. https://doi.org/10.1002/wcc.915

Saltelli, A., Melsen, L. A., & Puy, A. (2025). Digital twins of the earth between vision and fiction. *Minerva*. https://doi.org/10.1007/s11024-025-09581-3

Sanguet, A., Wyler, N., Guinaudeau, B., Waller, N., Urbina, L., Huber, L., Fischer, C., & Lehmann, A. (2023). Mapping ecological infrastructure in a cross-border regional context. *Land*, *12*(11), 2010. 2010. https://www.mdpi.com/2073-445X/12/11/2010

Sanguet, A., Wyler, N., Petitpierre, B., Honeck, E., Poussin, C., Martin, P., & Lehmann, A. (2022). Beyond topo-climatic predictors: Does habitats distribution and remote sensing information improve predictions of species distribution models? *Global Ecology and Conservation*, *39*, e02286. https://doi.org/10.1016/j.gecco.2022.e02286

Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., & Wyborn, L. (2019). Make scientific data FAIR. *Nature*, *570*(7759), 27–29. https://doi.org/10.1038/d41586-019-01720-7

Tagarakis, A. C., Benos, L., Kyriakarakos, G., Pearson, S., Sørensen, C. G., & Bochtis, D. (2024). Digital twins in agriculture and forestry: A review. *Sensors*, *24*(10), 3117. https://doi.org/10.3390/s24103117

UN. (2015). *Transforming our world: The 2030 agenda for sustainable development*. https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf

UNFCCC. (2015). *Adoption of the Paris Agreement*. https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

Urbina, L., Fischer, C., Ray, N., & Lehmann, A. (2023). Modeling red deer functional connectivity at a regional scale in a human-dominated landscape [original research]. *Frontiers in Environmental Science*, *11*, Volume11–2023. https://doi.org/10.3389/fenvs.2023.1198168

Urbina, L., Lehmann, A., Huber, L., & Fischer, C. (2024). Combining multi-species connectivity modelling with expert knowledge to inform the green infrastructure design. *Journal for Nature Conservation*, *81*, 126654. https://doi.org/10.1016/j.jnc.2024.126654

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Zhao, J. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18