

The Time Machine framework

Eastwood, Niamh; Stubbings, William A.; Abdallah, Mohamed Abou-Elwafa ; Durance, Isabelle; Paavola, Jouni; Dallimer, Martin; Pantel, Jelena H.; Johnson, Samuel; Zhou, Jiarui; Hosking, J. Scott; Brown, James; Ullah, Sami; Krause, Stefan; Hannah, David; Crawford, Sarah; Widmann, Martin; Orsini, Luisa

DOI:

[10.1016/j.tree.2021.09.008](https://doi.org/10.1016/j.tree.2021.09.008)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Eastwood, N, Stubbings, WA, Abdallah, MA-E, Durance, I, Paavola, J, Dallimer, M, Pantel, JH, Johnson, S, Zhou, J, Hosking, JS, Brown, J, Ullah, S, Krause, S, Hannah, D, Crawford, S, Widmann, M & Orsini, L 2022, 'The Time Machine framework: monitoring and prediction of biodiversity loss', *Trends in Ecology & Evolution*, vol. 37, no. 2, pp. 138-146. <https://doi.org/10.1016/j.tree.2021.09.008>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

The Time Machine framework: monitoring and prediction of biodiversity loss

Running title: Biodiversity and Ecosystem Services

N. Eastwood¹, W.A. Stubbings², M.A. Abou-Elwafa Abdallah², I. Durance³, J. Paavola⁴, M. Dallimer⁴, J.H. Pantel⁵, S. Johnson⁶, J. Zhou¹, J.S. Hosking^{7, 8}, J.B. Brown⁹, S. Ullah², S. Krause², D. M. Hannah², S.E. Crawford¹⁰, M. Widmann² and L. Orsini^{1, 8}.

¹Environmental Genomics Group, School of Biosciences, The University of Birmingham, Birmingham, B15 2TT, UK

²School of Geography, Earth & Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK

³School of Biosciences and Water Research Institute, Cardiff University, Cardiff, CF10 3AX, UK

⁴Sustainability Research Institute, School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, UK

⁵Department of Computer Science, Mathematics, and Environmental Science, The American University of Paris, 6 rue du Colonel Combes, 75007 Paris, France

⁶School of Mathematics, University of Birmingham, Birmingham, B15 2TT, UK; and Alan Turing Institute, London, NW1 2DB, UK

⁷British Antarctic Survey, Natural Environment Research Council, Cambridge, CB3 0ET, UK

⁸The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, UK

⁹Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

¹⁰Institute of Ecology, Evolution and Diversity, Dept. of Evolutionary Ecology and Environmental Toxicology, Goethe University Frankfurt, 60438, Germany

Keywords: ecosystem function, AI, time-series, climate, pollution, economic valuation

Abstract

Trans-disciplinary solutions are needed to achieve the sustainability of ecosystem services for future generations. We propose a framework to identify the causes of ecosystem function loss and to forecast the future of ecosystem services under different climate and pollution scenarios. The framework i) applies an artificial intelligence time-series analysis to identify relationships among environmental change, biodiversity dynamics and ecosystem functions; ii) validates relationships between loss of biodiversity and environmental change in fabricated ecosystems; and iii) forecasts the likely future of ecosystem services and their socio-economic impact under different pollution and climate scenarios. We illustrate the framework by applying it to watersheds, and provide system-level approaches that enable natural capital restoration by associating centennial biodiversity changes to chemical pollution.

The challenge of sustained ecosystem services

Biodiversity is directly linked to healthy ecosystems which provide provisioning (e.g. food), regulating (e.g. climate), supporting (e.g. nutrient cycling, primary production), and cultural (e.g. aesthetic and recreational) services [1-4]. At the global level, rapid and severe biodiversity loss has been identified as the main cause of deterioration of more than 60% of ecosystem services [1, 3]), affected by various factors [5]. Chemical pollution, habitat loss, unsustainable use of resources, invasive species and climate change are among the main acknowledged threats to biodiversity [6, 7].

The sustained delivery of ecosystem services in the face of these threats is challenging because natural capital is finite and the impact of human interventions on ecosystems is uncertain and/ or unknown across different spatial, temporal, and economic scales [8]. Ecosystem management that ensures the delivery of services while preserving natural capital is a complex, open-ended problem because of limited resources, competing objectives and the need for economic profitability [9]. This is because:

i) biodiversity loss happens on different spatial and temporal scales, and dynamics are context-dependent outcomes stemming from processes operating over many years [10, 11]. The cumulative effect of processes over time can cause delayed dynamics also known to cause so-called extinction debts [12]. Even state-of-the-art environmental monitoring fails to capture effects that may arise from cumulative effects over time of multiple threats [5]). Only by quantifying trajectories of abiotic, biotic and functional systemic change before, during and after pollution events, can we identify the causes of biodiversity and ecosystem function loss;

ii) research on biodiversity and ecosystem services is often constrained by disciplinary boundaries, whereas biodiversity loss has ecological, social and economic implications [13]. Discipline-constrained approaches may neglect process interactions, result in research undertaken at inappropriate or disconnected scales, or use discipline-specific tools that are inadequate to address cross-disciplinary questions [14].

iii) decision-making frameworks that enable the prioritization of interventions for the sustainable use of ecosystems typically require multiple lines of evidence from different disciplines, making decisions by stakeholders challenging, especially when relationships between socio-economic and ecological priorities are not linear (e.g. [15, 16]). While such decision-making frameworks are being developed, they still often fail to cover all types of ecosystem services, particularly the cultural ones [15].

We propose a ‘Time Machine’ framework that: 1) establishes spatiotemporal correlations among biotic, abiotic and ecosystem functional changes using multidecadal to millennial continuous data; 2) provides evidence for cause-effect relations through experimental validation in fabricated ecosystems from correlations identified in point 1; and 3) generates likely predictions of future ecosystem services under different pollution and climate scenarios, driven by localised and regional environmental change and mediated by changes in overall biotic interactions (Fig. 1).

The Time Machine framework

We illustrate the framework in five main *Steps* for applications in freshwater ecosystems, because they are diverse, geographically distributed and of high conservation value (Fig 1); they deliver important ecosystem services such as clean water, food and recreation, and are under increasing threat of destruction and degradation [2, 17].

Step 1 - *Sampling* through time and space, using Lake sedimentary archives (Fig. 1; *Sampling*). Sedimentary archives preserve biological and environmental signals, providing a

continuous record of changes from a reference baseline predating major human impact on climate and biodiversity (Anthropocene) to impacted environments [18]; these characteristics allow better understanding of temporal dynamics of biotic, abiotic and ecosystem functional data leading to current patterns [10]. To disentangle patterns driven by stochasticity from patterns driven by environmental change, semi-pristine sites (e.g. alpine lakes) can be used as reference. Whereas all natural communities experience changes over time, dynamics in semi-pristine sites will likely be driven by demographic and environmental stochasticity, which results in predictable community dynamics [19, 20]. Conversely, dynamics at impacted sites will likely be driven by exogenous environmental factors, which leave signatures that depart from stochasticity (e.g. [21]).

Step 2 - Biochemical and ecosystem functions *fingerprinting*. Spatiotemporal biotic assemblages are established by applying **metabarcoding to environmental DNA or eDNA** (see Glossary) [22] extracted from dated sediment layers of the biological archives (Fig. 1; *Fingerprinting*). DNA extracted from environmental matrices provides the means to collect continuous temporal data over time and space [23]. These data inform conservation biology by estimating human impact on biodiversity [24], invasion biology by identifying timing and severity of alien species invasion [25] and biodiversity science by enabling the census of species/taxa on a global scale in real time [26]. Through DNA sequence similarity, molecular operational taxonomic units (MOTUs) are identified by matching sequence similarity to records in public databases (e.g. NCBI, SILVA), allowing the analysis of compositional shifts and relative abundance of known and unknown taxonomic units. MOTUs enable the retrospective characterization of past community-level dynamics (e.g. microbes, plants and animals) without requiring specialist skills (e.g. light microscopy and taxonomy) and without being limited to taxonomic groups with well-preserved remains in environmental matrices (e.g. pollen, exoskeletons, bones) [27]. On the same sediment samples, abiotic changes (e.g. presence and concentrations of chemical mixtures) are quantified using high resolution mass spectrometry (MS), combining **suspect and non-target screening** of chemical pollutants in sediment samples [28] (Fig. 1; *Fingerprinting*). The spatiotemporal biochemical fingerprinting is complemented by the analysis of ecosystem functions [e.g. biogeochemical functions measured as the accumulation rate of total organic carbon (C), nitrogen (N) and phosphorus (P)] via bulk stoichiometry of sediments to help elucidate long-term dynamics in productivity as influenced by nutrient availability, and the relationships of stoichiometric ratios, productivity, decomposition and biological attributes. A potential caveat of reconstructing temporal community records from sediment matrices is the preservation state of DNA that may be affected by climatic conditions (e.g. warmer/humid climates influence mineralization; Outstanding Questions). However, metabarcoding has been successfully applied to fossil or remnant DNA as far back as the Holocene (e.g. [29]). In addition, only relative abundance of MOTUs can be quantified from community assemblages. This enables the reconstruction only of relative changes in MOTU abundance between sites and time points.

Step 3 - Establishing *associations* between past biodiversity, abiotic and functional changes with the use of Artificial Intelligence (AI) (Fig. 1; *Association*). Combining **explainable network models** with **multi-view learning** [30], co-varying elements (nodes) within and between networks are identified, where networks can represent MOTUs, environmental variables and pollutants (Box 1). This leads to the identification of interacting environmental factors putatively driving changes in MOTUs and ecosystem functions. These correlations are then validated experimentally in fabricated ecosystems as explained in Step 4. In a pilot study conducted on a natural lake, we applied the AI approach and determined that the decline in a specific taxonomic group of primary producers (e.g. green algae) was inversely

correlated with ten herbicides among the hundreds that were quantified in the sediment (Box 1).

Step 4 - Establishing *causality*. Fabricated ecosystems are used to establish causal links between the associations identified in Step 3 (Fig. 1; *Causality*). Focusing on short-term dynamics (weeks to months), fabricated ecosystems, such as artificial ponds, are used to isolate effects, explore controlled interactions among multiple factors, and determine causality among MOTUs' relative abundance (quantified via eDNA metabarcoding), ecosystem functions (P, N, C content measured through nutrient stoichiometry), climate variables and chemical pollutants (quantified with MS). Natural communities are used to inoculate fabricated ecosystems, which are then exposed to the environmental factors identified in Step 3 to co-vary with MOTUs. For example, the effect of global warming on biodiversity and ecosystem functions can be quantified by measuring MOTUs dynamics in fabricated ecosystems exposed to current temperature and in ecosystems exposed to temperature plus 2.5°C, representing the IPCC mid-range forecasts (RCP 6.0). While experimental results investigating short-term dynamics may not be directly comparable to natural dynamics, they provide a validation of observed trends in natural ecosystems for extrapolations to long-term dynamics using e.g. **machine learning** algorithms. The fabricated ecosystems serve a dual role of providing an experimental validation of a) observed past correlative patterns, and b) predictions of biodiversity and ecosystem functional changes in different climate and pollution scenarios (see Step 5).

Step 5 - *Forecasting* biodiversity and ecosystem services. **Ecological process-based models**, informed by the associations identified by the AI in Step 3, are used to generate predictions about projected future states of freshwater ecosystems (Fig. 1; *Forecasting*). These predictive models are incorporated into simulations that project solutions for ecosystem services (Fig. 2). Although it is likely that a range of complex processes interact to determine how environmental drivers and associated biodiversity shifts influence ecosystem functions, we propose to start with a simple model that illustrates the impacts of community composition and structure, species interactions, and the covariation of these with environmental factors on freshwater ecosystem functions. Process-based models with these components include PCLake [31], a process-based model that links species composition, environmental drivers (e.g. nutrient loading, temperature), and ecosystem responses (e.g. water quality; [32]). Alternatively, ELCOM-CAEDYM, a coupled hydrodynamic and biological model of phytoplankton dynamics and their impacts on water quality can be used [33]. For reviews on **process-based models** see (e.g. [34]). Inputs for these models can be provided from correlative approaches [35, 36], such as **correlative ecological models** (e.g. [37]), or AI-based inferences such as described in Step 3.

The Time Machine framework is widely applicable because it finds signatures that recapitulate community dynamics (e.g. loss of MOTUs) driven by environmental change (e.g. warming, pollution) that alter ecosystem function (e.g. nutrients ratio). For example, it is applicable to physical, chemical and biological long-term observations of oceanic and coastal areas available from Continuous Plankton Recorders Survey, collecting records from the North Atlantic and the North Sea since 1931 [38]. However, the ecological models that best describe these ecosystems may differ from the ones described here.

The Time Machine Emulator

Continuous temporal data are paramount to explain present-day patterns and to model relationships between biodiversity and environmental change [18]. In particular, palaeoecological data are a valuable source of temporal data and can inform the conservation management of future biodiversity (e.g. [39]). However, past dynamics may differ from that caused by future environmental changes or threats, introducing uncertainties in model estimates. To reduce uncertainty in forecasts, realistic dynamic interactions among several biotic and abiotic variables should be used (e.g. [40]). Generating predictions that account for all these variables in different scenarios (e.g. projected IPCC climate scenarios) is computationally intensive and time consuming. **Emulators** can provide robust predictions with calculated uncertainties across multiple scenarios while reducing computational cost and time. An ‘emulator’ is a low-order, computationally efficient model which emulates the specified output of a more complex model in function of its inputs and parameters. Emulators work with both structure-based methods, where the mathematical structure of the original model is manipulated to a simpler, more computationally efficient form; and data-based approaches, where the emulator is identified and estimated from empirical/experimental data [41]. Emulators are widely applied in big data science, such as i) climate science to generate predictions under different socio-economic scenarios in long-range simulations (e.g. [42] and references therein); ii) ecology to predict the status of ecological processes in changing environments using e.g. long-term remote sensing data [40]; and iii) environmental science to predict e.g. the hydrological status of water reservoirs [43].

We propose a *Time Machine Emulator* (TME) to generate forecasts of biodiversity and ecosystem functions. The TME ‘learns’ from past correlations, it is tested iteratively against long-term empirical data (e.g. collected in Step 3) and refined to predict the future biotic, abiotic and functional associations (Fig. 2). In our applications, it emulates ecological process-based models (e.g. [32, 44]) to generate predictions across a range of historical and future climates and ecosystem states, albeit in a more computationally scalable and efficient manner. Empirical data collected from sedimentary archives, including biodiversity changes over time and pollution, as well as climate data, are used to establish past correlations that inform ecological process-based models (Fig. 2A). These models integrate historical contamination trends that disrupt ecosystem functions mediated by changes in MOTUs, to identify risk trajectories, with measured uncertainties (Fig. 2B). The TME is applied to produce the same output as the process-based models without the time consuming and computationally expensive running of the vast number of possible model setups for every possible climate and pollution scenario (Fig. 2C). One of the biggest challenges for AI on medium to long-range timescales is the lack of high-quality data for training model predictions. The TME uses empirical past biodiversity-ecosystem dynamics (collected in Step 3), as well as temperature projections from regional climate simulations, as ‘training sets’ to reduce uncertainty in forecasting the future of ecosystem states under different climate scenarios (Fig. 2C). Predictions are also made over non-simulated regions and climates with higher uncertainty, informing strategies for additional sampling, in this way mitigating uncertainties when new data are introduced in the model (Fig. 2D). This directs sampling efforts improve the accuracy of forecasts. The output of the TME are predictions for changes in ecosystem functions driven by localised and regional climate change and mediated by changes in overall biotic/abiotic interactions, including on partial training sets. The robustness of the TME predictions is assessed by removing data (i.e. locations and conditions) from complete training sets and comparing predictions on whole and partial datasets. The future associations predicted by the TME are validated using fabricated ecosystems as described in Step 4. By identifying the MOTUs and ecosystem functions altered

by abiotic factors and/or climate variables, the TME provides probabilistic decision-support information for ecosystem services management (Fig. 2E).

Concluding remarks

Lack of understanding of the processes that underpin ecosystem services has often led to mismanagement with clear dis-benefits for the environment, the economy and human well-being. Systemic approaches, like the framework presented here, enable the prioritization of interventions that accelerate ecological restoration, and mitigate environmental factors that cause harm to MOTUs associated with key ecosystem functions and services. However, it requires critical changes in current environmental practice and a shift to whole-system evidence-based approaches. To overcome socio-technical barriers to adoption, stakeholder-enabled platforms are needed that include: 1) experimental protocols for the routine systemic biomonitoring of biodiversity; 2) a TME dashboard relying on Data Visualization Technology tested for ‘usability’ and ‘utility’; and 3) changes in legislation and regulatory practices to facilitate the adoption of novel tools and frameworks.

The use of the Time Machine framework comes with caveats. Process-based models cannot be easily validated in future climate regimes that have not been observed. To mitigate this, ecosystems over diverse climatic regions (e.g., warmer and drier environments) can be included as analogues of future climate states (see Outstanding Questions). In addition, known biotic interactions can be included in the model with information metrics, such as **Granger causality** [45]. This approach infers biotic interactions using the time series empirical data collected from sedimentary archives and/or other known associations (e.g. predator/prey interactions), with estimated uncertainties. Uncertainties in the TME can be reduced by constraining the models with available past and projected climate data collected from local weather stations, globally gridded ERA5-Land datasets [46], and Earth Observation data (<https://earthdata.nasa.gov/earth-observation-data>). Downscaling methods that use either physically-based regional climate models or statistical models to project future large-scale climate [47, 48], can be used to increase spatial granularity down to <10km and even to single point locations (e.g. [48, 49]). Although it is expected that uncertainty increases with increasing granularity, it may be reduced by using spatial scales for which the temporal variability is well simulated.

The TME applications can be, in principle, extended beyond predictions based on the ecological and functional status of ecosystems. By coupling ecological and economic modelling, the TME can also enable the alignment of socio-economic and ecological outcomes under different climate and pollution scenarios (see Outstanding Questions). To overcome adoption barriers by stakeholders, an AI-based Emulator dashboard can be developed, accessible to regulators and policy makers through data visualizations techniques. These tools can be adapted for probabilistic predictions of ecosystem services to aid decision-making and socio-economic trade-offs.

Acknowledgments

We thank William Scavone, Kestrel Studio, USA, for the artwork in Figure 1 and Chantal Jackson for the artwork in Figure 2. This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1, the Natural Environment Research Council (NE/N005716/1) and a MIBTP-BBSRC PhD fellowship to NE (BB/M01116X/1).

References

1. Cardinale, B.J. et al. (2012) Biodiversity loss and its impact on humanity. *Nature* 486 (7401), 59-67.
2. Durance, I. et al. (2016) The challenges of linking ecosystem services to biodiversity. *Adv Ecol Res* 54, 87-134.
3. Mace, G.M. et al. (2012) Biodiversity and ecosystem services: a multilayered relationship. *Trends Ecol Evol* 27 (1), 19-26.
4. Spaak, J.W. et al. (2017) Shifts of community composition and population density substantially affect ecosystem function despite invariant richness. *Ecol Lett* 20, 1315-1324.
5. Bonebrake, T.C. et al. (2019) Integrating Proximal and Horizon Threats to Biodiversity for Conservation. *Trends Ecol Evol* 34 (9), 781-788.
6. Backhaus, T. et al. (2012) The impact of chemical pollution on biodiversity and ecosystem services: the need for an improved understanding. *Integr Environ Assess Manag* 8 (4), 575-6.
7. Jackson, M.C. et al. (2016) Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Glob Change Biol* 22 (1), 180-189.
8. DeFries, R. and Nagendra, H. (2017) Ecosystem management as a wicked problem. *Science* 356, 265-270
9. Jax, K. et al. (2018) Handling a messy world: Lessons learned when trying to make the ecosystem services concept operational. *Ecosyst Serv* 29, 415-427.
10. Nogues-Bravo, D. et al. (2018) Cracking the Code of Biodiversity Responses to Past Climate Change. *Trends Ecol Evol* 33 (10), 765-776.
11. Orsini, L. et al. (2013) The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments. *Trends Ecol Evol* 28 (5), 274-82.
12. Figueiredo, L. et al. (2019) Understanding extinction debts: spatio-temporal scales, mechanisms and a roadmap for future research. *Ecography* 42, 1973-1990.
13. Barnosky, A.D. et al. (2016) Avoiding collapse: Grand challenges for science and society to solve by 2050. *Elementa* 4, 000094.
14. Roux, D.J. et al. (2017) Transdisciplinary research for systemic change: who to learn with, what to learn about and how to learn. *Sustain Sci* 12 (711-726).
15. Grêt-Regamey, A. et al. (2017) A tiered approach for ecosystem services mapping. In *Mapping Ecosystem Services* (Burkhard, B. and Maes, J. eds), Pensoft Publishers.
16. Pinto, R. et al. (2014) Mainstreaming Sustainable Decision-making for Ecosystems: Integrating Ecological and Socio-economic Targets within a Decision Support System. *Environ Proc* 1, 7-19.
17. Ruckelshaus, M.H. et al. (2020) The IPBES Global Assessment: Pathways to Action. *Trends Ecol Evol* 35 (5), 407-414.
18. Baert, J.M. et al. (2016) Per capita interactions and stress tolerance drive stress-induced changes in biodiversity effects on ecosystem functions. *Nat Commun* 7, 12486.
19. Shoemaker, L.G. et al. (2020) Integrating the underlying structure of stochasticity into community ecology. *Ecology* 101, e02922.
20. Vellend, M. et al. (2014) Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos* 123, 1420-1430.
21. Wang, Y. et al. (2018) Temporal effects of disturbance on community composition in simulated stage-structured plant communities. *Ecol Evol* 8 (1), 120-127.
22. Cristescu, M.E. and Hebert, P.D.N. (2018) Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation. *Annu Rev Ecol Evol S* 49, 209-230.
23. Balint, M. et al. (2018) Environmental DNA Time Series in Ecology. *Trends Ecol Evol* 33 (12), 945-957.

24. Tse, T.J. et al. (2018) Combining High-Throughput Sequencing of sedaDNA and Traditional Paleolimnological Techniques To Infer Historical Trends in Cyanobacterial Communities. *Environ Sci Technol* 52 (12), 6842-6853.
25. Ruppert, K.M. et al. (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *GECCO* 17, e00547.
26. Garlapati, D. et al. (2019) A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. *Rev Environ Sci Biotechnol* 18, 389–411.
27. Gillson, L. and Marchant, R. (2014) From myopia to clarity: sharpening the focus of ecosystem management through the lens of palaeoecology. *Trends Ecol Evol* 29 (6), 317-25.
28. Abdallah, M. et al. (2019) A single run, rapid polarity switching method for determination of 30 pharmaceuticals and personal care products in waste water using Q-Exactive Orbitrap high resolution accurate mass spectrometry. *J Chromatogr A* 1588, 68-76.
29. Hirai, J. et al. (2017) Cryptic Zooplankton Diversity Revealed by a Metagenetic Approach to Monitoring Metazoan Communities in the Coastal Waters of the Okhotsk Sea, Northeastern Hokkaido. *Front Mar Sci* 4.
30. Tenenhaus, A. et al. (2014) Variable selection for generalized canonical correlation analysis. *Biostatistics* 15 (3), 569-83.
31. Janssen, A.B.G. et al. (2019) PCLake+: A process-based ecological model to assess the trophic state of stratified and non-stratified freshwater lakes worldwide. *Ecol Model* 396, 23-32.
32. Janse, J.H. et al. (2010) Estimating the critical phosphorus loading of shallow lakes with the ecosystem model PCLake: sensitivity, calibration and uncertainty. *Ecol Model* 221, 654-665.
33. Robson, B.J. and Hamilton, D.P. (2004) Three-dimensional modelling of a Microcystis bloom event in the Swan River estuary, Western Australia. *Ecol Model* 174, 203-222.
34. Connolly, S.R. et al. (2017) Process, Mechanism, and Modeling in Macroecology. *Trends Ecol Evol* 32 (11), 835-844.
35. Brophy, C. et al. (2017) Biodiversity and ecosystem function: making sense of numerous species interactions in multi-species communities. *Ecology* 98 (7), 1771-1778.
36. Ovaskainen, O. et al. (2017) How are species interactions structured in species-rich communities? A new method for analysing time-series data. *P Roy Soc B* 284, 20170768.
37. Peterson, A.T. et al. (2015) Mechanistic and Correlative Models of Ecological Niches. *Eur J Ecol* 1, 28-38.
38. Survey, C., The continuous plankton recorder survey, in: Association, M.B. (Ed.) Marine Biological Association, Plymouth, UK, 2020.
39. Fordham, D.A. et al. (2020) Using paleo-archives to safeguard biodiversity under climate change. *Science* 369, eabc5654.
40. Leeds, W.B. et al. (2014) Emulator-assisted reduced-rank ecological data assimilation for nonlinear multivariate dynamical spatio-temporal processes. *Stat Methodol* 17, 126-138.
41. Castelletti, A. et al. (2012) A general framework for Dynamic Emulation Modelling in environmental problems. *Environ Model Softw* 34, 5e18.
42. Chantry, M. et al. (2021) Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. *Philos Trans A* 379, 20200083.
43. Asher, M.J. et al. (2015) A review of surrogate models and their application to groundwater modeling. *Water Resour Res* 51, 5957–5973.
44. Mooij, W.M. et al. (2010) Challenges and opportunities for integrating lake ecosystem modelling approaches. *Aquat Ecol* 44, 633-667.
45. De Vico Fallani, F. et al. (2014) Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philos Trans R Soc Lond B Biol Sci* 369 (1653).

- 376 46. Cao, B. et al. (2020) The ERA5-Land soil temperature bias in permafrost regions.
377 *Cryosphere* 14, 2581–2595.
- 378 47. Maraun, D. et al. (2017) Towards process-informed bias correction of climate change
379 simulations. *Nat Clim Change* 7, 764-773.
- 380 48. Maraun, D. and Widmann, M. (2018) Statistical downscaling and bias correction in
381 climate research, Cambridge University Press.
- 382 49. Gordon, J. et al., Convolutional Conditional Neural Processes *ICLR* New Orleans, 2019.
- 383 50. Li, Y. et al. (2019) A Survey of Multi-View Representation Learning. *IEEE Trans Knowl*
384 *Data Eng* 31, 1863–1883.
- 385 51. Cambronero, C.M. et al. (2018) Predictability of the impact of multiple stressors on the
386 keystone species *Daphnia* *Sci Rep* 8, 17572.

Box 1

Artificial Intelligence (AI) and eDNA metabarcoding on lake sediment: a case study

AI approaches that use **explainable network models** combined with **multi-view learning** [50] allow the simultaneous interrogation of different data matrices, to learn what components co-vary within a matrix (e.g. environmental variables), and among matrices (e.g. environmental variables and Molecular Operational Taxonomic Units or MOTUs) (Fig. I). These approaches are often used in a systems biology context in which e.g. protein-protein (within matrix) or protein-gene (between matrices) interaction are investigated. We use the AI approach on a pilot study to establish correlations between environmental factors and biodiversity, measured with eDNA metabarcoding. For this pilot study, eDNA data were obtained from the sedimentary archive of a watershed with a well-known history of human-driven environmental change (Lake Ring, Denmark; 55°57'51.83"N, 9°35'46.87"E) [51]. The history of Lake Ring can be separated into four main phases: semi-pristine (PR; <1950); eutrophication (E; 1960-1970); pesticides (P; 1980-1990); and partial recovery (R; >1999). The beta diversity (community diversity between each pair of sediment layers) from the dated sedimentary archive showed that whereas the prokaryotic community was shaped by the redox state of the sediment (Fig. IIA), the eukaryotic freshwater community composition changed with the major lake phases (Fig. IIB). Applying the explainable multi-view learning algorithm combined with matrix-on-matrix regression (Fig. I) we identified the top 10 herbicides with significant adverse effects on primary producers, specifically identifying green algae as target MOTUs (Fig. IIB). A top-down Pearson correlation analysis (vector-on-vector regression) validated this approach.

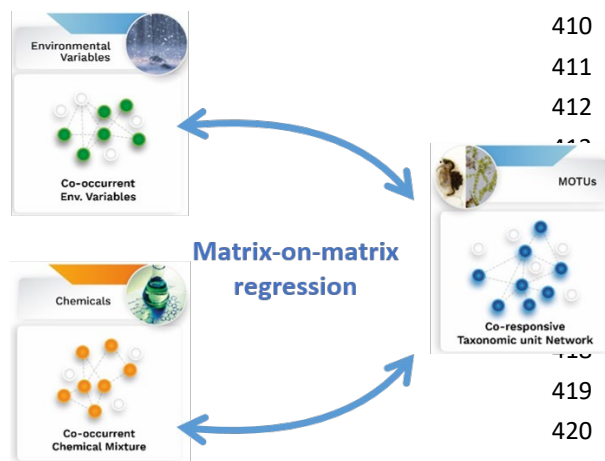


Figure I. Illustration of multi-view

learning combined with network analysis.

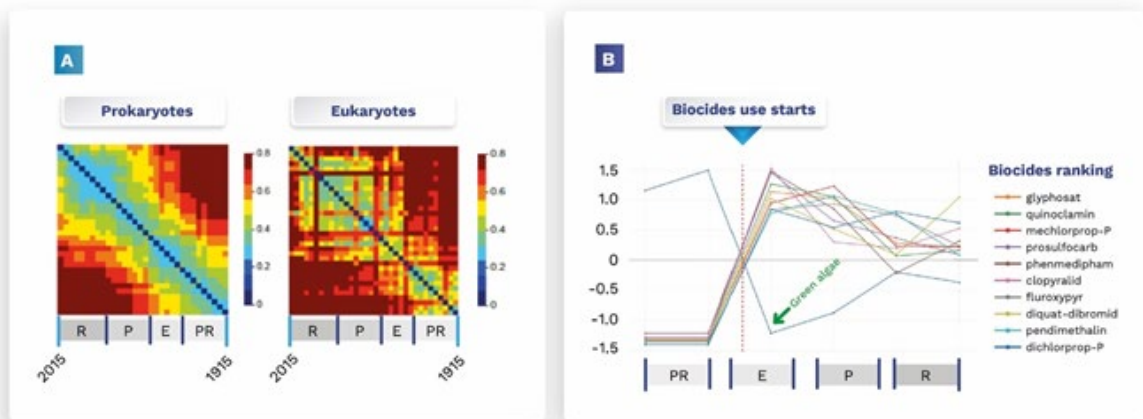
Combining multi-view learning and network analysis, symmetric matrix-on-matrix regressions are obtained. The matrices represented in Figure I are Molecular Operational Taxonomic Units (MOTUs); Environmental Variables (e.g., temperature); and chemical pollutants (chemicals). Co-varying elements within a matrix as well as co-varying elements between matrices are identified. For example each node in the

MOTUs network is a molecular taxonomic unit. Both co-variation in relative abundance of different MOTUs and their co-variation with environmental factors and chemical pollutants is identified using AI.

Figure II. Biodiversity changes through time and correlations with chemical pollution.

A) Eukaryote and prokaryote beta diversity through time (between each pair of layers) quantified with metabarcoding applied to eDNA extracted from lake sediment. Similarity in the composition of each sample (sediment layer) is measured through heat maps. Similarity decreases from blue to red. PR: pristine phase; E: eutrophication phase; P: pesticides phase; R: recovery phase; B) Inverse correlation between herbicides and primary producers (green

algae) established using the explainable multi-view learning algorithm combined with matrix-on-matrix regression shown in Fig. I.



Glossary

eDNA or environmental DNA is DNA extracted from environmental matrices, such as soil, sediment, water, ice and aerosol without any obvious signs of biological source material.

Metabarcoding is high throughput sequencing of PCR-amplified taxonomic marker genes.

MOTU or Molecular Operational Taxonomic Units. Units of taxonomic diversity that, differently from Operational Taxonomic Units, do not necessarily correspond to species, but they are treated as such when measuring diversity. Clustering within MOTU is based on similarity of DNA sequences.

Suspect and non-target screening of chemical compounds are techniques to screen for chemical pollutants using mass spectrometry (MS). The suspect screening uses a library of catalogued data such as chromatographic retention times in lieu of reference standards. Non-target screening uses no pre-existing knowledge for comparison before analysis.

Mass Spectrometry (MS) is an analytical technique used to quantify concentrations of chemical compounds in different matrices by measuring the mass-to-charge ratio of the chemicals.

AI or Artificial Intelligence refers to the simulation of human intelligence in machines.

Machine learning or ML is a subset of artificial intelligence and focuses on the development of algorithms that can access data and use it to learn for themselves

Explainable network model (ENM) is a recent advancement in Machine Learning algorithms designed to identify which features in the data are responsible for driving a certain output, providing more trustable predictions. A typical application of ENM is for diagnosing breast cancer based on observed patterns on pathology slides.

Multi-view learning is an emerging direction in machine learning, also called deep learning, which learns from multiple networks (or views). Typical applications of multi-view learning include systems biology where functional links between e.g. gene networks and metabolite networks are established.

An Emulator is a hardware or software that permits programs written for one computer to be run on another computer. This enables to increase the efficiency of time-consuming simulations by parallelising resources. In climate science, emulators are used to evaluate the realism of the warming signal in different models on both global and regional scales, by comparing global trends and regional response parameters to observations.

Granger causality is a statistical test for determining whether one time series is useful in forecasting another.

Process-based models are models that characterize changes in a system's state as explicit functions of the events that drive those state changes (*sensu* [34])

Correlative ecological models are models based on environmental associations derived from analyses of geographic occurrences of species.

Figure 1. The Time Machine framework

The Time Machine Framework is shown to establish past correlations between biodiversity, ecosystem functions and environmental change, and generate the likely forecasts of ecosystem services under different pollution and climate scenarios. Sedimentary archives from watersheds are sampled at continental scale and across climatic regimes. Sampled and dated sedimentary archives are used to *fingerprint* biodiversity, chemical pollution and ecosystem functions. Other environmental factors (e.g. temperature) are collected from weather stations or retrieved from public databases. Associations between biodiversity, ecosystem functions and environmental factors are established with the use of AI combined with explainable network models. Associations are then validated in fabricated ecosystems that are perturbed to identify cause-effect relations between biotic and abiotic changes. Model forecasts that ‘learn’ from past dynamics and feedback are tested iteratively against real data and refined to predict the future of biodiversity, ecosystems services and their economic value.

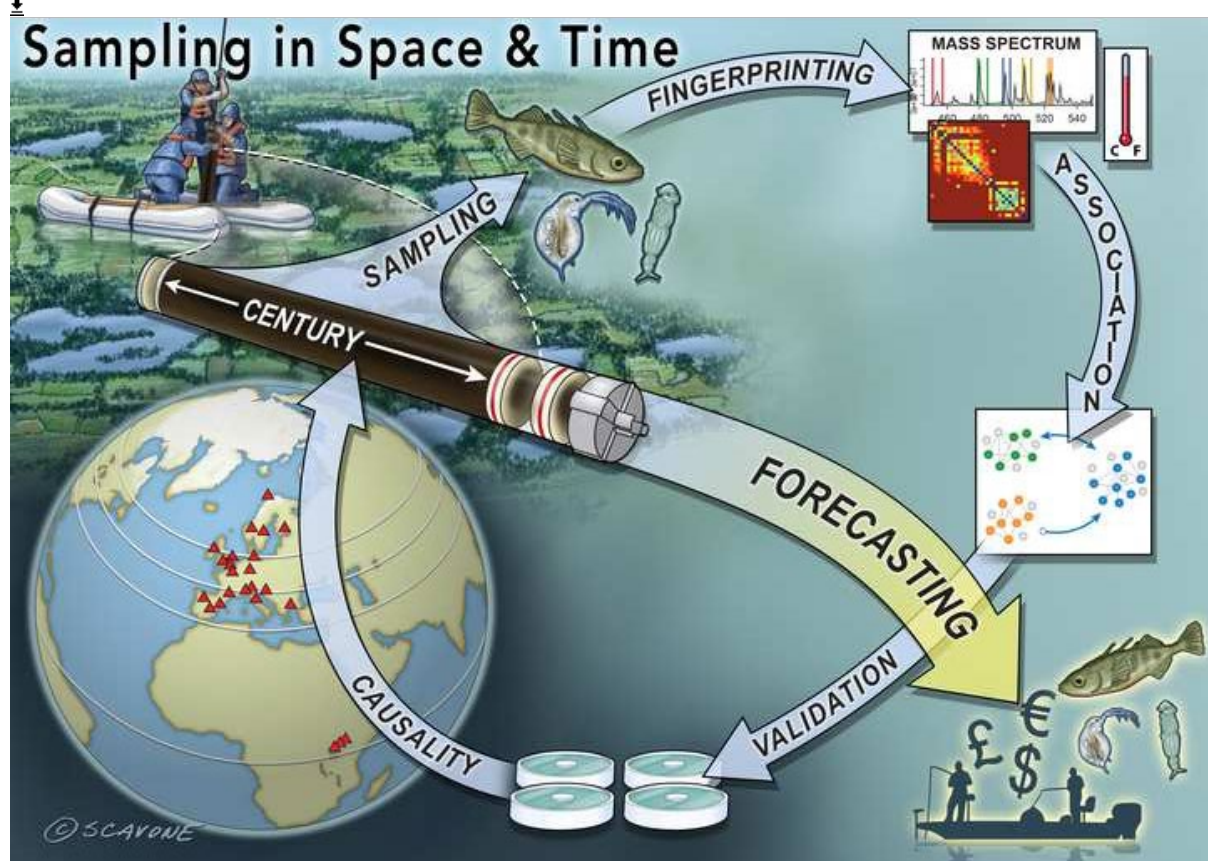
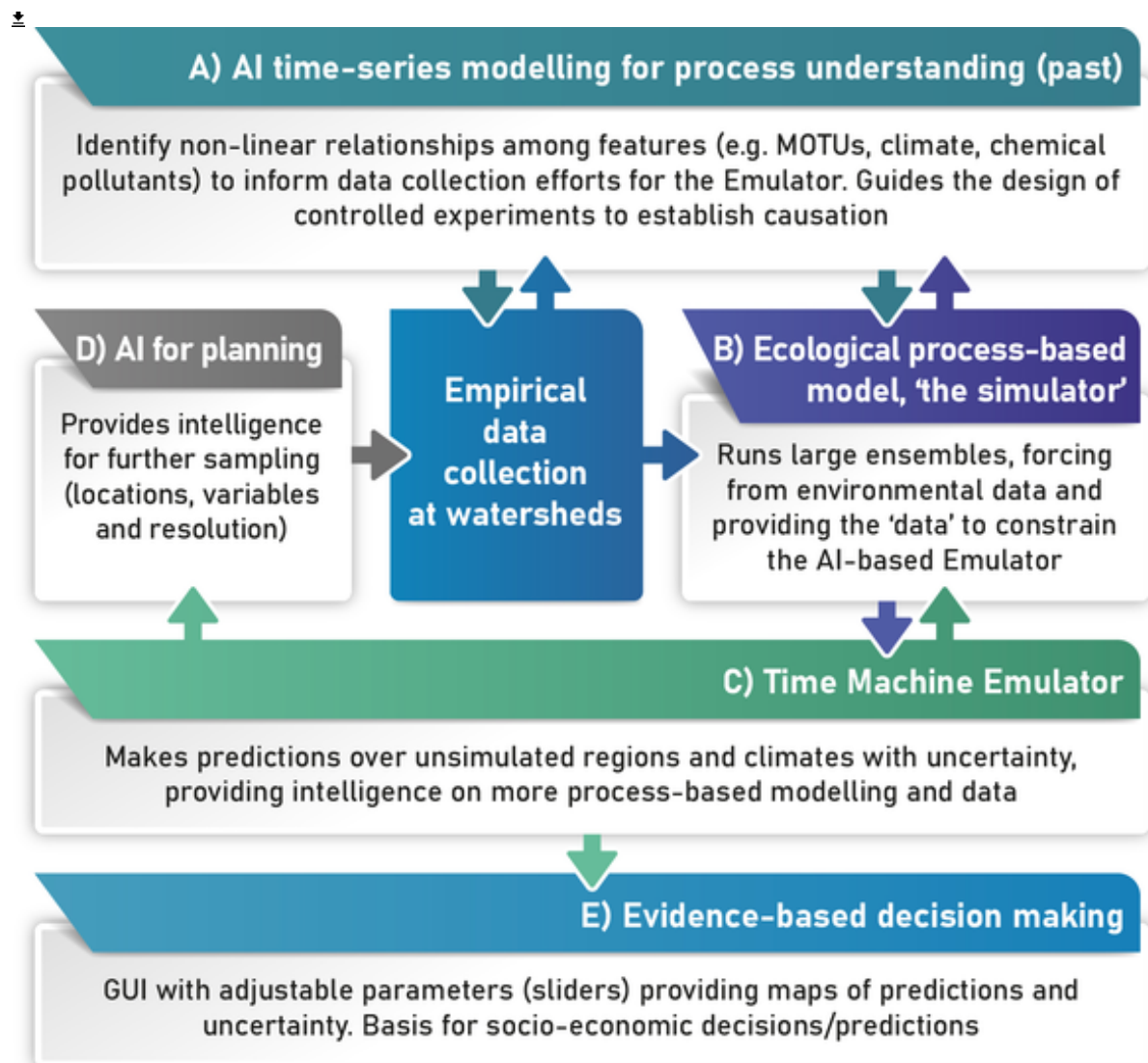


Figure 2. The Time Machine Emulator

Mapping out the various components of the data pipeline required for the Time Machine Emulator and the interoperability with data collection, process-based research, and the decision-making user interface. Arrows indicate the flow of data and information between the components. Empirical data (e.g. from watersheds) are analysed with an AI time series approach to establish non-linear relationships among multi-dimensional features (past) (A). The associations identified by the AI (A) and the process-based simulators (B) inform each other and are supported by empirical data. The Emulator (C) provides predictions and intelligence on regions that require further sampling to reduce prediction uncertainties (D)

and intelligence on more process-based modelling (B). A consensus model is obtained using a generalized additive framework that generates predictions through an AI-based emulator dashboard with a graphical user interface (GUI) (E).



505 Highlights

506 Chemical pollution and climate change are recognized as the two main causes of Earth's
507 ecosystem services deterioration and overuse, linked to the loss of biodiversity. Yet,
508 preventive interventions that mitigate this loss and preserve natural resources are inadequate
509 because the dynamics leading to biodiversity loss are context-dependent outcomes from
510 processes operating over many years.

511 We propose a framework that uses sedimentary archives from watersheds to establish causal
512 links between abiotic change and systemic loss of biodiversity, ecosystem functions and
513 services.

514 Just like a time machine, we go back in time and reconstruct a library of biological, chemical,
515 environmental and functional changes at a yearly resolution, enabling the understanding of
516 the spatiotemporal impacts of abiotic change on ecosystems and their services.

517 Outstanding questions

518 Is there a bias in the recovery of eDNA from fossil or remnant DNA in aged sediment and in
519 warm climates?

520 While we found stable composition of primary producers across 100 years, a bias might still
521 exist in sedimentary archives from warmer climates where mineralization of fossil remains is
522 influenced by higher temperature and humidity. For these archives shorter time series may be
523 used.

524 How to deal with high levels of uncertainty in the Emulator projections?

525 While the Emulator makes predictions over non-simulated regions and climates, guiding data
526 collection efforts, it carries a level of uncertainty. This has to be accounted for in decision-
527 making.

528 How does the Time Machine Emulator forecast the potential economic value of ecosystems?

529 Monetary and non-monetary estimates of values associated with ecosystem services can be
530 generated with the current monetary valuation strategies and the predicted changes in
531 ecosystem functions, uncertainty and risk provided by the Emulator. However, an outstanding
532 challenge for decision-making is incorporating how values held by society for different
533 services (e.g. carbon sequestration, food production, biodiversity, disease regulation) might
534 change through time.

535