

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/145501/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Eastwood, Niamh, Stubbings, William A., Abou-Elwafa Abdallah, Mohamed A., Durance, Isabelle , Paavola, Jouni, Dallimer, Martin, Pantel, Jelena H., Johnson, Samuel, Zhou, Jiarui, Hosking, J. Scott, Brown, James B., Ullah, Sami, Krause, Stephan, Hannah, David M., Crawford, Sarah E., Widmann, Martin and Orsini, Luisa 2022. The Time Machine framework: monitoring and prediction of biodiversity loss. *Trends in Ecology and Evolution* 37 (2) , pp. 138-146. 10.1016/j.tree.2021.09.008

Publishers page: <http://dx.doi.org/10.1016/j.tree.2021.09.008>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



1 **The Time Machine framework: monitoring and prediction of biodiversity loss**

2 Running title: Biodiversity and Ecosystem Services

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

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

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

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

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

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

16

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

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

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

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

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

25

26

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

28

29 **Abstract**

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

40

41 **The challenge of sustained ecosystem services**

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

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

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

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

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

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

80

81 **The Time Machine framework**

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

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

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

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

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

136 correlated with ten herbicides among the hundreds that were quantified in the sediment (Box
137 1).

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

156 Step 5 - *Forecasting* biodiversity and ecosystem services. **Ecological process-based**
157 **models**, informed by the associations identified by the AI in Step 3, are used to generate
158 predictions about projected future states of freshwater ecosystems (Fig. 1; *Forecasting*).
159 These predictive models are incorporated into simulations that project solutions for
160 ecosystem services (Fig. 2). Although it is likely that a range of complex processes interact to
161 determine how environmental drivers and associated biodiversity shifts influence ecosystem
162 functions, we propose to start with a simple model that illustrates the impacts of community
163 composition and structure, species interactions, and the covariation of these with
164 environmental factors on freshwater ecosystem functions. Process-based models with these
165 components include PCLake [31], a process-based model that links species composition,
166 environmental drivers (e.g. nutrient loading, temperature), and ecosystem responses (e.g.
167 water quality; [32]). Alternatively, ELCOM-CAEDYM, a coupled hydrodynamic and
168 biological model of phytoplankton dynamics and their impacts on water quality can be used
169 [33]. For reviews on **process-based models** see (e.g. [34]). Inputs for these models can be
170 provided from correlative approaches [35, 36], such as **correlative ecological models** (e.g.
171 [37]), or AI-based inferences such as described in Step 3.

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

179

180

181

182 The Time Machine Emulator

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

203

204 We propose a *Time Machine Emulator* (TME) to generate forecasts of biodiversity and
205 ecosystem functions. The TME ‘learns’ from past correlations, it is tested iteratively against
206 long-term empirical data (e.g. collected in Step 3) and refined to predict the future biotic,
207 abiotic and functional associations (Fig. 2). In our applications, it emulates ecological process-
208 based models (e.g. [32, 44]) to generate predictions across a range of historical and future
209 climates and ecosystem states, albeit in a more computationally scalable and efficient manner.
210 Empirical data collected from sedimentary archives, including biodiversity changes over time
211 and pollution, as well as climate data, are used to establish past correlations that inform
212 ecological process-based models (Fig. 2A). These models integrate historical contamination
213 trends that disrupt ecosystem functions mediated by changes in MOTUs, to identify risk
214 trajectories, with measured uncertainties (Fig. 2B). The TME is applied to produce the same
215 output as the process-based models without the time consuming and computationally expensive
216 running of the vast number of possible model setups for every possible climate and pollution
217 scenario (Fig. 2C). One of the biggest challenges for AI on medium to long-range timescales
218 is the lack of high-quality data for training model predictions. The TME uses empirical past
219 biodiversity-ecosystem dynamics (collected in Step 3), as well as temperature projections from
220 regional climate simulations, as ‘training sets’ to reduce uncertainty in forecasting the future
221 of ecosystem states under different climate scenarios (Fig. 2C). Predictions are also made over
222 non-simulated regions and climates with higher uncertainty, informing strategies for additional
223 sampling, in this way mitigating uncertainties when new data are introduced in the model (Fig.
224 2D). This directs sampling efforts improve the accuracy of forecasts. The output of the TME
225 are predictions for changes in ecosystem functions driven by localised and regional climate
226 change and mediated by changes in overall biotic/abiotic interactions, including on partial
227 training sets. The robustness of the TME predictions is assessed by removing data (i.e. locations
228 and conditions) from complete training sets and comparing predictions on whole and partial
229 datasets. The future associations predicted by the TME are validated using fabricated
230 ecosystems as described in Step 4. By identifying the MOTUs and ecosystem functions altered

231 by abiotic factors and/or climate variables, the TME provides probabilistic decision-support
232 information for ecosystem services management (Fig. 2E).
233

234 **Concluding remarks**

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

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

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

270

271 **Acknowledgments**

272 We thank William Scavone, Kestrel Studio, USA, for the artwork in Figure 1 and Chantal
273 Jackson for the artwork in Figure 2. This work was supported by The Alan Turing Institute
274 under the EPSRC grant EP/N510129/1, the Natural Environment Research Council
275 (NE/N005716/1) and a MIBTP-BBSRC PhD fellowship to NE (BB/M01116X/1).
276

277 **References**

- 278 1. Cardinale, B.J. et al. (2012) Biodiversity loss and its impact on humanity. *Nature* 486
279 (7401), 59-67.
- 280 2. Durance, I. et al. (2016) The challenges of linking ecosystem services to biodiversity. *Adv*
281 *Ecol Res* 54, 87-134.
- 282 3. Mace, G.M. et al. (2012) Biodiversity and ecosystem services: a multilayered relationship.
283 *Trends Ecol Evol* 27 (1), 19-26.
- 284 4. Spaak, J.W. et al. (2017) Shifts of community composition and population density
285 substantially affect ecosystem function despite invariant richness. *Ecol Lett* 20, 1315-1324.
- 286 5. Bonebrake, T.C. et al. (2019) Integrating Proximal and Horizon Threats to Biodiversity for
287 Conservation. *Trends Ecol Evol* 34 (9), 781-788.
- 288 6. Backhaus, T. et al. (2012) The impact of chemical pollution on biodiversity and ecosystem
289 services: the need for an improved understanding. *Integr Environ Assess Manag* 8 (4), 575-6.
- 290 7. Jackson, M.C. et al. (2016) Net effects of multiple stressors in freshwater ecosystems: a
291 meta-analysis. *Glob Change Biol* 22 (1), 180-189.
- 292 8. DeFries, R. and Nagendra, H. (2017) Ecosystem management as a wicked problem.
293 *Science* 356, 265-270
- 294 9. Jax, K. et al. (2018) Handling a messy world: Lessons learned when trying to make the
295 ecosystem services concept operational. *Ecosyst Serv* 29, 415-427.
- 296 10. Nogues-Bravo, D. et al. (2018) Cracking the Code of Biodiversity Responses to Past
297 Climate Change. *Trends Ecol Evol* 33 (10), 765-776.
- 298 11. Orsini, L. et al. (2013) The evolutionary time machine: using dormant propagules to
299 forecast how populations can adapt to changing environments. *Trends Ecol Evol* 28 (5), 274-
300 82.
- 301 12. Figueiredo, L. et al. (2019) Understanding extinction debts: spatio-temporal scales,
302 mechanisms and a roadmap for future research. *Ecography* 42, 1973-1990.
- 303 13. Barnosky, A.D. et al. (2016) Avoiding collapse: Grand challenges for science and society
304 to solve by 2050. *Elementa* 4, 000094.
- 305 14. Roux, D.J. et al. (2017) Transdisciplinary research for systemic change: who to learn
306 with, what to learn about and how to learn. *Sustain Sci* 12 (711-726).
- 307 15. Grêt-Regamey, A. et al. (2017) A tiered approach for ecosystem services mapping. In
308 Mapping Ecosystem Services (Burkhard, B. and Maes, J. eds), Pensoft Publishers.
- 309 16. Pinto, R. et al. (2014) Mainstreaming Sustainable Decision-making for Ecosystems:
310 Integrating Ecological and Socio-economic Targets within a Decision Support System.
311 *Environ Proc* 1, 7-19.
- 312 17. Ruckelshaus, M.H. et al. (2020) The IPBES Global Assessment: Pathways to Action.
313 *Trends Ecol Evol* 35 (5), 407-414.
- 314 18. Baert, J.M. et al. (2016) Per capita interactions and stress tolerance drive stress-induced
315 changes in biodiversity effects on ecosystem functions. *Nat Commun* 7, 12486.
- 316 19. Shoemaker, L.G. et al. (2020) Integrating the underlying structure of stochasticity into
317 community ecology. *Ecology* 101, e02922.
- 318 20. Vellend, M. et al. (2014) Assessing the relative importance of neutral stochasticity in
319 ecological communities. *Oikos* 123, 1420-1430.
- 320 21. Wang, Y. et al. (2018) Temporal effects of disturbance on community composition in
321 simulated stage-structured plant communities. *Ecol Evol* 8 (1), 120-127.
- 322 22. Cristescu, M.E. and Hebert, P.D.N. (2018) Uses and Misuses of Environmental DNA in
323 Biodiversity Science and Conservation. *Annu Rev Ecol Evol S* 49, 209-230.
- 324 23. Balint, M. et al. (2018) Environmental DNA Time Series in Ecology. *Trends Ecol Evol*
325 33 (12), 945-957.

- 326 24. Tse, T.J. et al. (2018) Combining High-Throughput Sequencing of sedaDNA and
327 Traditional Paleolimnological Techniques To Infer Historical Trends in Cyanobacterial
328 Communities. *Environ Sci Technol* 52 (12), 6842-6853.
- 329 25. Ruppert, K.M. et al. (2019) Past, present, and future perspectives of environmental DNA
330 (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of
331 global eDNA. *GECCO* 17, e00547.
- 332 26. Garlapati, D. et al. (2019) A review on the applications and recent advances in
333 environmental DNA (eDNA) metagenomics. *Rev Environ Sci Biotechnol* 18, 389–411.
- 334 27. Gillson, L. and Marchant, R. (2014) From myopia to clarity: sharpening the focus of
335 ecosystem management through the lens of palaeoecology. *Trends Ecol Evol* 29 (6), 317-25.
- 336 28. Abdallah, M. et al. (2019) A single run, rapid polarity switching method for
337 determination of 30 pharmaceuticals and personal care products in waste water using Q-
338 Exactive Orbitrap high resolution accurate mass spectrometry. *J Chromatogr A* 1588, 68-76.
- 339 29. Hirai, J. et al. (2017) Cryptic Zooplankton Diversity Revealed by a Metagenetic
340 Approach to Monitoring Metazoan Communities in the Coastal Waters of the Okhotsk Sea,
341 Northeastern Hokkaido. *Front Mar Sci* 4.
- 342 30. Tenenhaus, A. et al. (2014) Variable selection for generalized canonical correlation
343 analysis. *Biostatistics* 15 (3), 569-83.
- 344 31. Janssen, A.B.G. et al. (2019) PCLake+: A process-based ecological model to assess the
345 trophic state of stratified and non-stratified freshwater lakes worldwide. *Ecol Model* 396, 23-
346 32.
- 347 32. Janse, J.H. et al. (2010) Estimating the critical phosphorus loading of shallow lakes with
348 the ecosystem model PCLake: sensitivity, calibration and uncertainty. *Ecol Model* 221, 654-
349 665.
- 350 33. Robson, B.J. and Hamilton, D.P. (2004) Three-dimensional modelling of a *Microcystis*
351 bloom event in the Swan River estuary, Western Australia. *Ecol Model* 174, 203-222.
- 352 34. Connolly, S.R. et al. (2017) Process, Mechanism, and Modeling in Macroecology. *Trends*
353 *Ecol Evol* 32 (11), 835-844.
- 354 35. Brophy, C. et al. (2017) Biodiversity and ecosystem function: making sense of numerous
355 species interactions in multi-species communities. *Ecology* 98 (7), 1771-1778.
- 356 36. Ovaskainen, O. et al. (2017) How are species interactions structured in species-rich
357 communities? A new method for analysing time-series data. *P Roy Soc B* 284, 20170768.
- 358 37. Peterson, A.T. et al. (2015) Mechanistic and Correlative Models of Ecological Niches.
359 *Eur J Ecol* 1, 28-38.
- 360 38. Survey, C., The continuous plankton recorder survey, in: Association, M.B. (Ed.) Marine
361 Biological Association, Plymouth, UK, 2020.
- 362 39. Fordham, D.A. et al. (2020) Using paleo-archives to safeguard biodiversity under climate
363 change. *Science* 369, eabc5654.
- 364 40. Leeds, W.B. et al. (2014) Emulator-assisted reduced-rank ecological data assimilation for
365 nonlinear multivariate dynamical spatio-temporal processes. *Stat Methodol* 17, 126-138.
- 366 41. Castelletti, A. et al. (2012) A general framework for Dynamic Emulation Modelling in
367 environmental problems. *Environ Model Softw* 34 5e18.
- 368 42. Chantry, M. et al. (2021) Opportunities and challenges for machine learning in weather
369 and climate modelling: hard, medium and soft AI. *Philos Trans A* 379, 20200083.
- 370 43. Asher, M.J. et al. (2015) A review of surrogate models and their application to
371 groundwater modeling. *Water Resour Res* 51, 5957–5973.
- 372 44. Mooij, W.M. et al. (2010) Challenges and opportunities for integrating lake ecosystem
373 modelling approaches. *Aquat Ecol* 44, 633-667.
- 374 45. De Vico Fallani, F. et al. (2014) Graph analysis of functional brain networks: practical
375 issues in translational neuroscience. *Philos Trans R Soc Lond B Biol Sci* 369 (1653).

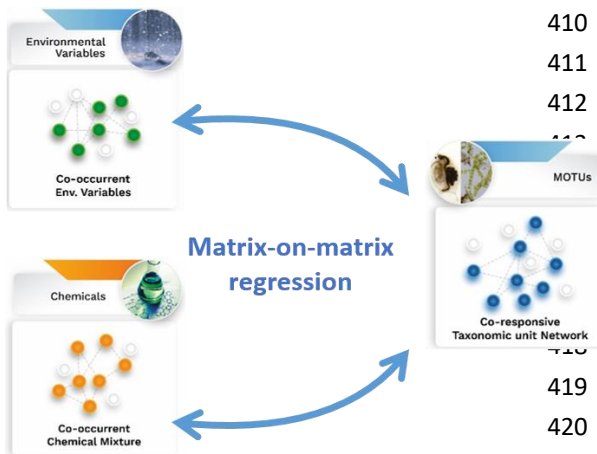
- 376 46. Cao, B. et al. (2020) The ERA5-Land soil temperature bias in permafrost regions.
377 *Cryosphere* 14, 2581–2595.
- 378 47. Maraun, D. et al. (2017) Towards process-informed bias correction of climate change
379 simulations. *Nat Clim Change* 7, 764-773.
- 380 48. Maraun, D. and Widmann, M. (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.

387

388 **Box 1**

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

390 AI approaches that use **explainable network models** combined with **multi-view learning**
391 [50] allow the simultaneous interrogation of different data matrices, to learn what
392 components co-vary within a matrix (e.g. environmental variables), and among matrices (e.g.
393 environmental variables and Molecular Operational Taxonomic Units or MOTUs) (Fig. I).
394 These approaches are often used in a systems biology context in which e.g. protein-protein
395 (within matrix) or protein-gene (between matrices) interaction are investigated. We use the
396 AI approach on a pilot study to establish correlations between environmental factors and
397 biodiversity, measured with eDNA metabarcoding. For this pilot study, eDNA data were
398 obtained from the sedimentary archive of a watershed with a well-known history of human-
399 driven environmental change (Lake Ring, Denmark; 55°57'51.83"N, 9°35'46.87"E) [51]. The
400 history of Lake Ring can be separated into four main phases: semi-pristine (PR; <1950);
401 eutrophication (E; 1960-1970); pesticides (P; 1980-1990); and partial recovery (R; >1999).
402 The beta diversity (community diversity between each pair of sediment layers) from the dated
403 sedimentary archive showed that whereas the prokaryotic community was shaped by the
404 redox state of the sediment (Fig. IIA), the eukaryotic freshwater community composition
405 changed with the major lake phases (Fig. IIB). Applying the explainable multi-view learning
406 algorithm combined with matrix-on-matrix regression (Fig. I) we identified the top 10
407 herbicides with significant adverse effects on primary producers, specifically identifying
408 green algae as target MOTUs (Fig. IIB). A top-down Pearson correlation analysis (vector-on-
409 vector regression) validated this approach.



410 **Figure I. Illustration of multi-view**
411 **learning combined with network analysis.**
412 Combining multi-view learning and network
413 analysis, symmetric matrix-on-matrix
414 regressions are obtained. The matrices
415 represented in Figure I are Molecular
416 Operational Taxonomic Units (MOTUs);
417 Environmental Variables (e.g., temperature);
418 and chemical pollutants (chemicals). Co-
419 varying elements within a matrix as well as
420 co-varying elements between matrices are
421 identified. For example each node in the

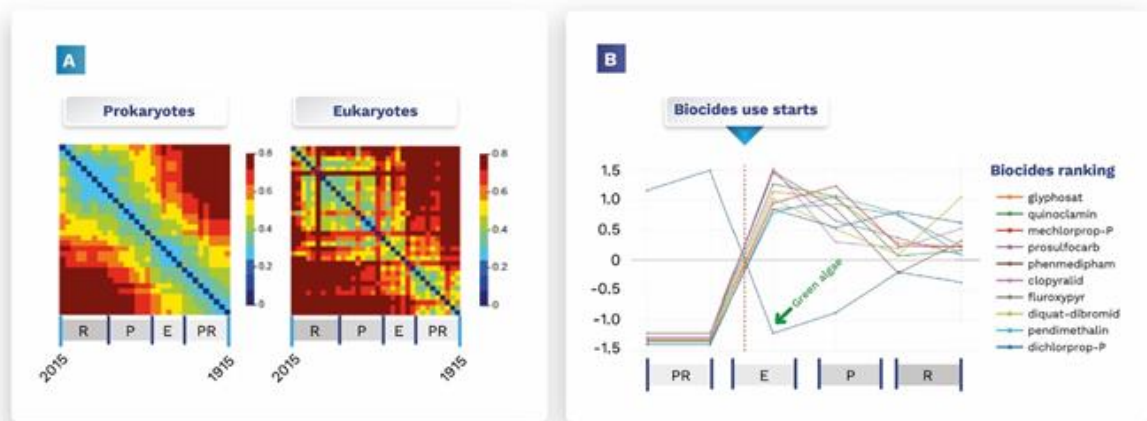
422 MOTUs network is a molecular taxonomic unit. Both co-variation in relative abundance of
423 different MOTUs and their co-variation with environmental factors and chemical pollutants is
424 identified using AI.

425

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

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

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



434

435

436 **Glossary**

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

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

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

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

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

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

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

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

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

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

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

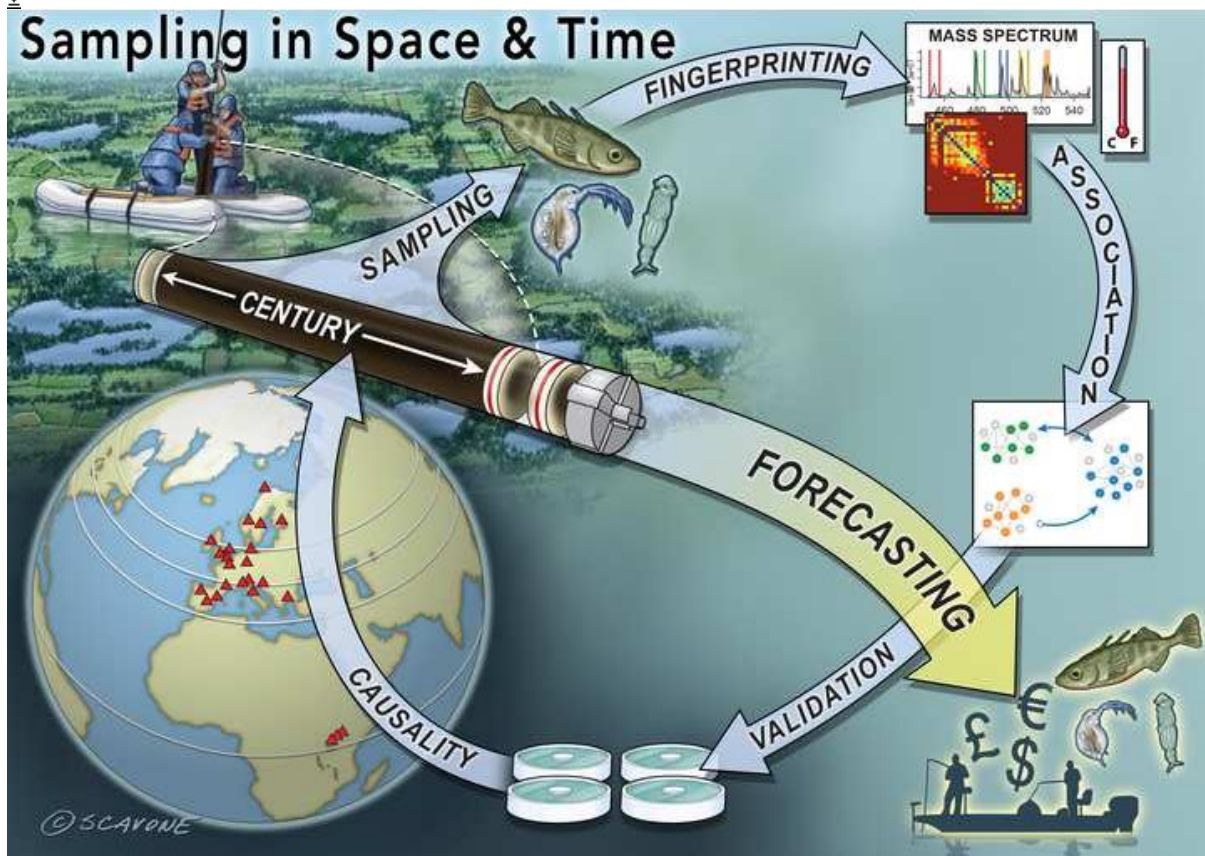
469 **Process-based models** are models that characterize changes in a system's state as explicit
470 functions of the events that drive those state changes (*sensu* [34])

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

473 **Figure 1. The Time Machine framework**

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

487



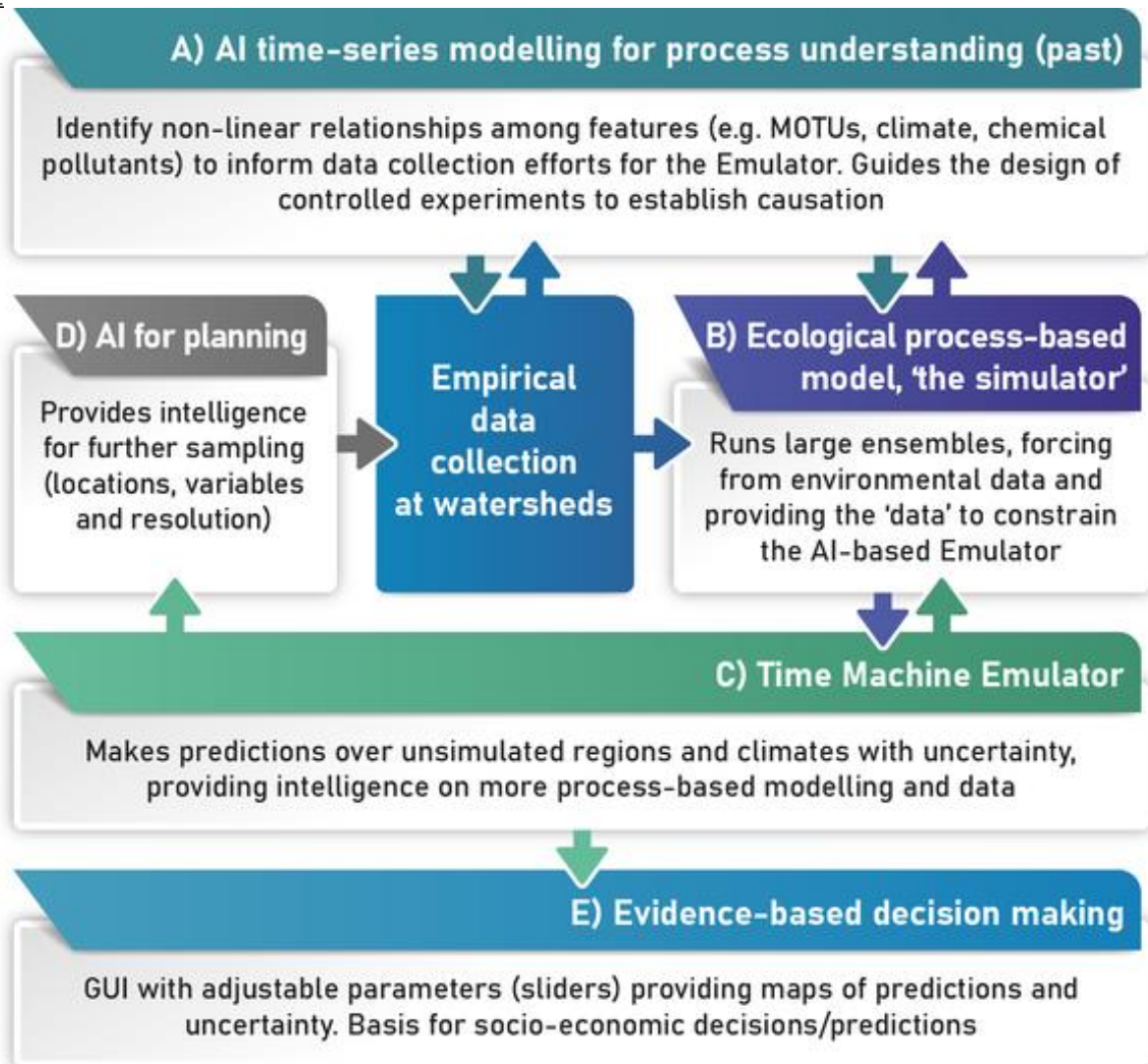
488
489

490 **Figure 2. The Time Machine Emulator**

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

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

502 *



503
 504

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