ARPHA Preprints

PENSOFT

Project Report

Author-formatted document posted on 23/05/2024

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e128109

D5.3 Past-to-present EBV integrated datasets and status indicators using Water Framework Directive data streams for selected freshwater ecosystem types in the Habitats Directive

Anne Lyche Solheim, Benoît O. L. Demars, Jannicke Moe, Sophie Mentzel, Jan-Erik Thrane, Jes Rasmussen, Benno Dillinger, Seppo Hellsten, Minna Kuoppala, Juha Riihimäki, Richard Johnson, Jens Fölster, Marit Mjelde

ARPHA Preprints *Author-formatted document posted on 23/05/2024*. DOI: https://doi.org/10.3897/arphapreprints.e128109

EUROPABON

Past-to-present EBV integrated datasets and status indicator using Water Framework Directive data streams for selected freshwater ecosystem types in the Habitats Directive

03/04/24

Lead beneficiary:

NORSK INSTITUT FOR VANNFORSKNING (NIVA)

This project receives funding from the European Commission's Horizon 2020 research and innovation programme, under Grant Agreement n.101003553
Prepared under contract from the European Commission
Grant agreement No. 101003553
EU Horizon 2020 Coordination and Support Action

Project acronym:	**EuropaBON**
Project full title:	**EUROPA BIODIVERSITY OBSERVATION NETWORK: INTEGRATING DATA STREAMS TO SUPPORT POLICY**
Start of the project:	01.12.2020
Duration:	36 months
Project coordinator:	Prof. Henrique Pereira Martin-Luther Universität Halle-Wittenberg (MLU) www.europabon.org
Type:	Coordination and Support Action
Call:	The Sc5-33-2020 Call: "Monitoring ecosystems through innovation and technology"

Project ref. no.	101003553
Project title	EUROPA BIODIVERSITY OBSERVATION NETWORK: INTEGRATING DATA STREAMS TO SUPPORT POLICY

Deliverable title	WFD synergies
Deliverable number	D5.3
Contractual date of delivery	31.03.2024
Actual date of delivery	03.04.2024 (2nd and final draft)
Type of deliverable	Report
Dissemination level	XXX
Work package number	WP5
Institution leading work package	MLU/NIVA
Task number	T5.3
Institution leading task	NIVA
Author(s)	Anne Lyche Solheim, Benoît O.L. Demars, Marit Mjelde, Jannicke Moe, Sophie Mentzel, Jan-Erik Thrane, Jes Rasmussen, Benno Dillinger (all NIVA) External data providers for the macrophyte species EBV: Seppo Hellsten (SYKE), Minna Kuoppala (SYKE), Juha Riihimäki (SYKE), Richard Johnson (SLU), Jens Fölster (SLU), External data providers for the community structure EBVs for EQR-values for benthic invertebrates in rivers and phytoplankton in lakes: EIONET National focal points & research centres (see preface for acknowledgments)
EC project officer	Laura Palomo-Rios

Deliverable description	This report is presenting the dataflows and outputs of three freshwater EBVs
Keywords	Macrophytes species distribution, Benthic invertebrates community structure, Phytoplankton community structure, lakes, rivers, modelling

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

EUROPABON

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

Contents

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

EUROPABON

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Preface

This report provides a showcase of selected EBVs related to the EU Water Framework Directive (WFD) and possible synergies with other EU policies. The report consists of two major parts: the species level EBV on macrophyte species in Nordic lakes and the community structure EBVs for phytoplankton in lakes and benthic invertebrates in rivers based on their deviation from natural conditions measured by Ecological Quality Ratios (EQR-values). The macrophytes species level part is produced by Benoit Demars at NIVA supported by a number of co-authors, including data providers (see frontpage), while the community structure EBVs based on EQR-values are produced by Jannicke Moe, Sophie Mentzel, Jan-Erik Thrane and Anne Lyche Solheim at NIVA. All EIONET data providers of EQR-values reported to WISE-2 are kindly acknowledged.

Executive summary

Introduction

The WFD monitoring has brought orders of magnitude more biological data from rivers and lakes than freshwater monitoring before the WFD. The implementation of the WFD monitoring programmes has largely increased the knowledge on major groups of freshwater biota, the so-called biological quality elements (BQEs) in rivers and lakes and how they respond to various human pressures.

There are several gaps in the WFD monitoring, in particular for very small water bodies, which are not identified as separate water bodies. These are rivers/streams with < 10 km^2 catchment area and lakes/ponds with surface area < 0.5 km^2 (50 ha). There are also spatial gaps in terms of the percentage of water bodies monitored for the different BQEs. The most monitored BQEs are benthic invertebrates in rivers and phytoplankton in lakes, which are monitored in one third of all the water bodies, while lake macrophytes are only monitored in 10% of all lake water bodies. Some freshwater biota is not monitored at all under the WFD, e.g. amphibians and zooplankton. This does not mean that there is no monitoring at all for those groups, but only that they are not required to be monitored by the WFD.

The monitoring raw data are normally stored in national or regional databases, but these are not reported to the European level. Some species data are reported to GBIF however. The raw data are used by the countries to calculate ecological quality ratios (EQR), which is the deviation from reference conditions in water bodies with undisturbed flora and fauna. The EQR values ranges from 0-1, where zero is a completely damaged ecosystem and one is a fully natural undisturbed ecosystem.

The status class data for each BQE are reported to the European level once every six years together with the updating of the river basin management plans (RBMPs). The EQR data are reported annually to EEA with the State-of-Environment (SoE) reporting and stored in the Water Information System for Europe, section on biological data (WISE-2). The EQR data can be used for trend analysis, as described below (chapter 5), but has geographical gaps, as not all countries report data to WISE-2.

The metadata from WISE-2 has been reported for each country to the EuropaBON metadatabase (EuropaBON WP3.1) and is further described in EuropaBON D3.1 Inventory of current European network for monitoring (Morán-Ordóñez et al. 2021). The workflows and bottlenecks are described in Morán-Ordóñez et al. (2023), while the gaps are further described in Santana et al. (2023).

Three of the groups of EBVs are particularly relevant to link to the WFD data:

1) Species distribution and abundance, incl. macrophytes, invertebrates, fish. These are providing the underlying data for the other freshwater EBVs.
2) Community composition, including ecological quality ratios (EQR) values for all biological quality elements (BQEs) required by the WFD.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

These are directly linked to the WFD obligations, as necessary for assessing ecological status class for each of the WFD biological quality elements according to the WFD Annex V requirements. The EQR data is reported by most European countries to EEA as part of the voluntary annual State-of-Environment data flow. The data are stored in the WISE-2 database. The metadata are also available in the EuropaBON database (https://monitoring.europabon.org/).
3) Ecosystem structure, including distribution of EUNIS inland water habitats are mostly comparable to the WFD broad surface water body types (Lyche Solheim et al., 2019).
The revised EUNIS inland water habitats include descriptions of characteristic, common and dominant taxa for each of the BQEs in each of the EUNIS habitats.
A fourth group of EBVs is Ecosystem function, which includes harmful algal blooms (i.e., cyanobacteria biomass), is reported by several European countries to EEA as part of the voluntary annual State-of-Environment data flow. The data are stored in the WISE-6 database.

The goals of this showcase for the freshwater EBVs are to describe and explain three of the freshwater EBVs in terms of datasets, dataflows and modelling:

1) Species distribution of macrophytes in lakes (https://github.com/EuropaBON/EBV-Descriptions/wiki/Freshwater-Species-distributions-of-freshwater-macrophytes)
2) Community structure EBVs for phytoplankton in lakes based on EQR values (https://github.com/EuropaBON/EBV-Descriptions/wiki/Freshwater-Community-composition-of-phytoplankton)
3) Community structure EBVs for benthic invertebrates in rivers based on EQR values (https://github.com/EuropaBON/EBV-Descriptions/wiki/Freshwater-Community-composition-of-benthic-invertebrates)

The following text summarizes the goals, datasets, approaches, outputs, lessons learned, as well as challenges and proposed solutions for these three EBVs, grouped into two main chapters:

a) The species level EBV for macrophytes in lakes and
b) The community structure EBVs for phytoplankton in lakes and benthic invertebrates in rivers.

Macrophyte species EBV

The goal has been to use raw data at species level together with environmental abiotic data from national WFD monitoring across the Nordic countries to develop species distribution models.

Altogether, raw data were compiled from 2059 lakes across Norway, Sweden and Finland. Altogether 97 species were registered. In this project, we focused on two species with different environmental requirements, but similar approach can be used for all those species in future projects. The most important environmental variables used in the models were alkalinity, colour (humic substances), nutrients (Total nitrogen and Total phosphorus), lake area and temperature (annual sum of degree days). These environmental variables were available for ca. half of all the lakes.

Various modelling approaches were tested to link the species to the environmental variables, e.g. general linear models (glm), generalized additive models (gam) and ensemble models.

The modelling was done for two macrophyte species, *Isoëtes lacustris* and *Potamogeton lucens*, and can be repeated for any of the species in the dataset. For each of the two species, the models estimated their habitat preferences (niche) based on their co-occurrence with alkalinity, water colour (humic substances), temperature (annual sum of degree days) and total phosphorus. The two species were found to occupy very different habitats: *Isoëtes lacustris* occurs mainly in low alkalinity, clear and relatively cool lakes with low/moderate phosphorus concentration, while *Potamogeton lucens* is found only in high alkalinity lakes, prefers warmer lakes, can tolerate both low and high levels of humic substances and is less sensitive to changes in phosphorus concentration. These habitat preferences

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

were used to model their probability of occurrence across in a much larger dataset with only environmental data from 3693 lakes across the three countries.

The models were also used to estimate the impacts of nutrient enrichment on the probability of occurrence for each of the two species in Fennoscandian lakes. These model results can therefore be used to estimate effects of nutrient reduction measures on the occurrence of single species, e.g. how would the species distribution be changed if lake phosphorus concentrations were halved following pollution abatement measures in agriculture or in better waste water treatment?

When such modelling have been done for all macrophyte species, they can also be used to estimate species richness for the whole macrophyte community across the whole geographic region. The stacking of predicted species distribution therefore opens the possibility to derive community and habitat status assessment, offering an alternative pathway to the production of ecological quality ratios (EQRs) in the implementation of the EU WFD.

The species distribution models (EBV cubes) can be used to characterise the health of key habitats (e.g. EU Habitats Directive, EUNIS habitats) from characteristic species and their behaviour under varying scenarios. Thereby these models can contribute to more confident and comparable assessments of conservation status for the Habitats Directive.

These species distribution models based on lake co-variates therefore offer a powerful approach for management, conservation and a range of environmental policies.

Community structure EBVs based on EQR-values for benthic invertebrates in rivers and phytoplankton in lakes.

The main goal has been to explain the rationale and existing data flow for these EBVs and show how their trends can be analysed and presented, also pointing out their benefits for water managers and other stakeholders responsible for the implementation of the Water Framework Directive and other policies, including the Habitats Directive and Nature Restoration Law.

These EBVs are based on Ecological Quality Ratio (EQR) data, measuring the current condition of different biological communities in rivers and lakes. The EQR-value is the deviation from reference conditions found in undisturbed biological communities and is given on a harmonised scale from 0 (largest deviation) to 1 (no deviation = reference conditions). EQR-values > 0.6 represent rivers and lakes which are in good or high status for a biological community, e.g. benthic invertebrates in rivers or phytoplankton in lakes. In those rivers and lakes, the biological communities mainly consist of species found in undisturbed conditions. In contrast, EQR-values < 0.6 represent rivers and lakes with moderate, poor or bad status, in which the biological communities are dominated by species that are more tolerant to disturbance by pollution or hydro-morphological habitat alteration.

The community structure (= species composition) is measured by national indices developed from knowledge on species occurrences and preferences along pollution gradients, e.g. nutrient gradients in different types of rivers and lakes. These national indices and their thresholds between the high & good, as well as good & moderate status classes have been intercalibrated across countries sharing common types within a European geographical region, e.g. Northern, Central, Eastern, Mediterranean. This intercalibration provides comparable EQR-data across countries.

The dataflow for the EQR-values started in 2008 after consultation with stakeholders in major European networks (EIONET and ECOSTAT). The ETC-Water working for EEA developed the guidance for data providers at national level and supported EEA in data handling and analysis. This support work was continued in the ETC-ICM and lastly in the ETC-BE.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

A series of data products have been developed showing the spatial distribution of status classes for each biological community and trends over time, including the amount of change over a certain period (2015-2021). The spatial distribution of status classes are shown as interactive maps allowing the user to select the area and year, as well as water category (rivers or lakes for example) and the biological community they want to see. The trend results show that the condition for benthic invertebrates in rivers and phytoplankton in lakes have significantly improved during this period in rivers and lakes that were in poor or bad status in 2015 (EQR-values < 0.4), most likely in response to pollution reduction measures taken in the years before 2015. However, they have still not achieved good status, but are approaching moderate status. Such an improvement would not be possible to show if only the WFD status class data were used, but can be highly motivating for stakeholders, in terms of showing positive effects of costly measures taken to reduce pollution.

There are major spatial gaps in the data flow due to some large countries not reporting EQR-values, e.g. France and Germany. However, the status class distribution across all the 16 countries reporting EQR-values is quite representative for all the countries reporting only status class data once every 6 years in the mandatory reporting required by the WFD. Temporal data gaps can be closed by interpolation up to 3 years, which is normal practice in other EEA indicators (e.g. the nutrients indicator). This practice is particularly important for biological communities that are not monitored annually, but only once every 2-3 years.

These community structure EBVs have recently been shown to correlate well with land-use variables, showing deterioration (lower EQR-values) with increasing proportion of agriculture and urban areas in the catchment of the rivers and lakes. Through regression tree analysis, the actual proportion of such areas that would decrease the EQR-values were quantified. This would in the next step facilitate modelling of EQR-values based on land-use data in areas without biological observations.

In conclusion, the community structure EBVs based on EQR-values have been demonstrated as powerful trend indicators, allowing early trend detection and quantification of change for each BQE in each single water body where EQR-values have been reported. The EBVs can therefore capture impacts of pressures and measures on the condition of freshwater biodiversity.

These EBVs can also be used as indicators for ecosystem structure and function within the conservation status assessments of the Habitats Directive for all types of water bodies that match the habitat types listed for inland waters in the HD Annex 1.

Motivating national data providers to report EQR-data in the coming years is essential to maintain and strengthen this data flow in the coming years. Visualisation of results and webinars with data providers must therefore be continued.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

1. Introduction

1.1. State of the art in WFD monitoring in Europe

1.1.1 The WFD monitoring system

The WFD monitoring has brought orders of magnitude more biological data from rivers and lakes than freshwater monitoring before the WFD, when most of the monitoring was mostly chemical. The WFD monitoring is a legal obligation described in Article 8 with details in Annex V. The emphasis is on biological quality elements (BQEs), although also the supporting physico-chemical and hydromorphological quality elements are included.

The WFD monitoring is divided into three major types:

- Surveillance monitoring
- Operational monitoring
- Investigative monitoring

The surveillance monitoring shall include all large water bodies and all the BQEs and supporting QEs. It shall be representative of all types of water bodies in rivers and lakes (as well as in transitional and coastal waters). The minimum frequency is quite low (< once every six years), as the aim is to capture wide-spread and long-term impacts on ecological and chemical status, e.g. climate change and long-range transboundary air pollution. Another aim is to monitor pristine water bodies to set the reference values for each BQE and supporting QE, which is the baseline for assessing ecological status.

The operational monitoring shall include a representative subset of water bodies having similar human pressures. This monitoring can be limited to the most sensitive BQE and supporting QEs that are relevant to capture the impact of those pressures on ecological status. The minimum frequency is higher than for surveillance monitoring and is given in Annex V, section 1.3.4: once every 6 months for phytoplankton, once every 3 years for the other BQEs, as well as once every 3 months for the supporting physico-chemical QEs (e.g. nutrients, oxygen), once every 6 years for river continuity and morphology and continuous or monthly for hydrology.

The investigative monitoring shall be done in water bodies which are failing the environmental objectives, but the cause of failure is uncertain, as well as to ascertain the magnitude and impacts of accidental pollution. There are no specifications of which BQEs and supporting QEs that should be included.

The implementation of the WFD monitoring programmes has largely increased the data and knowledge on BQEs in rivers and lakes and how they respond to various human pressures (e.g. Lyche Solheim et al. 2013, Birk et al. 2023).

1.1.2. Gaps

There are several gaps in the WFD monitoring, in particular for very small water bodies, which are not identified as separate water bodies. These are rivers/streams with < 10 km^2 catchment area and lakes/ponds with surface area < 0.5 km^2 (50 ha). There are also gaps in terms of the percentage of water bodies monitored for the different BQEs (Table 1). The most monitored BQEs are benthic invertebrates in rivers and phytoplankton in lakes, which are monitored in one third of all the water bodies. These data are used to assess ecological status class for each BQE. In addition, other methods like grouping and expert judgement are also used to assess ecological status class for each BQE for one third of the water bodies without monitoring data.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Table 1. Overview of the number and percentage of water bodies monitored in rivers and lakes based on information reported with the WFD 2nd river basin management plans: Workbook: WISE_SOW_SWB_QualityElement_geMonitoringResults (europa.eu)

BQE	Rivers	Lakes
Phytoplankton	3768 (3%)	7747 (30%)
Macrophytes	12192 (11%)	2541 (10%)
Phytobenthos (algae)	20280 (18%)	1294 (5%)
Benthic invertebrates	38120 (33%)	2622 (10%)
Fish	21893 (19%)	2560 (10%)

Moreover, some freshwater biota is not monitored at all under the WFD, e.g. amphibians and zooplankton. This does not mean that there is no monitoring at all for those groups, but only that they are not required to be monitored by the WFD.

The monitoring raw data are normally stored in national or regional databases, but these are not reported to the European level. Some species data are reported to GBIF however.

The raw data are used by the countries to calculate ecological quality ratios (EQR), which is the deviation from reference conditions in water bodies with undisturbed flora and fauna. The EQR values ranges from 0-1, where zero is a completely damaged ecosystem and one is a fully natural undisturbed ecosystem. The EQR is calculated for each BQE in each monitored water body, which are further used to assess the ecological status class for each BQE. For some of the non-monitored water bodies, the ecological status class is determined by other methods: grouping or expert judgement (see the link in Table 1). It is not clear whether the latter methods are based on modelling or simply on qualitative information available in maps (e.g. land-use, population density etc.). For most of the BQEs, more than half of all water bodies have unknown status class (see the link in Table 1, column "unpopulated"). The most monitored BQEs have unknown status class for 30-40 % of all the water bodies.

The status class data for each BQE are reported to the European level once every six years together with the updating of the river basin management plans (RBMPs). The EQR data are reported annually to EEA with the State-of-Environment (SoE) reporting and stored in the Water Information System for Europe, section on biological data (WISE-2). The EQR data can be used for trend analysis, as described below (chapter 5), but has geographical gaps, as not all countries report data to WISE-2.

The metadata from WISE-2 has been reported for each country to the EuropaBON metadatabase (EuropaBON_WP3.1) and is further described in EuropaBON D3.1 Inventory of current European network for monitoring (Morán-Ordóñez et al. 2021). The workflows and bottlenecks are described in Morán-Ordóñez et al. (2023), while the gaps are further described in Santana et al. (2023).

1.2. Showcase goals for freshwater EBVs

The concept of Essential Biodiversity Variables (EBVs) was introduced to structure biodiversity monitoring globally, and to harmonize and standardize biodiversity data from disparate sources to capture a minimum set of critical variables required to study, report and manage biodiversity change (Kissling et al., 2018). The list of EBVs has been used as an analytical framework to help bridge the gap between biodiversity data and policy reporting needs (Geijzendorffer et al., 2016).

Within EuropaBON, EBVs have been defined for terrestrial, freshwater and marine ecosystems by the research partners in close dialogue with stakeholders (Moersberger et al., 2022), and is further specified in another deliverable from the EuropaBON project (D4.1) (Junker et al., 2023). The links between EBVs and different policies are available here: EBVs-policies-160623.xlsx - Google Sheets

Three of the groups of EBVs are particularly relevant to link to the WFD data:

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

1) Species distribution and abundance, incl. macrophytes, invertebrates, fish. These are providing the underlying data for the other freshwater EBVs.

2) Community composition, including ecological quality ratios (EQR) values for all biological quality elements (BQEs) required by the WFD.

These are directly linked to the WFD obligations, as necessary for assessing ecological status class for each of the WFD biological quality elements according to the WFD Annex V requirements. The EQR data is reported by most European countries to EEA as part of the voluntary annual State-of-Environment data flow. The data are stored in the WISE-2 database. The metadata are also available in the EuropaBON database (https://monitoring.europabon.org/).

3) Ecosystem structure, including distribution of EUNIS inland water habitats are mostly comparable to the WFD broad surface water body types (Lyche Solheim et al., 2019).

The revised EUNIS inland water habitats include descriptions of characteristic, common and dominant taxa for each of the BQEs in each of the EUNIS habitats.

A fourth group of EBVs is Ecosystem function, which includes harmful algal blooms (i.e., cyanobacteria biomass), is reported by several European countries to EEA as part of the voluntary annual State-of-Environment data flow. The data are stored in the WISE-6 database.

The goals of this showcase for the freshwater EBVs are to describe and explain three of the freshwater EBVs in terms of datasets, dataflows and modelling:

1) Species distribution of macrophytes in lakes (https://github.com/EuropaBON/EBV-Descriptions/wiki/Freshwater-Species-distributions-of-freshwater-macrophytes)
2) Community structure EBVs for phytoplankton in lakes based on EQR values (https://github.com/EuropaBON/EBV-Descriptions/wiki/Freshwater-Community-composition-of-phytoplankton)
3) Community structure EBVs for benthic invertebrates in rivers based on EQR values (https://github.com/EuropaBON/EBV-Descriptions/wiki/Freshwater-Community-composition-of-benthic-invertebrates)

1.2.1. Goals for species distribution of macrophytes

The main goals are:

1) To illustrate how raw data from national monitoring, including monitoring for the Water Framework Directive (WFD, 2000/60/EC), may be mobilised to assess species status
2) To generate a conceptual workflow from raw data to status assessment using the WFD reference condition concept (Figure 1).
3) To illustrate the workflow with a compilation of lake data from Norway, Sweden and Finland on aquatic macrophytes and environmental variables (water quality, climate data).
4) To present species distribution models for two key species (Figure 2) with different prevalence representing important European habitat types (siliceous and calcareous lakes).
5) To fill spatial biodiversity data gaps using the northern European lake survey of water chemistry in over 5000 lakes.
6) To define reference conditions for total phosphorus (pressure) across Fennoscandian lakes, and project species distribution under current and reference conditions for species and lake status assessment, thereby creating a new typology free intercalibration tool for the WFD.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024*. DOI: https://doi.org/10.3897/arphapreprints.e128109

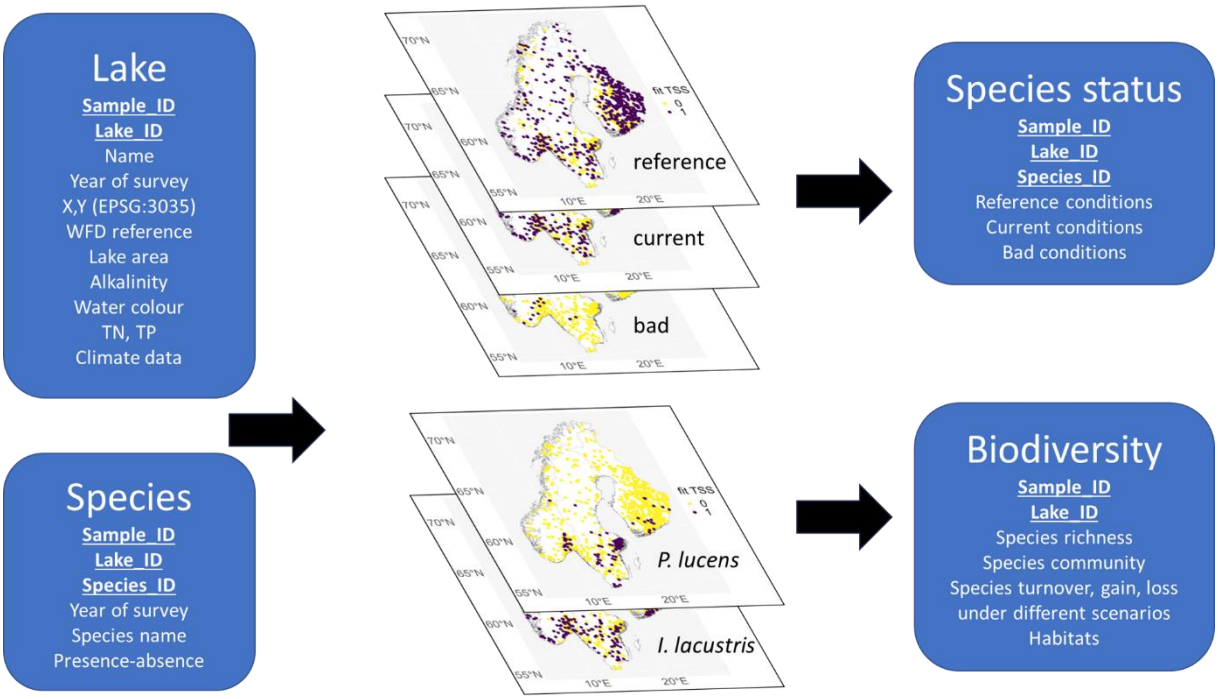


Figure 1. Workflow: EBV datasets -> spatial EBV cubes -> derived and modelled EBV data

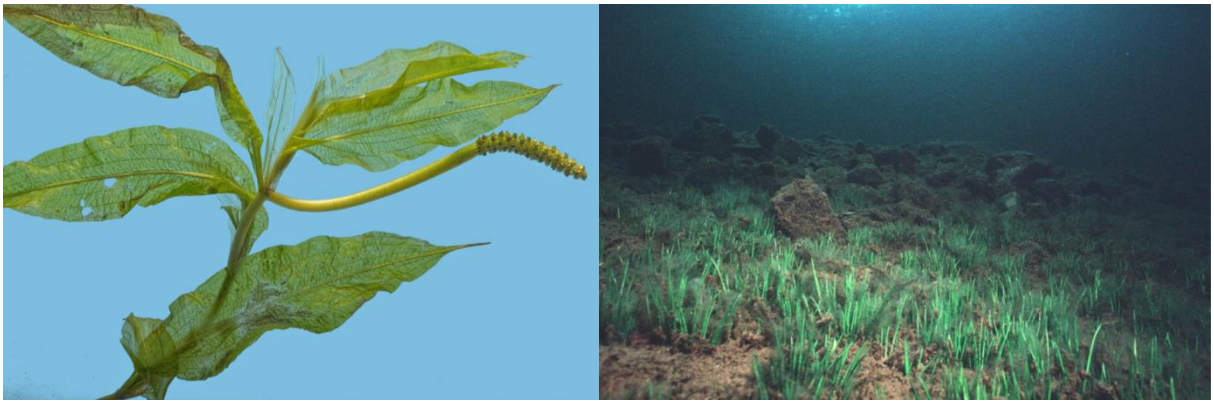


Figure 2. Potamogeton lucens L. (left) and Isoëtes lacustris L. (right), from Mjelde et al 2022 Photoflora (https://www.niva.no/en/featured-pages/photoflora-for-norwegian-aquatic-plants)

Goals for the Community structure EBVs based on EQR values

The goals for the EQR-based EBVs are:

1) To show the deviation of the current community structure from that found in reference conditions, in which the biological community is undisturbed by human pressures. Thereby the EBVs can show the current ecological condition of each of the biological quality elements required by the WFD.
2) To show how this deviation changes over time to answer the key question on whether the freshwater communities are becoming restored in response to restoration measures to reduce human pressures on freshwater habitats. These EBVs also aim to enable assessments of potential deterioration of the communities in response to climate change or increasing human pressures.
3) To obtain a spatially representative picture of the ecological condition of freshwater communities across Europe.
4) To motivate national stakeholders across Europe to improve their monitoring and reporting of the freshwater biological communities to allow reliable trends to be assessed at EU level.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

2. Showcase participatory design

2.1. Stakeholders' engagement process: Macrophyte species

2.1.1 Methodology

This task aimed to collect a representative set of data from three Nordic countries. The range of stakeholders was limited to those essential for the completion of the task since it is really time consuming to assemble data across countries (independently of the source). We directly contacted experts involved in the WFD in Norway (NIVA), Finland (SYKE) and Sweden (SLU) to identify geo-referenced raw data availability in water chemistry and species presence-absence in lakes. We also reviewed the lake macrophyte survey literature, including local or national reports.

2.1.2. Key stakeholders: users and data providers

All stakeholders were data providers and willing to share those data for this task of the EuropaBon project.

2.1.3. Stakeholder representativeness & gaps

The stakeholders were not representative of the EU. NIVA and SYKE are national research institutes. The role of SYKE is mainly to coordinate environmental monitoring in Finland. NIVA is a private company monitoring water chemistry and aquatic biodiversity through commissioned work and research projects with (or for) environmental agencies and industries. In Sweden, SLU (university) is responsible for designing the monitoring programme and compiling / making available data collected by consultancies. In Europe, freshwater data are also held in environmental and conservation agencies, natural history museums, botanical societies, and a wide variety of freshwater initiatives such as the freshwater platform (http://www.freshwaterplatform.eu/) or Freshwater BON (https://geobon.org/bons/thematic-bon/freshwater-bon/).

2.2. Key inputs from stakeholders: Macrophyte species

The access of large freshwater raw datasets, accumulated over decades, held by research organisations in Europe, have generally been restricted to single projects (e.g. EU projects REBECCA, WISER) or initiatives (e.g. WFD intercalibration exercise, JRC hold a copy of the data). All stakeholders provided important historical context and limitations of sampling methods. SYKE has invested in data portal (https://www.syke.fi/en-US/Research__development/Water/Monitoring). NIVA has been reporting a lot of its environmental and biodiversity data through national portals (https://vannmiljo.miljodirektoratet.no/ ; https://artsdatabanken.no/ reporting to GBIF). Sweden (SLU) have made a large amount of recent data available through an open access portal (https://miljodata.slu.se/mvm/). More data are being made public via LifeWatch ERIC initiative (https://www.lifewatch.eu/).

2.2. Stakeholders' engagement process: Community structure EBVs

The key stakeholders are the national focal point (NFPs) and national research centres (NRCs) included in EIONET (European Information and Observation Network; https://www.eionet.europa.eu/). The EIONET is established and run by the EEA to "gather and develop data, knowledge, and advice to policy makers about Europe's environment". The EIONET is also supported by the European Topic Centre on Biodiversity and Ecosystems (ETC BE), who is helping EEA to organise webinars with the NFPs and NRCs. (The predecessors of ETC-BE were named "ETC Waters" and "ETC/ICM" (Inland, Costal and Marine waters. For simplicity, this chain of consortia will be labelled "ETC" onwards).

The NFPs for freshwater are mainly the national environment agencies of Europe who are responsible for the WFD monitoring, assessments and reporting. The NRCs are national research institutes who are supporting the NFPs to design and do the monitoring and assessments related to ecological status

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

in rivers and lakes. There are specific representatives of the NFPs and NRCs who are engaged in the data generation and reporting of the community structure EBVs based on EQR values. These are called the WISE-2 stakeholders, who are invited to annual WISE-2 webinars for data providers, in which technical details of the reporting and also the use of the data are clarified and demonstrated. However, the people attending the WISE-2 webinars are often the data reporters (focusing on the technical work), rather than thematic experts or decision-makers, within each country. The WISE-2 data reporters can also contact EEA any time through the WISE-SoE helpdesk. A more systematic feedback on technical issues regarding the data deliveries is given to each country shortly after the annual reporting deadline (mid-January).

An additional group of key stakeholders are the WFD-CIS working group members of ECOSTAT, which are representatives responsible in their countries for the methodological development and intercalibration/harmonisation of indices/metrics for assessing ecological status for the BQEs and supporting QEs. These representatives are mostly representatives of the same organisations as in EIONET, however having more specific responsibilities for the methodological development.

The stakeholders are fully representative for Europe because all EEA member states are invited to the WISE-2 webinars and most countries attend, even those who do not report any data. There are no obvious gaps in the stakeholders.

2.3. Key inputs from stakeholders: Community structure EBVs

The key inputs from stakeholders started in an EIONET workshop at EEA premises back in 2006, in which Ireland was asking why EEA did not have any biological indicators, which appeared as a weakness in light of the WFD emphasis on BQEs. The EEA agreed to consider this in their follow-up actions after that workshop. The EEA therefore asked the ETC to produce a first draft of a guidance for community structure EBVs based on EQR values. This draft was consulted with the national stakeholders in EIONET and ECOSTAT to test their willingness to report such data to EEA. The key stakeholders supported this guidance, which was finalised in 2007 together with a data dictionary specifying all the details in the reporting.

A test reporting was started in 2008, in which many countries reported EQR values from a small proportion of water bodies. In 2009 the EEA/ETC did an evaluation of the data and had a dialogue with the countries to improve the representativeness of the reporting in terms of spatial coverage and ecological status class distribution. The first official data reporting started in 2010, and has been repeated almost every year since then with an increasing number of water bodies reported.

3. Policy targets

3.1. Reporting needs and data gaps

3.1.1 WFD reporting and data gaps

- **Mandatory WFD reporting and the state-of-water assessments**

The EU countries and EFTA countries Norway and Iceland are obliged to report their River Basin Management Plans (RBMPs) once every six years for each river basin district. The EFTA countries lag one six-year cycle behind the EU countries due to adoption of the WFD only in 2007. The RBMPs include information on pressures and status for all water bodies in rivers, lakes, transitional and coastal waters, as well as a programme of measures to reduce the pressures needed to achieve the objectives, which are good ecological and chemical status for all water bodies unless exemptions are well justified (e.g. extract of the 3rd RBMPs for Germany Water Framework Directive - The Status of German Waters in

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

2021 | Umweltbundesamt). The ecological status class for each BQE and supporting elements is not usually included in the RBMPs, although some countries may include such information.

The EC/EEA have also provided a data platform in WISE (Water Information System for Europe) (WISE-WFD), now under Reportnet 3 (europa.eu), allowing electronic data reporting following a very detailed and comprehensive reporting guidance (Water Framework Directive - River Basin Management Plans - 2022 Reporting (europa.eu)) developed by the WFD-CIS working group DIS (Data and Information Sharing). This includes also status class for each BQE and supporting QE. The electronic data reporting is not mandatory, but strongly encouraged. The links to various outputs are given in WISE-Freshwater (WISE Freshwater (europa.eu)), e.g. Surface water quality element status (europa.eu)

- **Voluntary reporting under the annual State-of-Environment (SoE) reporting to EEA:**

EEA agreed with EIONET in the 1990-ies to establish several voluntary dataflows based on annual data requests and reporting guidance documents to allow European overviews of the State-of-Environment. The data reported are/can be used to develop indicators showing trends that EEA uses in their SoE-reports once every five years (e.g. The European environment — state and outlook 2020 — European Environment Agency (europa.eu)).

The reported freshwater-biology SoE data based on EQR values is the basis for the community structure EBVs presented in chapter 5. These data are stored in Waterbase - Biology (europa.eu) and are also now directly available for download: EEA geospatial data catalogue (europa.eu). These data are well-suited for trend analysis due to the annual reporting on a continuous scale from 0-1, in contrast to the categorical status class data reported once every six years (WFD-data).

- **Remaining challenges in current reporting streams**

There are several challenges in the current reporting streams.

For the mandatory reporting, several countries are seriously delayed with the adoption and reporting of their 3rd RBMPs (or 2nd RBMPs for the EFTA countries) (
Figure 3).

More countries are delayed with the non-legally binding electronic data-reporting underlying their RBMPs (e.g. Bulgaria, Finland, Poland, Sweden), and a few countries (e.g. Finland) has rejected this data-reporting now in the 3rd cycle of RBMPs. This prevents a complete spatial coverage of current status class and change in status class between RBMP cycles.

The comparability of ecological status between countries is quite low due to different practices of combining BQEs and supporting QEs for assessing overall ecological status. This is why a recommendation was given to focus on comparing status class for only single BQEs between countries if the intention is to illustrate progress (Lyche Solheim et al. 2020).

Another challenge is that only a minority of all the water bodies are assessed based on monitoring data, more water bodies are assessed based on grouping and expert judgement (Table 1), but it is unclear which methods are used for the grouping and expert judgements. A major issue is also that a large proportion of water bodies have no assessment for the different BQEs.

The WFD data reporting once every 6 years is not fit for trend analysis.

For the voluntary SoE reporting, the main challenge for the biological EQR data reported to WISE2 is the lack of complete spatial coverage due to several large countries not reporting (e.g. France and Germany). It is therefore uncertain whether the current data are representative for Europe.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

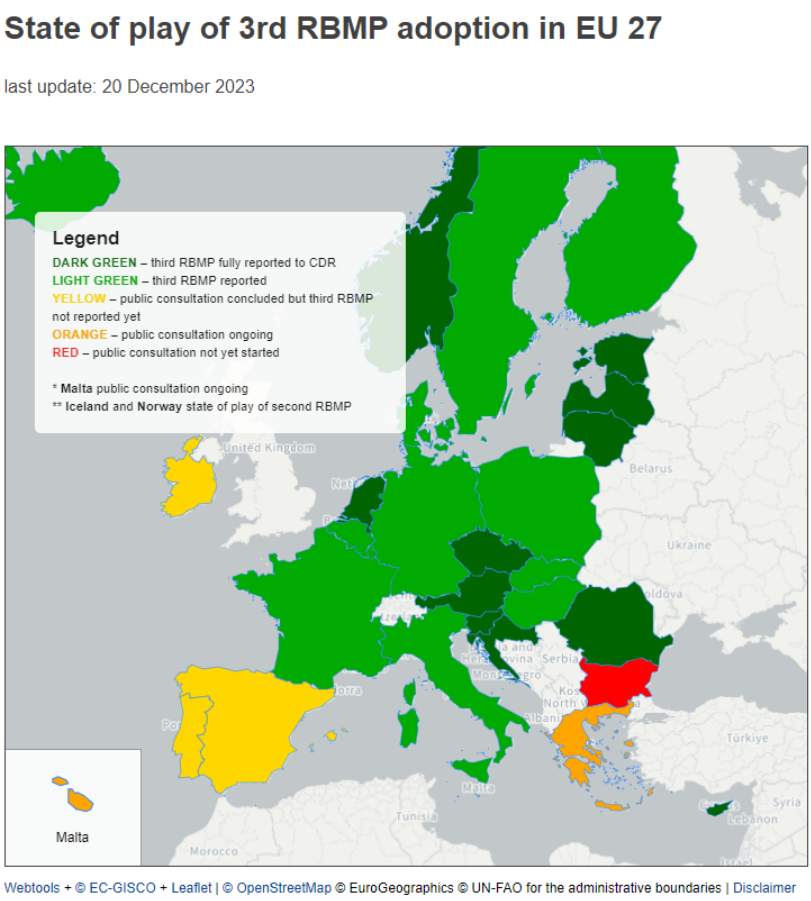


Figure 3. Latest update of the reporting of RBMPs (copied from https://environment.ec.europa.eu/topics/water/water-framework-directive_en)

- ***Expected contributions of the planned Essential Biodiversity Variable(s) for improving reporting and assessments.***

The main contributions of the EBVs described in chapters 4 and 5 below are several:

For the macrophyte species EBV, the main benefit is to demonstrate what we can do if the raw monitoring data were available. Having access to single species data from over 2000 lakes across the Fennoscandian countries, it is possible to model current species distribution from abiotic data (water quality and climate) and derive species distribution under varying scenarios such as ecosystem conditions (reference to bad status) or climate change.

For the community structure EBVs based on EQR values reported to WISE-2, the main benefit is to allow trends to be assessed for different groups of water bodies, depending on their initial EQR value at a chosen starting year. These trends can also later be linked to abiotic data and land-use data to explain the trends. Those EBVs are able to capture change also within ecological status classes, which is not possible with only categorical status class data. Thereby positive effects of costly measures taken to reduce pressures can be seen much sooner than what is possible with the WFD status class data from the RBMPs. Potentially negative impacts of climate change can also be seen, especially for water bodies in high or good status today, thereby triggering further research and actions to counteract such impacts.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

3.2. Cross-policy contributions

3.2.1. Links to mandatory reporting and state-of-nature assessments under the Habitats Directive.

Remaining challenges in current reporting streams

The mandatory reporting of conservation status for freshwater habitats under HD Annex 1 is composed of several metrics that are not harmonised across countries. This lack of harmonisation questions the comparability of results from different countries. Another challenge is that some of the freshwater habitats are very wide (e.g. Fennoscandian rivers or Water courses of plain to montane levels), while others are very narrow (e.g. Lakes of gypsum karst or Turloughs). The spatial resolution is also quite limited as the reporting unit is for each habitat aggregated to each biogeographic region in each country. The temporal resolution is also limited to once every six years, thereby not suited for trend analysis. Nevertheless, the conservation status includes trends as one of the metrics to be assessed. It is highly unclear how trends are assessed in the different countries.

HD conservation status also has to be assessed for all freshwater species listed in Annex 2, which includes only the species that need special protection. Other freshwater species are not considered, thus many commonly occurring species are simply ignored. Another challenge is that the monitoring of the Annex 2 species is quite limited due to constraints related to taxonomic expertise and funding of monitoring programmes, raising doubts about the realism in the assessment.

Relevance of WFD data for HD freshwater habitats and species, incl. revised EUNIS types:

In spite of the challenges listed above, there are several HD habitat types that can be linked to WFD types (Lyche Solheim et al. 2015, 2019), including the inland waters EUNIS types that are currently under revision by EEA supported by the ETC (draft report sent EEA in January 2024). Therefore, the ecological status of the WFD can be partially linked to the conservation status of the HD, for the component Structure and Function. Thus, there is a synergy between the WFD data and assessments and the HD data and assessments for comparable types of habitats.

Expected contributions of the planned Essential Biodiversity Variable(s) for improving reporting and assessments

The macrophytes species EBV can motivate countries to report more species data to GBIF or other potential raw data receiving bodies at EU level, by showing what is feasible to do if such data are available. The modelling can help assessing species distribution patterns and habitat preferences for the different species. The modelling for this EBV can also contribute to increase the spatial coverage of where single species are likely to occur and thereby the confidence in HD conservation status component on species habitat extension and range for macrophytes species listed in Annex II or IV.

The community structure EBVs based on EQR values can stimulate further data reporting of EQR values by demonstrating how trends can be detected long before a shift of status class. This increased data reporting will contribute to close the current spatial gaps in the data and thereby make the EBVs more representative for Europe as a whole. These EBVs can also support assessments of the State-of-Water by showing effects of measures taken to reduce pressures. This will be motivating for the countries to justify and enhance implementation of costly measures to further restore Europe's waters. Thereby, these EBVs will also support the implementation of the Nature Restoration Law, when/if this is endorsed by the European Parliament and Council.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

4. Essential Biodiversity Variables: Macrophyte species

4.1. EBV design characteristics

We assembled a database of 2079 georeferenced lakes for a predefined list of aquatic plant species and key water quality parameters from NIVA, SLU and SYKE. The spatial extent of the core dataset is Norway, Sweden and Finland (about 1,173,647 km^2). Data were collected over several decades and we defined a reference period 1980-2021 against which temporal changes may be assessed, through historical data or model scenarios. This long reference period allowed to have a representative and comprehensive spatial coverage on which species distribution models may be derived. Repeated lake surveys were not included in this database to avoid issues of temporal autocorrelations. In addition, a few lakes with the same geographical coordinates were also excluded. So, the final database was 2059 lakes.

4.2. Input biodiversity data

4.2.1 EBV-usable data sets: overall description of biodiversity data sources

Data on aquatic plants and some key environmental variables were already synthesised in previous EU projects such as REBECCA and WISER. While NIVA and JRC hold copies of the databases, access is restricted by member states and the data not readily usable. The most recent 2012 EU WISER database is now more than a decade old and the data on aquatic macrophytes used for the intercalibration of the WFD indices (EQRs) also had several limitations. Many lakes were surveyed a long time ago (problematic for spatial analyses) and did not have a full set of environmental variables, e.g. Lohammar 1938 did not have water colour, Lundh 1951 did not have TP or TN. Many lakes in WISER had no geographical coordinates (e.g. Finland, Kolada et al 2014). Water colour was only included as a categorical variable (clear lakes <30 mg Pt L-1, humic lakes 30-90 mg Pt L-1, poly-humic lakes > 90 mg Pt L-1). Elevation was introduced as a surrogate for climate data.

Here we compiled a novel synthesis on aquatic plants and water quality in lakes from a joint effort by NIVA, SYKE and SLU – based on a multitude of local, national and international projects. The details were compiled in Tables 2-4 below.

4.2.2. Raw data quality and assurance

Quality insurance varies between monitoring programs. Finland introduced ring tests for surveyors (Kuoppala et al 2008). Sweden had a period (2007-2014) when sampling effort per lake was not sufficient (Ecke 2014), so lakes with less than six transects were not selected and more recent surveys using an updated protocol (Ecke 2015) were selected whenever possible. The Norwegian data were compiled and checked by Marit Mjelde with 40 years' experience of aquatic plant recording in Norway. The water quality data followed established protocols and were from the same period (or year) of the plant survey.

Undetermined taxa (not identified to species) and hybrids were excluded. Bryophytes and helophytes were also excluded as generally not recorded in Norway. The taxonomy was cross-checked and authorities updated to produce a single list of species for Fennoscandia including tracheophytes (vascular plants) and charophytes. All sub-species and varieties were aggregated to species level: e.g. *Nymphaea alba* subp. *alba* and *N. alba* subsp. *candida*. Three pairs of species were aggregated as they cannot be confidently identified down to species without the reproductive parts: *Nitella opaca / flexilis*, *Utricularia vulgaris / australis, U. stygia / ochroleuca*. *Ranunculus peltatus* included *R. schwalhausenii* and *R. peltatus subsp. baudotii*. *Ranunculus aquatilis* in Sweden likely included *R. trichophyllus* (Hong 1991) and could not be harmonised.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

4.3. Data handling

The Norwegian dataset (Excel spreadsheet) was cleaned from artefacts '*', '?', '/', " ", '+/-', brackets, commas and spaces. A couple of river sections (more than lakes) and a few brackish water (salinity > 0.5 psu) lakes were removed. Records of continuous environmental variables reported as category were deleted (e.g. water colour, clear lakes: <30 mg Pt L^{-1}; humic lakes: >30 mg Pt L^{-1}). Few values below detection limits were deleted. A handful of obvious outliers when plotting the raw data were also deleted. Alkalinity units were all converted to mEq L^{-1}, and missing alkalinity values were imputed from calcium concentrations (mg Ca L^{-1}) with alkalinity = 0.0494*Ca +0.0128. Only one survey per lake was kept, preferentially the first survey or the survey with the most co-variates available. The dataset was reduced from 1378 lake-years to 952 lake surveys, including 137 reference lakes.

The Finnish dataset was made of two Excel spreadsheets: macrophyte species and associated lake data including average water chemistry. Multiple waterbodies for large lakes were merged. A few lakes without lake ID were removed. Taxa names were replaced by species code according to our selected Fennoscandian list of species. Replicate surveys and surveys prior to 1970 were removed.
Macrophyte data and lake data were merged. The macrophyte dataset was reduced from 1145 plant surveys to 610 lakes with both macrophyte and lake co-variates, including 130 reference lakes.

The Swedish dataset combined data from the former intercalibration exercise for the WFD (a lot of old data from literature review) and more recent data from SLU (2007 onwards). The SLU macrophyte data were downloaded directly from the portal (as Excel spreadsheet) and the co-variates were provided separately by SLU (Excel spreadsheet). Elevation was added to the Swedish data using rgbif (Chamberlain et al 2024) and Google Earth. The water colour units in absorbance at 420 nm (filtered water) were multiplied by 500 to convert them into mg Pt L^{-1}. The WFD and SLU macrophyte databases held 254 (single surveys, 71 without year) and 696 plant surveys (in 417 lakes), respectively. Multiple waterbodies for large lakes were merged. Duplicates, repeated surveys (retaining the most recent surveys less affected by sampling effort), and plant surveys without matching co-variates were removed. The final Swedish dataset had 517 lakes, including 139 reference lakes.

The sf package (Pebesma 2018) was used to harmonise the national geographical coordinates to either latitude-longitude WGS84 EU EPSG:3035 or projected coordinates ETRS89-extended / LAEA Europe EU EPSG:3035.

Climate data were retrieved from E-OBS (E-OBS gridded dataset (ecad.eu)), Cornes et al 2018)

For statistical modelling, independent climate variables/predictors comprised daily mean temperature TG, daily minimum temperature TN, daily maximum temperature TX and daily precipitation sum RR from the 0.1 degrees regular grid E-OBS ensemble dataset (v26.0e). Climate normal period was defined as 1980 - 2010 (31-year period).

All operations on climate variables were carried out using the CDO (Climate Data Operators) software tool (Overview - CDO - Project Management Service (mpg.de)). Climate variables were bilinearly interpolated to the 10 km EEA reference grid (Europe shapefile 10 km 100 km — European Environment Agency (europa.eu)), and monthly average temperatures and precipitation for the 30-year climate normal period were calculated. The sum of degree days per year for the climate normal period was derived from the sum of degree Celsius above a daily base temperature of +5 degree Celsius divided by the number of years (31). Lakes near the coast could not always be interpolated, hence some missing values (representing 13% of the lakes).

As part of data preparation, further data imputation filled all the missing values for water quality parameters, but the imputed data were not used in the present study. Multivariate interpolation in

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

form of inverse distance weighting (IDW) from R package 'gstat' (Pebesma 2004) was applied to water chemistry parameters alkalinity, colour, total nitrogen, total phosphorous, electric conductivity and chlorophyll a. IDW parameters in function 'gstat' were set to an inverse distance weighting power 'idp' = 0.5, number of nearest observations used 'nmax' = 10, and maximum distance 'maxdist' = Inf.

R script available on github:

r-spatial/gstat: Spatial and spatio-temporal geostatistical modelling, prediction and simulation (github.com)

Representativeness assessment

There is a good geographical coverage for species distribution models and the dataset is large enough to include uncommon species (present in 2-5% of the lakes). Plants from lake shore (helophytes) were not included and this restrict conservation assessment to a shorter list, but submerged and floating plants are good indicators of lake conditions. While tracheophytes (vascular plants) and charophytes (macroalgae) were included, bryophytes were excluded as not recorded in Norway. The derivation of community status assessment from individual species distribution model will not be reliable for mountain and northern lakes rich in bryophytes.

Product: EBV-ready data set for Fennoscandia

The full list of species is reported in Appendix 1. Table 5 lists the co-variates available in the EBV-ready dataset. The dataset and R scripts have been made public on Github: benoit-demars73/EuropaBon (github.com).

Table 2. Norway raw data availability and access

DATASET TITLE	NIVA lake macrophyte
Raw data collection design	Aquatic plant surveys were mostly carried out by NIVA following a predetermined species list. All macrophyte data were recorded based on comparable survey methods, mainly using a bathyscope (underwater viewer) and dragging a casting rake from the boat at different localities (random "walk") covering the diversity of habitats around the lakes. Environmental data were acquired at the time of plant surveys. See Mjelde et al 2023 for more details.
Monitoring programs	The data were acquired through individual projects to assess lake quality, and more recently through the EU WFD monitoring (Direktoratsgruppa vanndirektivet 2018).
Types of data access	Restricted access until now, opening it with the publication of this work
Data repositories	NIVA lake macrophyte database, NIVA water chemistry AquaMonitor
Persistent identifier(s)	NA
Metadata description	Species list with authorities: https://www.niva.no/omradesider/fotoflora-for-norske-vannplanter
Other provenance information	Lake area and elevation: NVE lake database (https://atlas.nve.no/Html5Viewer/index.html?viewer=nveatlas#) Climate data: rainfall, average monthly temperature (minimum. mean and maximum), derived annual sum of degree days from Cornes et al (2018)

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Table 3. Finland raw data availability and access

DATASET TITLE	SYKE lake macrophyte
Raw data collection design	Aquatic plants were recorded in phyto-littoral inventories and along transects (see Hellsten et al 2014), selected surveys span 1972-2022. Both methods include visits at several sites and with survey ranging from uppermost littoral (helophyte zone) to deepest growing points of submerged plants. Water chemistry was an average of sampling points per lake during June - September and for the period 1990 – 2015.
Monitoring programs	The surveys were listed in Leka 2008, cited in Hellsten 2014. EU WFD monitoring in place now
Types of data access	Open access now
Data repositories	SYKE lake macrophyte database, SYKE water chemistry database
Persistent identifier(s)	NA
Metadata description	Species list with authorities: Kuoppala et al 2008, Appendix 3
Other provenance information	Climate data: rainfall, average monthly temperature (minimum. mean and maximum), derived annual sum of degree days from Cornes et al (2018)

Table 4. Sweden raw data availability and access

DATASET TITLE	SLU lake macrophyte
Raw data collection design	Aquatic plants were predominantly surveyed from whole lake surveys using different methods (1920s-2005), and then along transects (Ecke 2014)
Monitoring programs	The macrophyte monitoring programs and available data were reviewed in Ecke 2007. EU WFD monitoring now in place (Ecke 2015). Water quality data followed standard collection (surface water) and analytical protocols (e.g. Folster et al 2014). Annual median of all sampling locations within a lake was used.
Types of data access	Literature review (Lohammar 1938, Lundh 1951, Andersson 1999, Ecke 2007, Johnson & Toprak 2021) and SLU open access portal (https://miljodata.slu.se/mvm/). SLU provided the water chemistry and lake area data from their database (data open access).
Data repositories	Restricted access to REBECCA WFD intercalibration (JRC copy holder) and WISER macrophyte database (NIVA copy holder).
Persistent identifier(s)	NA
Metadata description	No current list of species for the survey. List of hydrophyte scoring taxa for the WFD (Characeae, Bryophyta, Tracheophyta). Helophytes not included in WFD index.
Other provenance information	Climate data: rainfall, average monthly temperature (minimum. mean and maximum), derived annual sum of degree days from Cornes et al (2018)

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Table 5. Data ready co-variates in Fennoscandian lakes

Lake surveys	Number of records	completion
EuropaBon_ID	2059	100 %
Reference lake	403	NA
Country	2059	100 %
Year	2010	98 %
X_3035	2059	100 %
Y_3035	2059	100 %
Lake area_km2	2053	100 %
Elevation	2046	99 %
Alkalinity	1854	90 %
Water colour	1348	65 %
TN	1602	78 %
TP	1824	89 %
Annual sum degree days	1780	86 %
Average annual rainfall	1780	86 %
Electric conductivity	1111	54 %
Chlorophyll_*a*	1052	51 %

4.4. The EBV model

4.4.1. Data preparation

- A complete dataset (no missing values) of key co-variates lake surface area, alkalinity, colour, total phosphorus and nitrogen and annual sum of degree days was available for 1081 lakes (out of 2059 lakes). Very strong environmental gradients were available (Table 6) and not strongly correlated ($R < 0.7$; Figure 4).
- Plant species were recorded as present-absent. The species pool is 97 species, of which 67 occur in at least 20 lakes. The species rank abundance shows the number of lakes where *Potamogeton lucens* (calcareous species) and *Isoëtes lacustris* (siliceous species) are found – (Figure 5).
- The selection of co-variables was based on ecological knowledge and parsimony (limited collinearity), so simple stepwise selection (using AIC) of co-variates was considered appropriate. More complicated methods were not implemented (e.g. variable selection with lasso, ridge, covsel, or random stratified sub-sampling along gradients to "un-correlate" co-variates)
- The reference lakes were distributed across Fennoscandia (Figure 6). The initial thought was to run species distribution models on reference lakes separately from the rest of the lakes, but there were not enough reference lakes to include less common species (e.g. *Potamogeton lucens* was only present in 5 reference lakes). Total phosphorus under reference conditions (TP_ref) was highly related (explained deviance 77.4%) to key co-variates using gam (package mgcv, thin-plate regression splines; Wood 2003, 2017) as follows:

  log(TP_ref) ~ log(Area_km2) + s(log1p(Alk)) + s(log(Colour)) + s(X_3035, Y_3035).

  TP_ref increased with lake area, alkalinity, colour (organic P) and increased from North-west to south-east (Figure 7). TP_ref was then predicted at all lakes knowing lake area, alkalinity, colour and geography.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

Table 6. Selected environmental co-variates for the 1081 Fennoscandian lakes

Co-variates	units	Min	25th	50th	75th	Max
Lake area	km^2	0.0001	0.53	1.52	5.80	3086
Elevation	m	1	67	108	175	1234
Alkalinity	mEq L^{-1}	-0.05	0.11	0.19	0.48	9.03
Colour	mg Pt L^{-1}	1	22	44	83	364
TN	µg N L^{-1}	30	326	474	720	4460
TP	µg P L^{-1}	1	8	15	31	380
TN:TP		2	20	29	41	661
Annual sum of degree days	°C	245	1049	1236	1377	1711

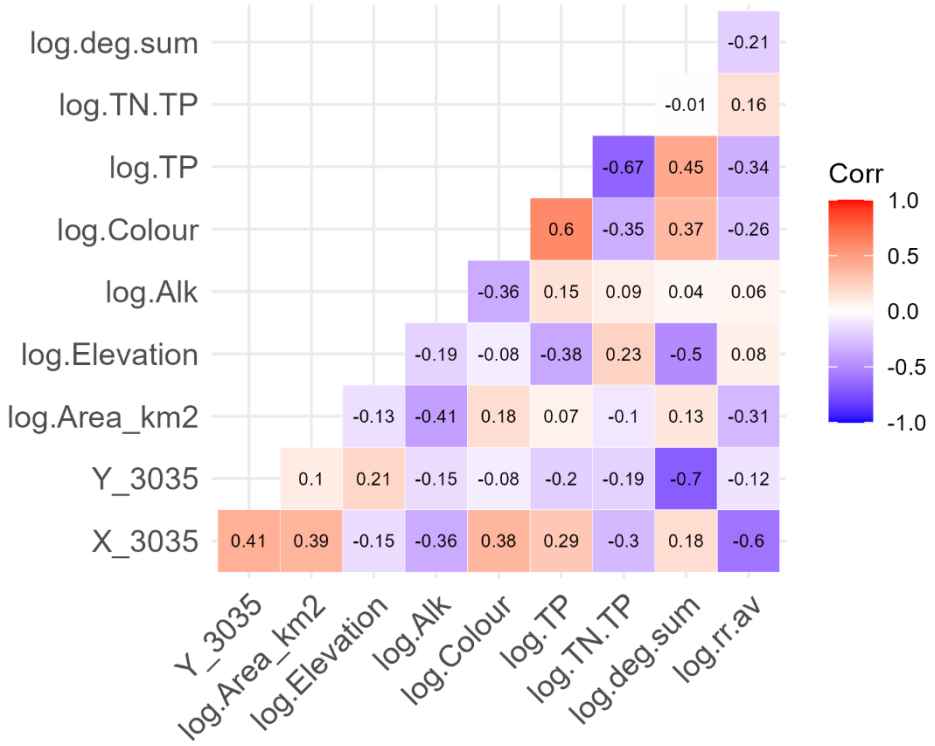


Figure 4. Correlation matrix of spatial and environmental co-variates. Matrix drawn with ggcorrplot (Kassambara, 2023) More detailed correlation matrix and density plots available in R script (section 4.3).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

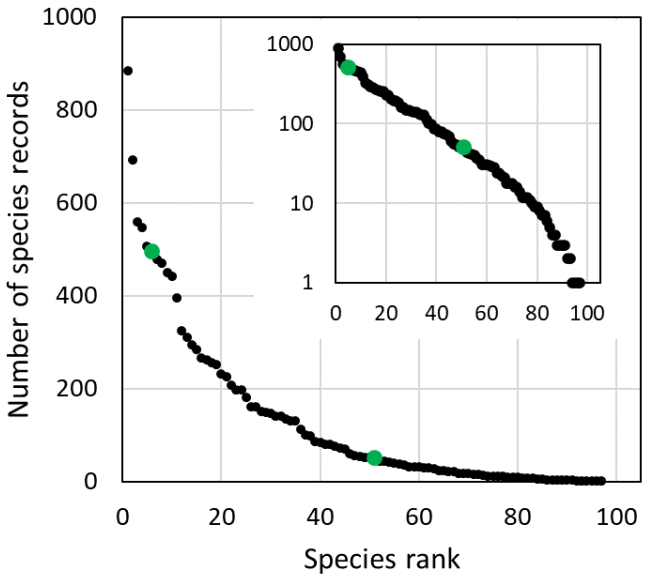


Figure 5. Species rank abundance for the 1081 Fennoscandian lakes. Inset: log transformed y axis. The position of Isoëtes lacustris (n=495) and Potamogeton lucens (n=50) is shown with larger green dots. Graph drawn in Excel (Microsoft).

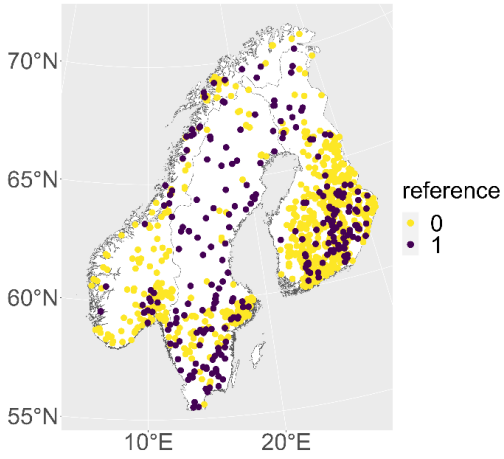


Figure 6. Distribution of the 254 reference lakes (dark dots) in the 1081 Fennoscandian lakes (all dots) with a complete set of co-variates. Maps made with ggplot2, sf and raster packages in R. Custom made map with occurrences brought to the front, using sf (Pebesma, 2018), raster (Hijmans 2023) and ggplot2 (Wickham 2016) packages in R, see R scripts (section 4.3).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

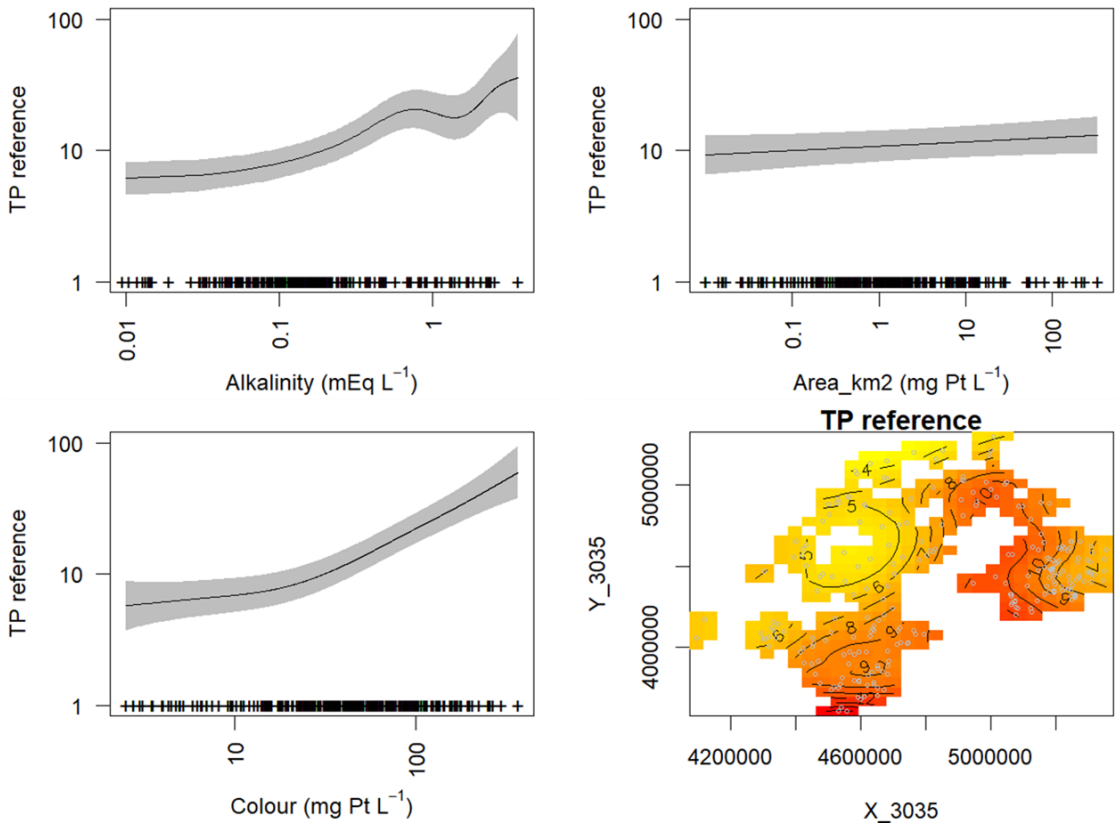


Figure 7. Lake specific predicted total phosphorus (TP) reference conditions in Fennoscandia as a function of alkalinity, lake area, water colour and geography based on 254 reference lakes. Partial plots with alkalinity, lake area, water colour set at their geometric means, and xy coordinates at their means when not on the plot. All data were back-transformed to their original scale for the plots (TP in µg P L^{-1}). Custom made graphs with the plot function in base R (R Core Team, 2023), and modified vis.gam from mgcv package in R (Wood 2017), see R script (section 4.3).

4.4.2. Statistical workflow

The statistical workflow followed recommendations from general statistical textbooks, notably Jongman et al 1995; Legendre & Legendre 1998; Guisan et al 2017.

Step 1. Spatial gam (mgcv R package, Wood 2003, 2017)

- manual model check (convergence, explained deviance, coefficients, effective degree of freedom) and concurvity (collinearity) check,
- manual stepwise model selection (AIC),
- autocorrelation tests of model residuals using spatial spline correlogram (ncf package, Bjornstad 2022): non-directional (Figure 8) and directional (time consuming to run). In the two example species, spatial auto-correlation was not or barely significant in the residual of the model. If significant re-run gam with neighbourhood cross validation to minimise short range spatial autocorrelation,
- Investigate effect of individual co-variate on species response : plot of partial response (continuous probability of species occurrence 0-1) against individual co-variates while setting other co-variates to their geometric mean (since our data were log-transformed) – except for alkalinity set at 1.5 and 0.1 mEq L^{-1} for *Potamogeton lucens* and *Isoëtes lacustris* to get more relevant plots using alkalinity values near the optimum of the species (*Figure* 9). Note how uncertain the *Potamogeton lucens* optimum for TP is between the fit of the model and the uncertainty bounds. All plots were custom made, see R script (section 4.3).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

- calculate threshold (TSS, Kappa) to convert continuous fitted probabilities in binary probabilities (presence or absence) and estimate model performance through the area under the ROC curve (ROC = Receiver Operating Characteristic) using the ecospat package in R (Broennimann et al 2023).
- extract model fit and residual: fitted continuous response, standard error, calculate and rescale lower and upper 95% confidence interval, range of confidence interval on response (0-1) scale (to produce uncertainty maps), fitted probabilities minus observed values, predicted presence-absence using the cut-off thresholds determined by maximum TSS or Kappa (change continuous probability into binary format); save and write file to produce maps
- Produce custom made maps using sf (Pebesma, 2018), raster (Hijmans 2023) and ggplot2 (Wickham 2016) packages in R: observations, fitted continuous probabilities (0-1 scale), uncertainties (range of 95% confidence interval on response scale 0-1), predicted occurence from TSS or Kappa (binary) – see Figure 10-Figure 13.

Step 2. Non-spatial glm (R Core Team, 2023) or spatial glm (mgcv R package Wood 2003, 2017), following same flow-path as gam model, use of *non-orthogonal* second-degree polynomial to allow derivation of species maximum, optimum and tolerance for each selected co-variate (except for lake area set with a simple linear coefficient, assuming probability of species occurence linearly increase with log sampling area). Note the coefficient of the quadratic term should be negative (otherwise the quadratic term may be dropped, Roy et al 2000, Demars & Trémolières 2009). When the optimum is outside the range, then the minimum or maximum (may be better 2.5 or 97.5 centile) should be applied. Some co-variates may not be retained after the step selection and simply won't be an indicator; otherwise it could be extracted from the full (sub-optimal) model, knowing that the regression coefficients of the remaining co-variates do change a little after the stepwise selection – see Table 7 and R script for the semi-automated calculation details (section 4.3). These optima provide quantitative species indicator values (habitat preference or ecological traits) with their uncertainties (niche tolerance) which allow the calculation of community indices with a simple weighed average method (ter Braak & Barendregt 1986; Demars & Trémolières 2009). The calculation of these binary predictions for all species will allow to derive a predicted list of species per lake to calculate community EBVs such as species richness and species turnover along environmental gradients, and ecological quality ratios (EQRs) derived. Custom made maps, see text and R script (section 4.3).

Step 3. Ensemble modelling (biomod2 4.2-4 R package, Thuiller et al 2023)

- Model tuning (glm worked well, but some issues with gam).
- While gam run smoothly and quickly in step 1 above, gam did not converge for *Potamogeton lucens in biomod2* even when simply using the default mgcv algorithm.
- Model evaluation through random split sampling using 70% of the data for training and 30% for validation, 10 runs per algorithms, display boxplots of calibration and validation runs for the different algorithms and performance indicator (TSS, Kappa, ROC), display partial plots of the 10 runs of species response per algorithms (see Figure 14-Figure 15
- Individual model performance (TSS, Kappa, ROC): GAM, GLM
- Ensemble modelling (weighted average of all runs by the TSS of the validation runs)
- Extract all model results and produce custom made maps (as for step 1 and 2), see R script (section 4.3) for more details (Figure 16).

Step 4. Save and share data, metadata, R scripts and maps (see Github link in section 4.3)

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

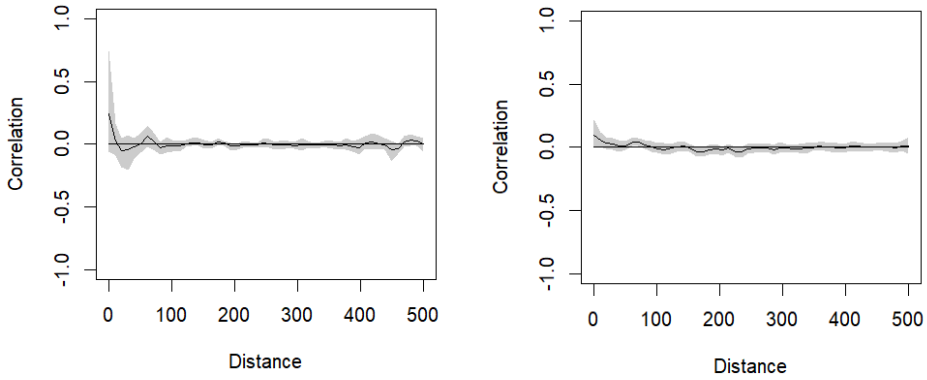


Figure 8. Potamogeton lucens (left) and Isoëtes lacustris (right) spatial autocorrelation of the model residuals (glm, same results for gam not shown) estimated every 10 km over 0-500 km distance. Auto-correlogram drawn with ncf package in R (Bjornstad 2022).

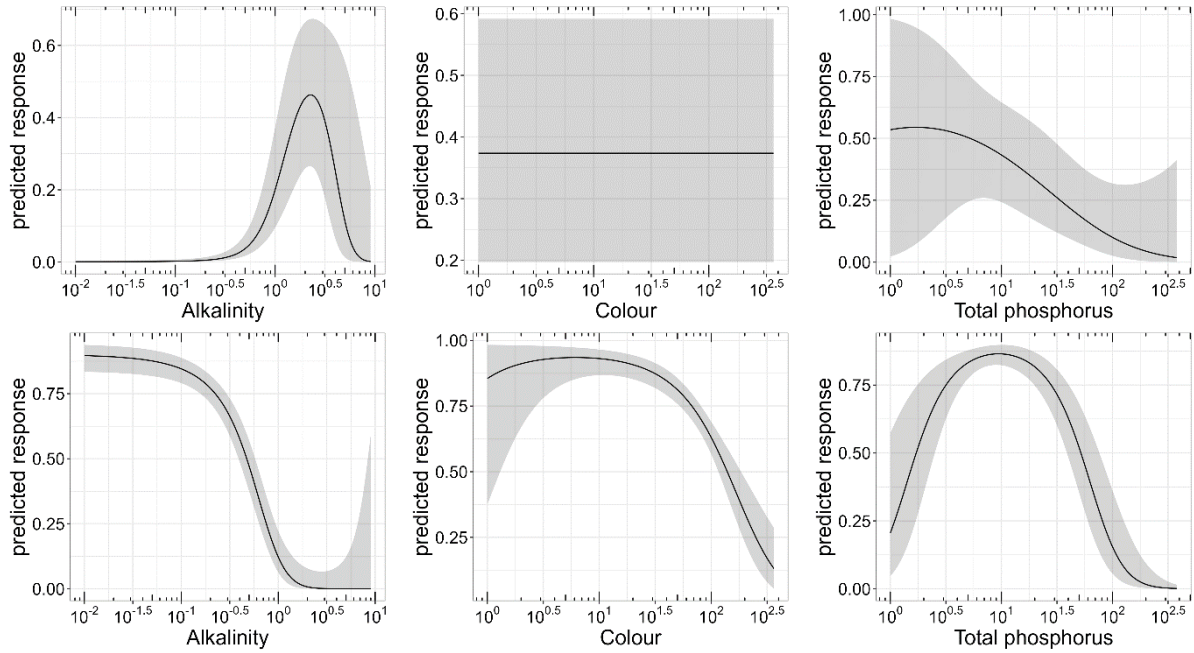


Figure 9. Partial plot response of Potamogeton lucens (top) and Isoëtes lacustris (bottom) from the glm models (similar results with gam not shown). Note water colour was not selected in the glm for Potamogeton lucens, but selected by the gam and biomod2 (ensemble models). Custom made graphs with ggplot2 in R (Wickham 2016), see R script (section 4.3).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

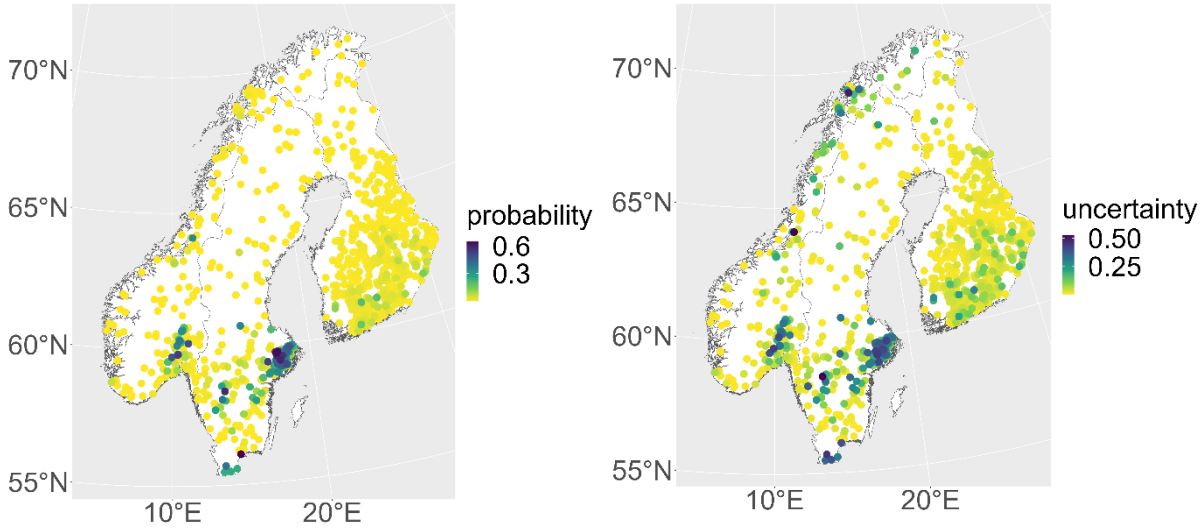


Figure 10. Potamogeton lucens probability of occurrence (glm model) and uncertainty (95% confidence interval) Model adjusted explained deviance D2=0.39. Custom made maps (see text) with highest probabilities and uncertainties brought to the front, see R script (section 4.3).

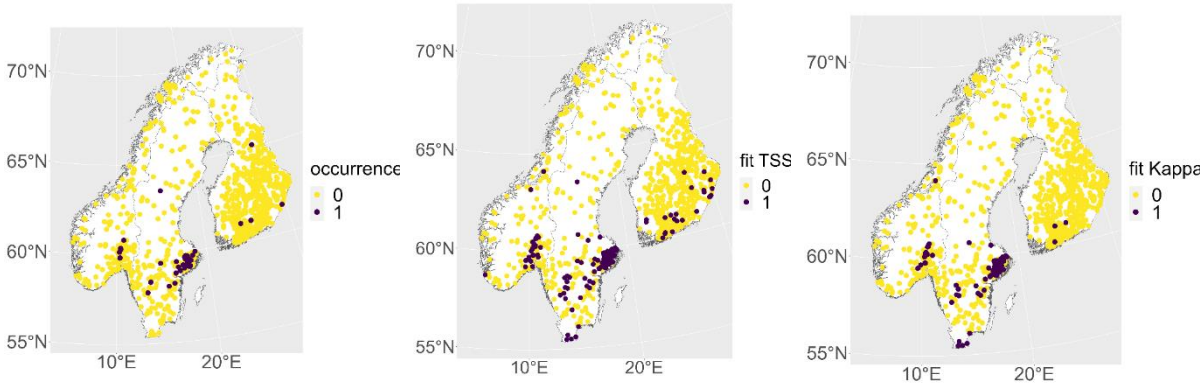


Figure 11. Potamogeton lucens observed (left, occurrence in 50 lakes) and fitted presence-absence after filtering the continuous probabilities of the glm model with thresholds: TSS=0.80 with response threshold 0.05 (middle, occurrence in 189 lakes) and Kappa=0.42 with threshold response 0.14 (right, occurrence in 120 lakes). Model performance AUC = 0.93.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

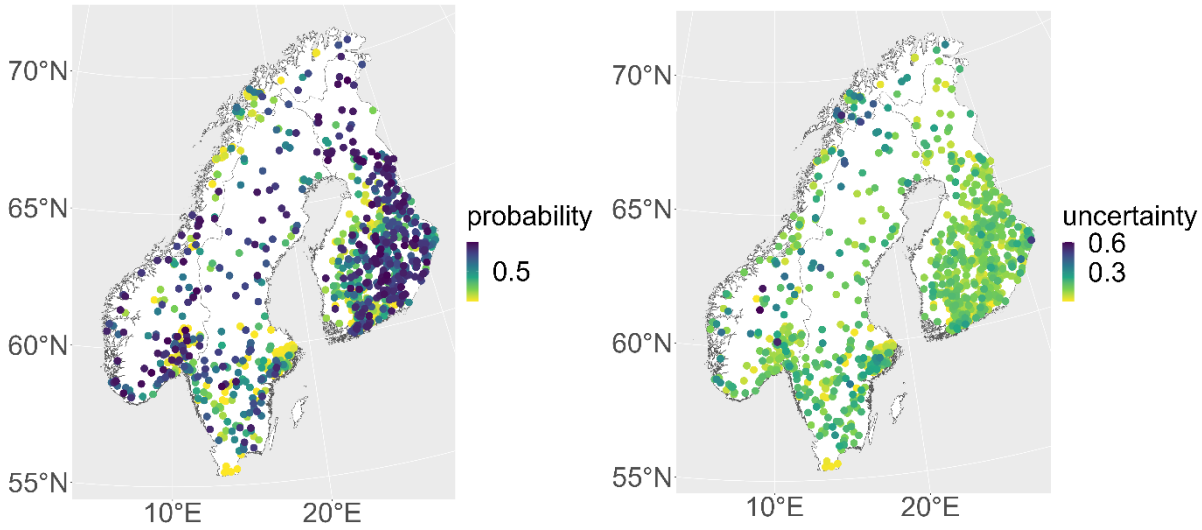


Figure 12. Isoëtes lacustris probability of occurrence (glm model) and uncertainty (95% confidence interval). Model adjusted explained deviance D2=0.42. Custom made maps, see text and R script (section 4.3).

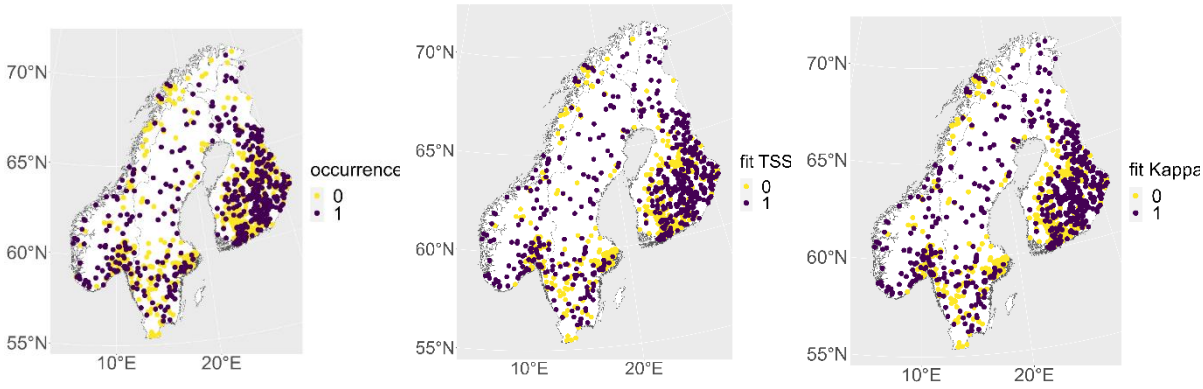


Figure 13. Isoëtes lacustris observed (left, occurence in 495 lakes) glm model with thresholds: TSS=0.65 with response threshold 0.42 (middle, occurrence in 580 lakes) and Kappa=0.64 with threshold response 0.42 (right, occurrence in 580 lakes. Model performance AUC = 0.90. Custom made maps, see text and R script (section 4.3).

*Table 7. Species optima and tolerance derived from the glm model coefficients. *set to minimum as optimum was outside the range of observed values (-0.9 ±0.1 mEq L^{-1}). ** when selected in sub-optimal model.*

	Alkalinity	TP	Water colour	Annual sum of degree days
units	mEq L^{-1}	µg P L^{-1}	mg Pt L^{-1}	°C
Isoëtes lacustris	*0.01±0.1	9 (4-23)	6 (2-24)	877 (566-1359)
Potamogeton lucens	2.3 (1.4-3.5)	2 (0.3-11)	**28 (11-75)	1430 (1140-1790)

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

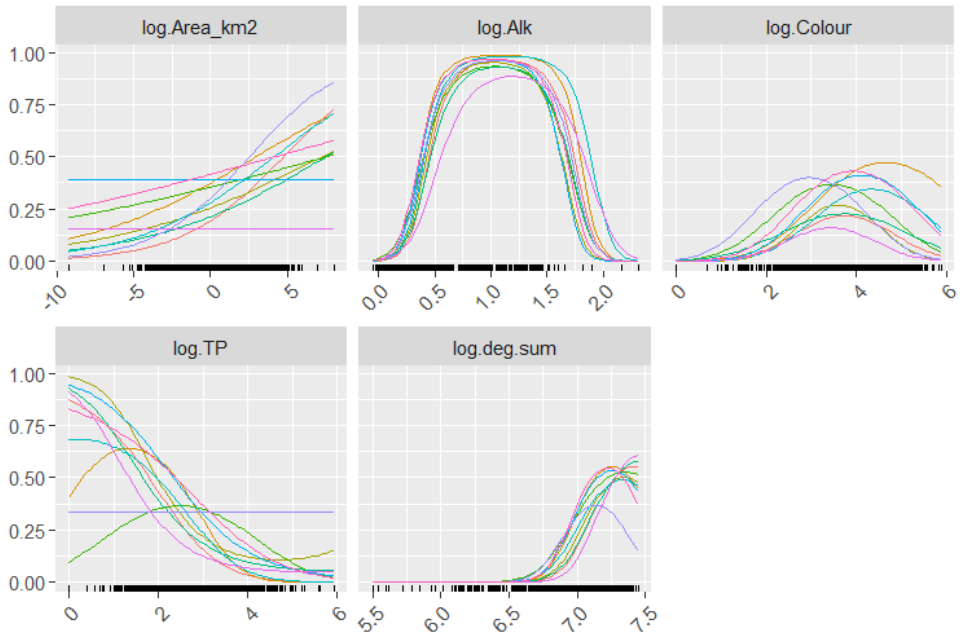


Figure 14. Partial plot responses of Potamogeton lucens using random split sampling, 70% calibration, 30% for validation from biomod2 ensemble model package. 10 runs with the same algorithm (glm). Plots from biomod2 package in R (Thuiller et al 2023). The y axis scale is bounded 0-1 as for the other graphs above, but the response curves were automatically rescaled to facilitate comparison between algorithms (Elith et al 2005). The x axis is the natural logarithm of the covariate and ln(alkalinity+1) (note gam models did not converge in biomod2 for this species)

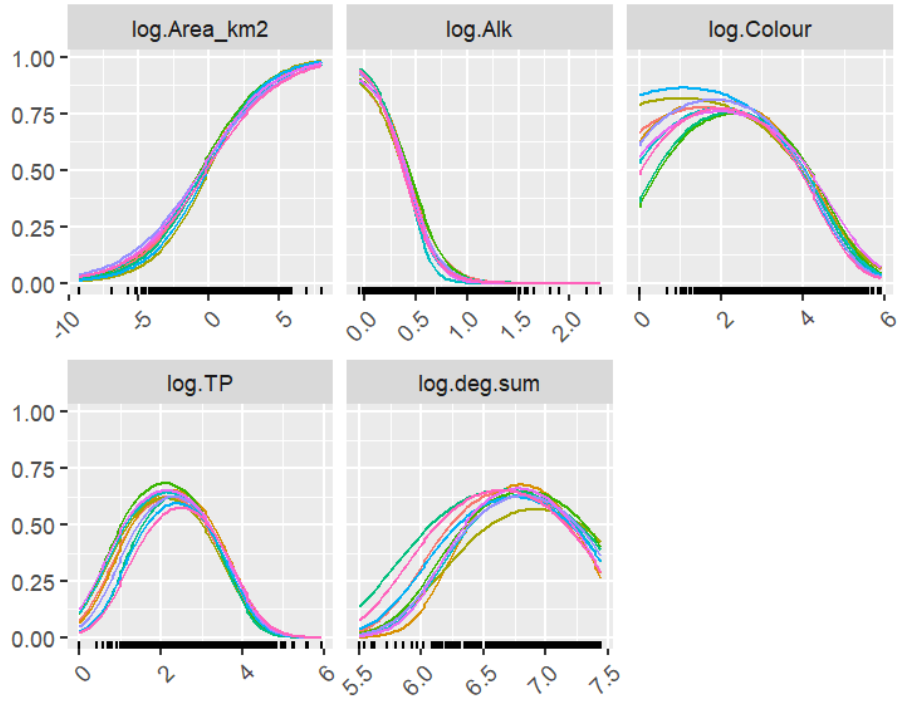


Figure 15. Partial plot response of Isoëtes lacustris using random split sampling, 70% calibration, 30% validation from biomod2 ensemble model package in R (Thuiller et al 2023). 10 runs with the same algorithm (glm). The gam model produced similar results (not shown). Biomod2 package graph (same legend as previous figure).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

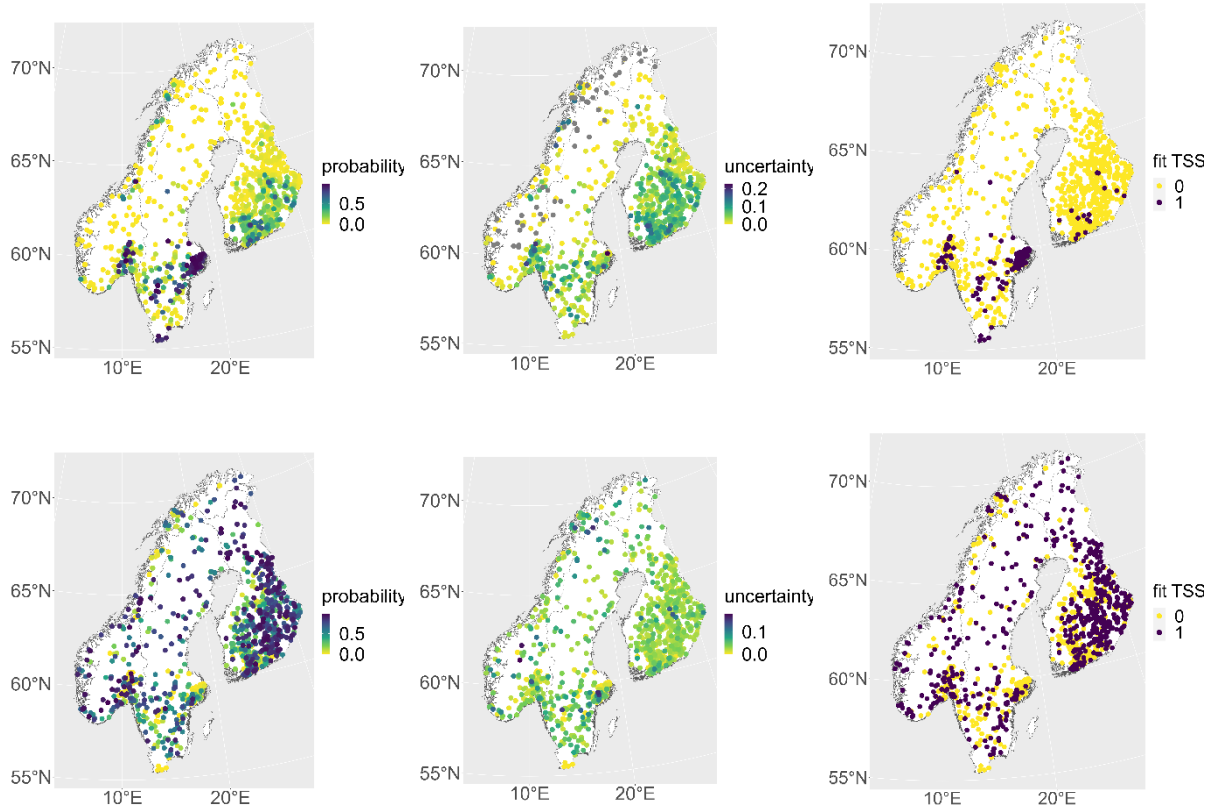


Figure 16. Potamogeton lucens (top) and Isoëtes lacustris (bottom) probability of occurrence (weighted mean), uncertainty (standard deviation) and binary fit with TSS threshold using biomod2 ensemble models. The mean was weighted by the validation run performance using TSS. The uncertainties relate to differences in model runs (glm only for P. lucens; glm and gam for I. lacustris). Custom made maps, see text and R script (section 4.3).

4.4.3. Spatial biodiversity data gaps

- **Data preparation.** Nearly 5000 lakes were surveyed for water chemistry in Finland (873), Norway (1006) and Sweden (3075) as part of the georeferenced northern European lake survey of acidification in 1995, representing 3.5% of the 128624 lakes over 0.04 km^2 (Henriksen et al 1998). The survey followed a stratified random survey design, with a local bias towards lakes more likely to be sensitive to acidification (base poor water). Fish (people) surveys were also carried out at the same time from 1995 to 1997 in 3821 lakes (Tammi et al 2003). There were no dedicated simultaneous aquatic plant surveys, but the chemical co-variates can be used to project species distribution models trained on the Fennoscandian macrophyte dataset. NIVA has only curated the Norwegian data (available in NIVA's Aquamonitor database). Fortunately, a copy of the whole dataset for a subset of the water chemistry including base cations (Ca, Mg, K, Na in µEq L^{-1}; total phosphorus (µg P L-1), lake surface area) was found and also extended with catchment variables and air temperature by Anders Finstad (Department of Natural History NTNU University Museum) and colleagues (Perrin et al. 2022). Here we added a more comprehensive set of climate data corresponding to our Fennoscandian macrophyte dataset (temperature, precipitation and annual sum of degree days). We used a more comprehensive Norwegian lake dataset for water chemistry (1000 lake survey from 2019) to relate the sum of base cation molarity to alkalinity (µEq L^{-1}) with [alkalinity]=0.1926

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

[cations]^1.0719, R^2=0.82. Water colour and total nitrogen were unfortunately not available in the copy of the data we have, thus further work between NIVA, SLU and SYKE is necessary.

- **Species distribution models**. For now, we simplified our model (glm) with the set of co-variates available (lake surface area, imputed alkalinity, total phosphorus and annual sum of degree days) using 1259 lakes from our Fennoscandian macrophyte database. We then projected the species distribution of *Potamogeton lucens* and *Isoëtes lacustris* at 3693 lakes with a high degree of confidence in the geographic position and complete set of co-variates (*Figure* 17). The species projection using the 1995 northern Europe lake surveys has enabled to considerably increase the coverage. The projection of *Isoëtes lacustris* is similar to its known distribution, but *Potamogeton lucens* was not as well predicted, largely because calcareous lakes were missed (truncated alkalinity gradient) from this massive survey more focused at the time on acidification issues for fish. When it comes to biodiversity, it will be important to combine a good geographical and environmental representation, making sure rare habitats such as calcareous lakes in Fennoscandia are sampled adequately.
- **Wall-to-wall biodiversity.** Further work on the 1995 northern European lake survey of water chemistry would allow to include more lakes and more co-variates. We also tried to impute missing data in our 2059 lake macrophyte database (see section 4.2 Input Biodiversity Data), but limited our effort to the use of the water chemistry of the northern European lake survey, due to lack of time. The imputation method we used is based on autocorrelation, and likely will smooth a lot the data making rare habitats “invisible”, a common issue with implication for the application of species distribution models.
- **Gridded data.** Species distribution models could also be gridded in the final step to make them comparable to e.g. open access atlas data. Gridded biodiversity data may have its use at large spatial extent (national to continental and global). If the lake co-variates are gridded spatial heterogeneity is smoothed, rare habitats are less ‘visible’ and the conservation of rare species using species distribution models could be biased (textbook knowledge, see e.g. Guisan et al 2017). The species distribution models based on lake co-variates, from a comprehensive sampling design and good taxonomic recording, offers a more powerful approach for management, conservation and a range of environmental policies.

This project receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

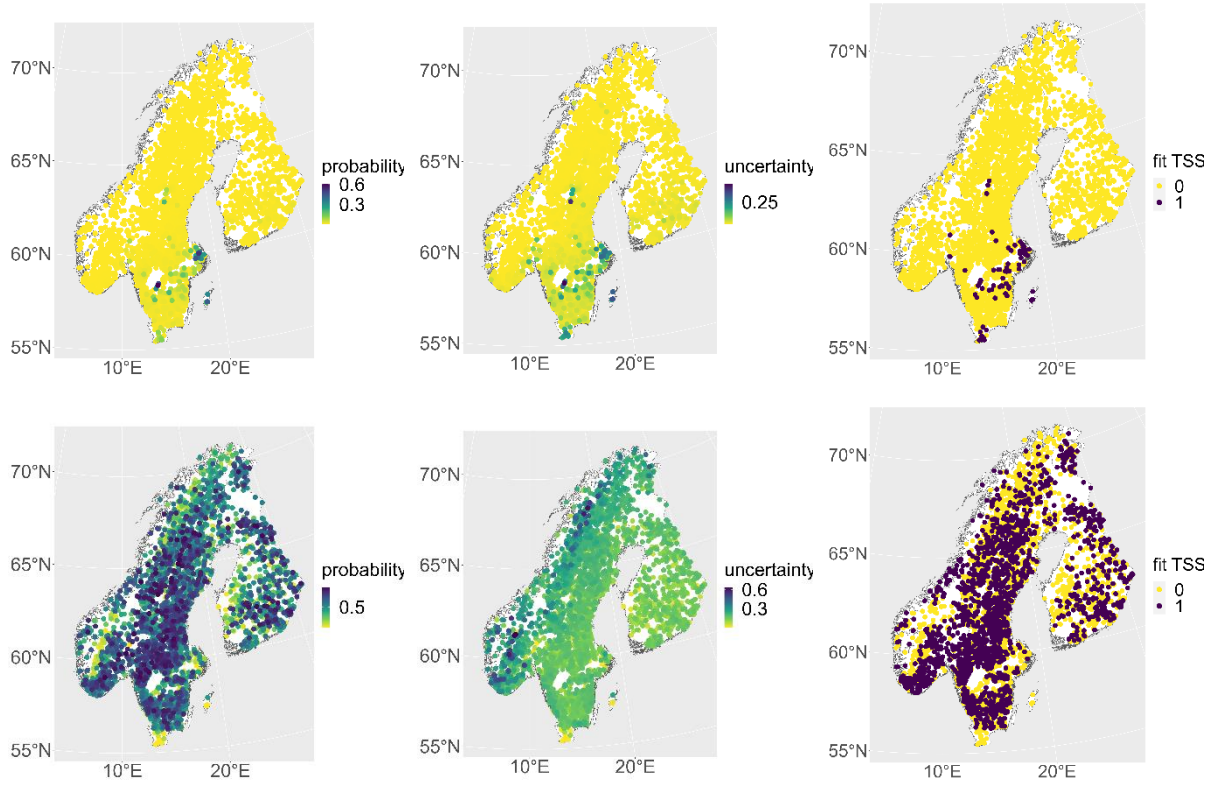


Figure 17. Projected species distributions (glm) in 3693 Fennoscandian lakes from the 1995 northern European lake water chemistry survey: Potamogeton lucens (top) and Isoëtes lacustris (bottom) with continuous probability (left), 95% confidence range of uncertainty (middle) and species projected occupancy (right).

4.4.4. Spatial status assessment

There are known gaps between the requirements of the EU Water Framework Directive and the EU Habitats Directive (Irvine 2009, Ecke et al 2010). The above species distribution models can be used to make predictions under varying environmental conditions within the same geographical extent and environmental space (avoiding extrapolation). The species status was estimated through the comparison of the species probability of occurrence in current projections relative to TP reference conditions (Figure 18). The occurrence of *Potamogeton lucens* was predicted to be lower under current projections (n=189) than under reference conditions (n=213). The occupancy of *Isoëtes lacustris* was also predicted to be lower under current conditions (n=580) than under reference conditions (n=641). The change in occupancy in current projections relative to TP reference conditions can be illustrated by maps displaying the number of records in **common** (species occupancy in current projections and under TP reference conditions), **gain** (species occupancy in current projections but not under TP reference conditions) and **loss** (no species occupancy in current projections but occupancy under TP reference conditions). These differences may be expressed as simple indices (scale 0-1): Sørensen β-diversity ($\beta_{sør}$) = turnover + nestedness; and together with maps form a powerful visual understanding (Figure 18). Both *Potamogeton lucens* and *Isoëtes lacustris* had low $\beta_{sør}$: 0.08 and 0.06 respectively dominated by a loss (Figure 18) in species occupancies (nestedness = 0.06 and 0.04, respectively). Thus, both species were generally close to what would be expected under reference conditions in Fennoscandia, as expected from their low TP optima 2 and 9 µg P L^{-1}, respectively. The maps highlight lakes with potential concerns, i.e. loss in occupancies (Figure 18).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

The above species and lake status assessment only worked because the species optima for TP tended towards lake TP reference values (recall that the lake TP reference values vary with other state variables, see Figure 7). Many species have optima away from reference conditions, e.g. free-floating species thrive in lakes with high TP concentrations. So, we need to separate species status assessment from lake status assessment. Current species projections should be compared against species projections under their TP optimum conditions: this provides a status assessment from the species' eyes (independently of what humans regard as reference conditions). The lake status assessment needs to integrate how the species optimum relate to reference conditions. A general equation for lake status assessment in Fennoscandia from where species *i* occurs (L_i) is thus:

$$L_i = \frac{O_i(current)}{O_i(optimum)} \times \frac{O_i(reference)}{O_i(optimum)} = S_i \times \frac{O_i(reference)}{O_i(optimum)}$$

with O_i the number of projected occupancies (continuous probability exceed TSS threshold) of species *i* under current, reference and TP optimum conditions. The first term is the species status assessment from the species' eye (S_i). The second term expresses how far apart the projections in species occupancies are between species optimum and lake ecological reference conditions. Since $O_i(optimum)$ corresponds to the maximum probability of occurrence along the pressure gradient for a given set of state variables, both terms scale 0-1, and L_i scales 0-1. When $O_i(optimum)$ tends to $O_i(reference)$, then the second term tends to 1 and the lake status assessment tends to the species status assessment, as in the special cases of *Potamogeton lucens* and *Isoëtes lacustris* illustrated above. L_i may be interpreted as a turnover in species occupancy between current projection and reference conditions weighted by the distance of the species optimum to reference conditions.

The changes in occupancy for individual species between current and reference conditions illustrate the contribution of individual species to community level lake status assessment as illustrated in Figure 18. If the species optimum is at the opposite end of the TP gradient relative to TP reference conditions, then differences between current and reference conditions will mostly translate in gain of occupancies: 'good' from the species point of view (S_i), 'bad' from the lake reference conditions point of view (L_i).

The use of species distribution models (EBV cubes) allows to project counterfactual scenarios, e.g. how does species occupancy change along the pressure gradient (Figure 19); how would the species distribution be changed if lake TP concentrations in moderate to bad conditions were halved following pollution abatement measures?

Note the two species selected to illustrate the EBV cube approach also represent important freshwater habitats in Fennoscandia. The stacking of predicted species distribution (spatial EBV cube) opens the possibility to derive community (richness, turnover) and habitat (through characteristic species) status assessment, offering an alternative pathway to the production of ecological quality ratios (EQRs) in the implementation of the EU WFD.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

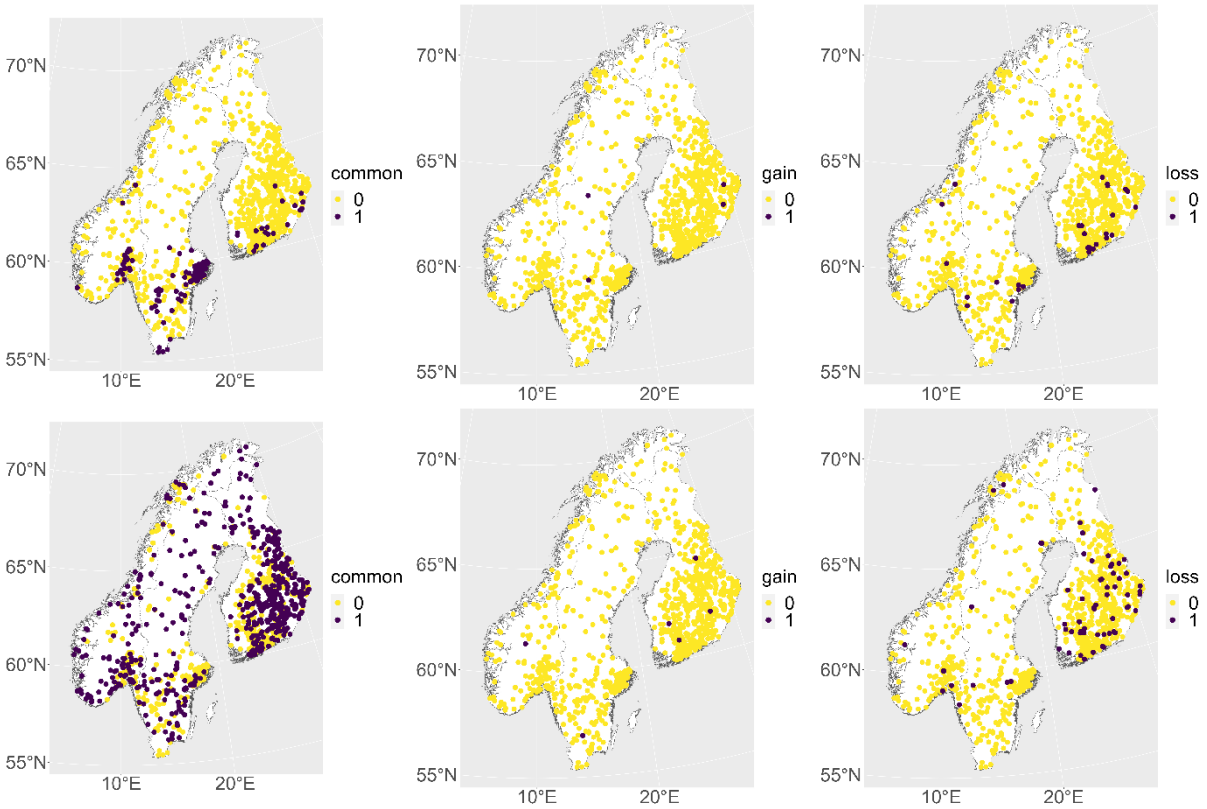


Figure 18. Species occupancy status assessment: species occupancy in common (left), gained (middle) and lost (right) in current species projections against reference TP conditions for Potamogeton lucens (top) and Isoëtes lacustris (bottom). Binary probability of occurrence (glm model, TSS threshold). Custom made maps, see text and R script (section 4.3).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

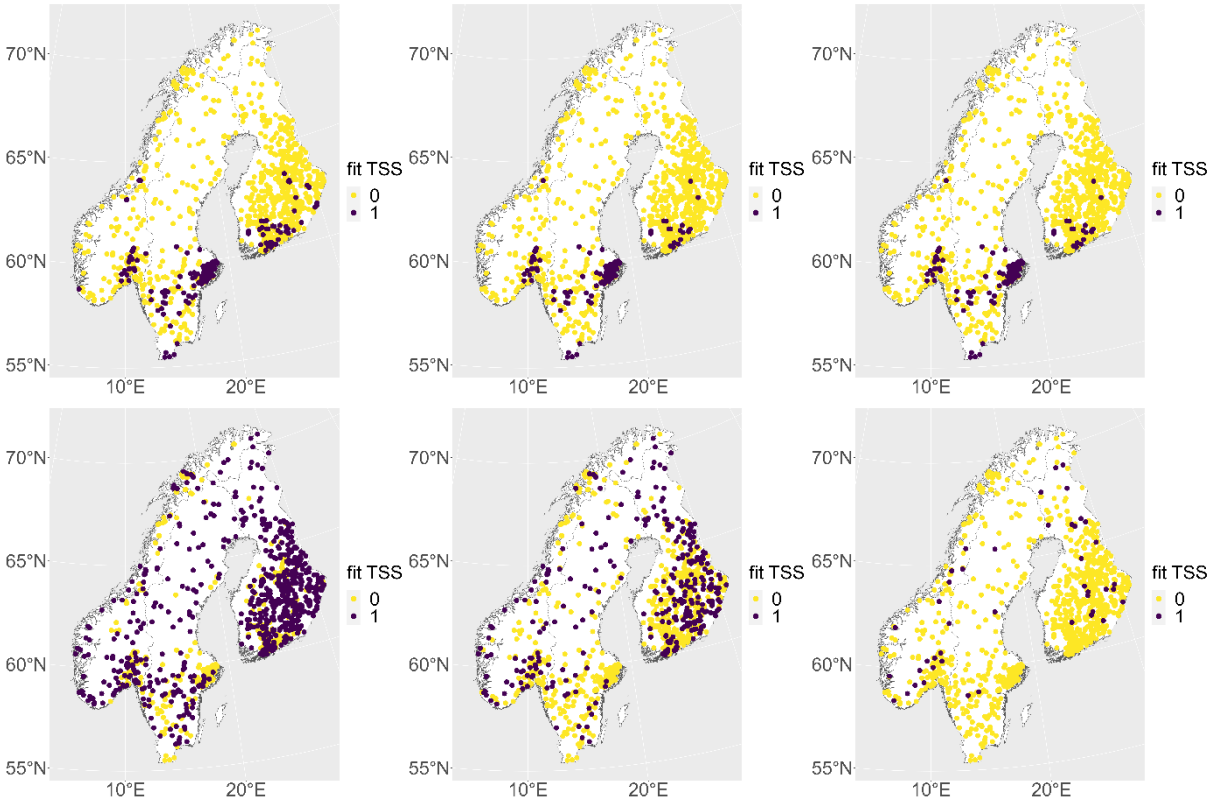


Figure 19. Projected species occupancy under varying TP scenarios: Potamogeton lucens (top) and Isoëtes lacustris (bottom) under 15 (left), 50 (middle) and 100 (right) µg P L^{-1}. These scenarios (EBV cubes) illustrate the contrasting response of two species. Potamogeton lucens is a tall and fast growing species able to tolerate high TP concentrations, while Isoëtes lacustris is a short perennial and slow growing species unable to reach sufficient light under high TP concentrations promoting phytoplankton growth.

4.5. Performance of EBV-derived policy indicators

Here we appraise the relevance of the new EBV product for policy reporting as compared to customarily used data flows.

- Current data flows vary greatly between countries in Europe. Species and associated environmental data (co-variates) are generally hosted in different databases, some have been dedicated specifically to freshwater (e.g. Sweden SLU portal). It is hard to know to what extent national open access portal for biodiversity report to GBIF can provide wider cross-country overviews of species distribution ranges. Here we brought together species (aquatic plants) and geo-referenced co-variates (lake surface area, water quality and climate data) to project species distribution.
- The species assessment method (spatial EBV cube) developed here is entirely free of hierarchies, i.e. it works with continuous co-variates rather than lake typologies, removing some inherent artificial uncertainties from categorical grouping. It provides lake specific maps of common, gain and loss of species occupancy between current projections and reference conditions. We also developed species indices to derive species and lake ecological status assessment (not necessarily responding in the same direction) based on current, optimum and reference projections of species occupancies – ecological quality ratios (EQRs) at the species level.
- Species optima (habitat preferences) and tolerances (uncertainties) were derived *concurrently* for multiple gradients ensuring a better characterisation of the realised species niche. The optima and tolerances can be used to derive quantitative community indices with ecological uncertainties (rarely if ever included in WFD EQRs). Note our novel approach through species projection goes beyond the simple use of species optima (not integrating individual species sensitivity along the

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

pressure gradient) and tolerance (propagating uncertainties along the environmental pressure gradient).

- Status assessment of current conditions relative to reference conditions was derived from species distribution models under the full environmental gradients of the state variables and predicted pressure value at reference condition (the later derived from the reference lakes). This quantitative approach propagates ecological and modelling uncertainties into species status assessment while not relying on too few lakes per lake type to derive ecological reference conditions. The only expert judgement remaining is the definition of reference conditions.
- The species distribution model approach allows to stack species (EBV cubes) to form predicted communities on which biodiversity indices may be derived (richness, turnover) for reference and current conditions as well as alternative scenarios, hindcasting and forecasting (e.g. assessing the effect of pollution abatement programs, climate change). EBV cubes can therefore be useful for assessing effects of restoration measures under the Nature Restoration Law and WFD.
- The use of species distribution models provides an alternative to the intercalibration approach of the WFD and integrate all conceptual approaches: Norway community indices based on species sensitivity (the slope in our models), Sweden index based on species phosphorus optima (the optima and tolerance in our models) and Finland index based on species turnover (can be derived from stacking our species distribution model) – *cf* Hellsten et al 2014.
- The availability of data and transparency of our approach (data and R scripts provided section 4.3) remove two major stumbling blocks for ecological status assessment of waterbodies, in sharp contrast to the WFD Macrophyte Northern GIG *intercalibration* exercise which did not report the data or the calculations and had no geographical display (Hellsten et al 2014). The WFD *intercalibration* Excel spreadsheets are currently held at JRC with restricted access and assessed by JRC of moderate to poor quality since “further assistance/explanation from the originator [consultant] would almost certainly be required”. Our approach could make it possible for anyone to use and improve or to cross-check improved national community indices (EQRs) or new countries to compare their approach with the WFD intercalibration.
- The species distribution models (EBV cubes) can be used to characterise the health of key habitats (e.g. EU Habitats Directive, EUNIS habitats) from characteristic species and their behaviour under varying scenarios. Thereby EBV cubes can contribute to more confident and comparable assessments of conservation status for the Habitats Directive.

4.6. Discussion

4.6.1. Advantages and caveats of the EBV result

- The production of species distribution models under varying scenarios (EBV cube) is a powerful tool for status assessment of both individual species and species assemblages. The communication of the *results and uncertainties* in geographical space (maps) is visually easier to process than results in environmental space using multivariate statistics (even for researchers).
- The species distribution models are built directly from continuous state variables (co-variates) and not lake typology, reducing modelling uncertainties.
- Stacking species distribution models (EBV cubes) integrates multiple national concepts in developing WFD indices (EQRs): Norway species *sensitivity*, Sweden species *optima* and Finland species *turnover*. Thus, EBV cubes can provide a more integrative inter-calibration procedure.
- The species distribution models are empirical models (based on the concept of realised niche) and cannot be applied beyond the spatial extent and environmental space. Scenarios of climate change may be applied to the northern species, but a more southward extent would be necessary to predict species response to warming in Fennoscandia. One alternative may be to develop species distribution models based on biological species traits to increase predictive abilities based on biological mechanisms, despite limitations in trait coverage and imputation methods.
- In principle, helophyte and bryophyte species available from Sweden and Finland could be incorporated to make predictions for Norway where additional fieldwork will be needed. This

This project receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101003553.

would be particularly relevant as both helophytes and bryophytes have different elevational and latitudinal gradients compared to submerged tracheophytes (vascular plants). However, there are current issues with the reliability of taxonomic identification, there being too few bryophyte experts. DNA extraction from collected bryophytes could be a solution.

4.6.2. Breakthroughs and lessons learned

- The first and most important breakthrough was to gather decades of monitoring efforts and learn from each other's across disciplines and generations. The curation and common understanding of the data are essential priors to any statistical analyses if we are to make sense of the results. This step is too often taken for granted by R users (many examples with macrophytes), but it is the single biggest hurdle in improving our understanding of species distribution.
- More resources are needed for data curation and data flow. Section 4 of the deliverable required 1.5 person months (PM) to gather, harmonise and quality control the data, 1 PM to workshop, reading and designing the approach and 1.5 PM to data analyses (writing scripts, producing and writing up results). That's about 100,000 €.
- The cost effectiveness could certainly be improved. Data curation and harvesting and statistical analyses could be automated (e.g. SLU portal, although additional data was made available by colleagues from SLU for this chapter). Species distribution models including ensemble models benefit from fine tuning (Valavi et al 2022), but as the number of species, cross-validation runs, and algorithms increases, then fully automated workflows become indispensable (e.g. biomod2, Thuiller et al 2019). Here we opted to use just two algorithms and mostly presented results from glm with coefficients easily interpreted in ecological terms (similar results were produced for gam and ensemble models). Since the species list of macrophytes is relatively short, fine tuning of models can run efficiently on a laptop computer with 16GB installed RAM (Intel® Core™ i5-6440HQ CPU @ 2.60GHz).
- The stacked species distribution models (EBV cubes) provide a powerful, flexible and integrative tool for the status assessment of species, habitats and ecosystems using the WFD concept of reference conditions.
- The representation of complicated statistical analyses through custom made maps allows to easily communicate the results and interact with stakeholders.

4.6.3. Outstanding challenges and proposed solutions

It remains to run the models on a larger range of species for status assessment at community level (species richness and turnover between current and reference conditions). It should not take too long now that semi-automated workflows are in place. Check the match between observed and predicted rank-abundance distribution.

The most outstanding challenge at European extent is to have sufficient resources to increase data coverage (the existence of field surveys), curation (national databases) and sharing (EU and global level). Lake surveys are expensive and time consuming, especially in remote locations. Thus, a European survey design to optimise data acquisition is important in space and time. The present study established species distribution models from a relatively long reference period (40 years) in three *data rich* countries. It is important to gather more homogeneous data through time and the WFD monitoring is of great help but does not cover all types of standing waters (e.g. temporary pools, ponds, oxbow lakes). Access to the raw species data (and associated co-variates available in WISE-6) is currently not supported by member states through the WFD workflows only harvesting community EQRs on a voluntary basis through WISE-2). Yet, those raw data could help the status assessment of habitats and ecosystems as required by other EU policies such as the Habitats Directive and EU 2030

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Biodiversity Strategy. In addition, raw species data could help estimate the effects of restoration such as pollution abatement programs (Nature Restoration law).

5. Essential Biodiversity Variables: Community structure (EQR)

5.1. EBV design characteristics

5.1.1. Community composition quantified as ecological quality ratio (EQR)

The EQR value quantifies the ecological condition (status) of the community composition for a freshwater biological quality element (BQE) required for assessing ecological status for that BQE. The EQR value indicates the deviation of the currently observed community from a reference condition given on a scale from zero (worst condition) and one (best condition). The reference condition describes the community composition of a natural community found in undisturbed rivers or lakes, and is specific for different types, depending on various typology descriptors, e.g. altitude, catchment size, geology (alkalinity, humic substances), mean depth. Most indices used for assessing EQRs are based on knowledge of environmental requirements of single species, including their occurrence or relative abundance along gradients of human pressure, such as nutrient pollution or organic pollution.

The EQR data are based on taxonomic analysis of species present in samples from river or lake stations used for WFD surveillance and/or operational monitoring (section 1.1.1). The single species occurrence or relative abundance and their indicator value along gradients of human pressure is combined for all species found in the sample to provide a community composition metric or index, wich can be used to assess the reference conditions in pristine water bodies and the current conditions in impacted water bodies as a basis for calculating the EQR values (Birk et al. 2012, Brucet et al. 2013) (Figure 20).

In the WISE-2 data flow (described below), information on the metric or index used by national reporters as the basis for EQR calculation (i.e., the left bar in Figure 20) can be provided in the field "ClassificationSystem" (http://dd.eionet.europa.eu/dataelements/latest/procedureClassificationSystem) for each individual EQR value. For example, an index for benthic invertebrates in rivers can be ASPT – Average Score Per Taxon (https://dd.eionet.europa.eu/vocabularyconcept/wise/ClassificationSystem/ASPT/view). The WISE-2 system does not contain further details about specific indices, but this information can still be useful as a pointer towards further information on the underlying species abundance data for each record of EQR values. For example, information on the species or higher taxa used in calculation of ASPT can be found in national classification guidances or in more general scientific literature.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

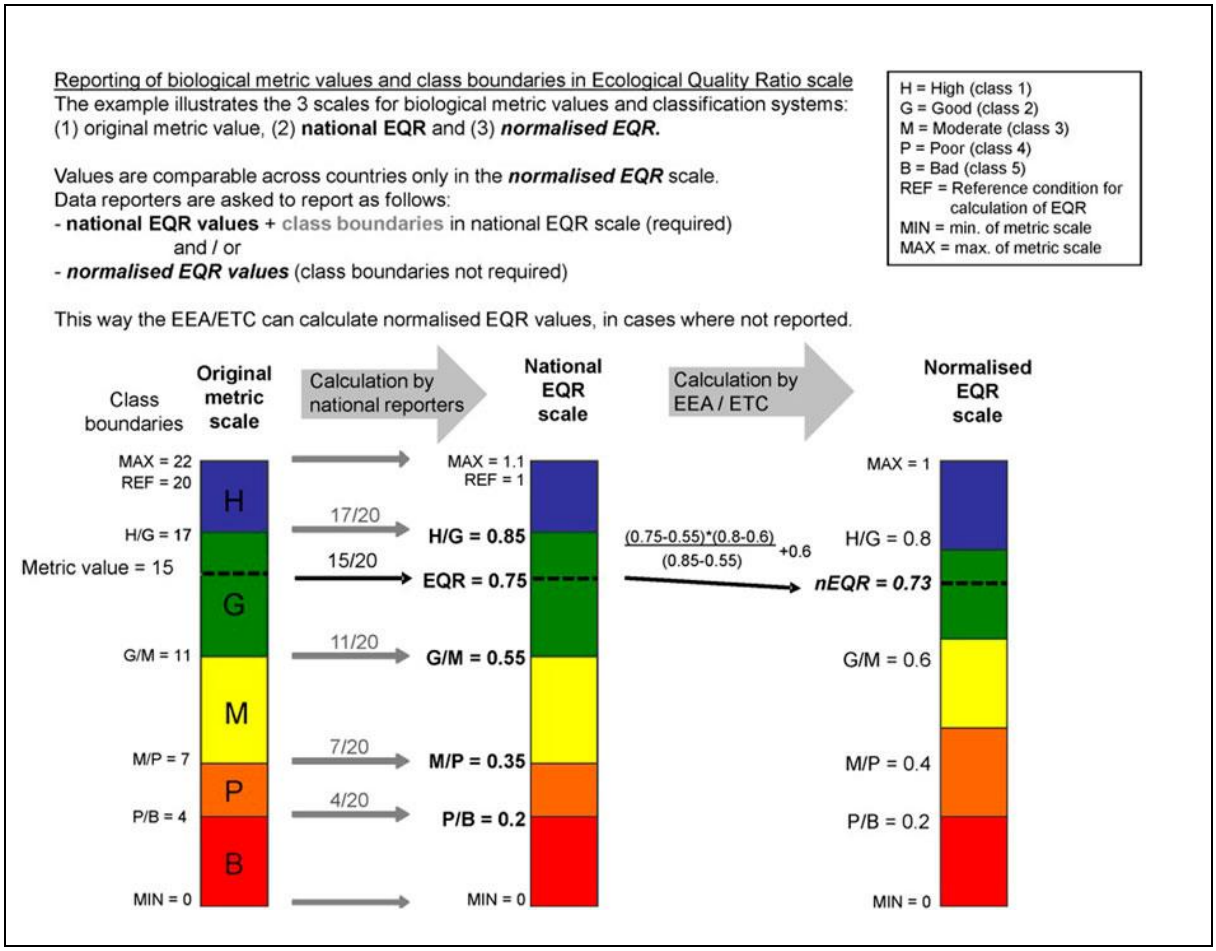


Figure 20. Example of calculation of normalised ecological quality ratios (EQRs) from national EQRs. Colour codes represent ecological status classes: blue = H(igh), green = G(ood), yellow = M(oderate), orange = P(oor), red = B(ad). REF = reference condition, defined as the metric value of unimpacted water bodies. Note that the maximum value of the metric scale can exceed the reference condition, therefore the national EQR value can exceed 1. Source: https://dd.eionet.europa.eu/datasets/latest/WISE-SoE_Biology/tables/BiologyEQRData.

If the underlying raw data at species level become accessible, they would support the species level EBV group 1 above as a basis for further species modelling at the European scale (Moe et al. 2023) An example of relevant species modelling is illustrated for lakes macrophyte species in Chapter 4. Essential Biodiversity Variables: Macrophyte species). An important activity in EuropaBON has therefore been to provide metadata for the individual (country-based) datasets underlying the EQR values in the EuropaBON database (Morán-Ordóñez et al. 2023), as illustrated in Figure 21. For this purpose, the EuropaBON project and database was presented at the EEA webinar for WISE-2 reporters in October 2022, followed by a questionnaire where we asked WISE-2 reporters for permission to store their names and contact information in the EuropaBON database. However, only a few of the reporters have responded positively to this request so far.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

(a) (b)

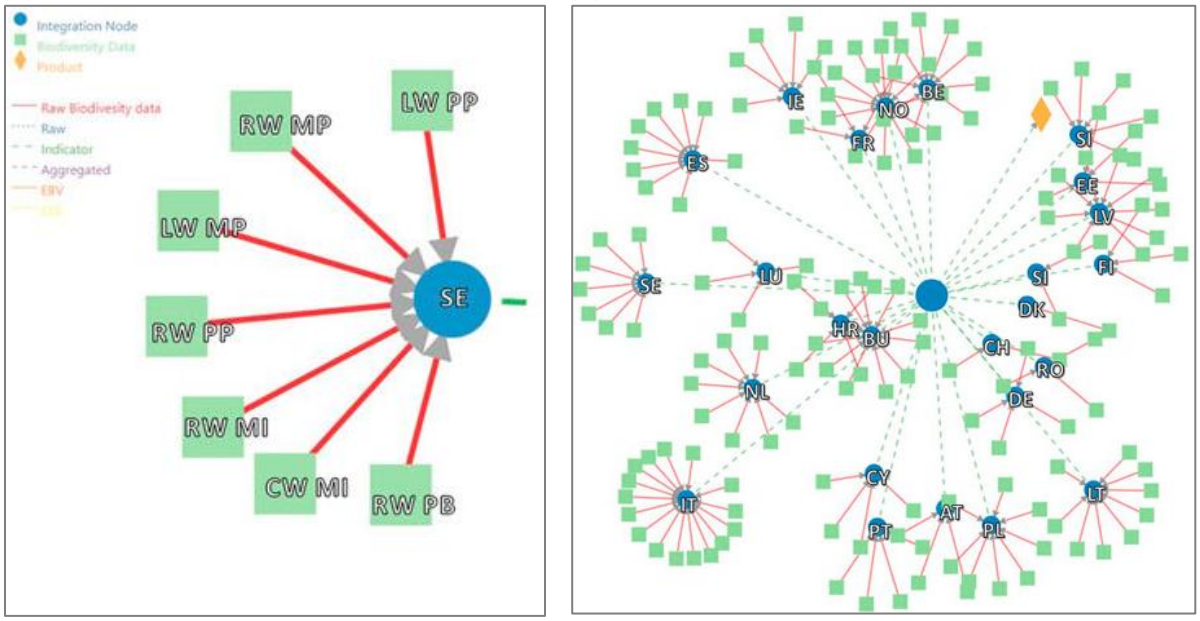


Figure 21. Representation of the WISE-2 dataflow in the EuropaBON database. (a) National EQR values are calculated (cf.

5.1.2. Community composition of benthic invertebrates in rivers

For benthic invertebrates in rivers, the EQR value is based on indices that combine the indicator value of each species (or higher taxonomic level) across all species found in a sample. The indicator value of each species or higher taxonomic level is based on its optimum occurrence along the gradient of organic pollution. As example, many stoneflies (*Plecoptera*) mainly occur in quite unpolluted rivers where oxygen conditions are good, while chironomids and oligochaetes mainly occur in large numbers in quite polluted environments due to their tolerance and adaptability to low oxygen concentrations.

5.1.3. Community composition of phytoplankton in lakes

For phytoplankton in lakes, both biomass and species composition are included in the EQR value by averaging the EQR values for biomass and for species composition. The latter is based on the indicator value of each species along the nutrient gradient in European lakes. As example, many chrysophytes mainly occur in quite pristine lakes with very low nutrient concentrations, while many chlorophytes and cyanobacteria mainly occur in quite eutrophied lakes with high nutrient concentrations.

5.1.4. Spatial and temporal resolution and extent, related to the intended use

The added value of the WISE-SoE biology data (WISE-2) in comparison to other main relevant data sources is illustrated schematically (Figure 22). Other data sources such as national monitoring data can provide even longer time series and higher resolution of information, such as abundance per species. However, compilation of raw species data is beyond the scope of EEA. While the WISE-2 data flow has a lower spatial coverage than the WFD reporting, the former has several benefits in the context of biodiversity information:

(1) More frequent reporting: annual WISE-2 data calls vs. WFD reporting every 6 years.
(2) Quantitative data: EQR (Ecological quality ratio) values on continuous scale (0-1) vs. categorical (5 status classes).
(3) More informative determinands: While both WFD and WISE-2 data represent the same biological item (BQE), the determinands of WISE-2 can be related to physical or chemical pressures and impacts (e.g., eutrophication vs. general degradation).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

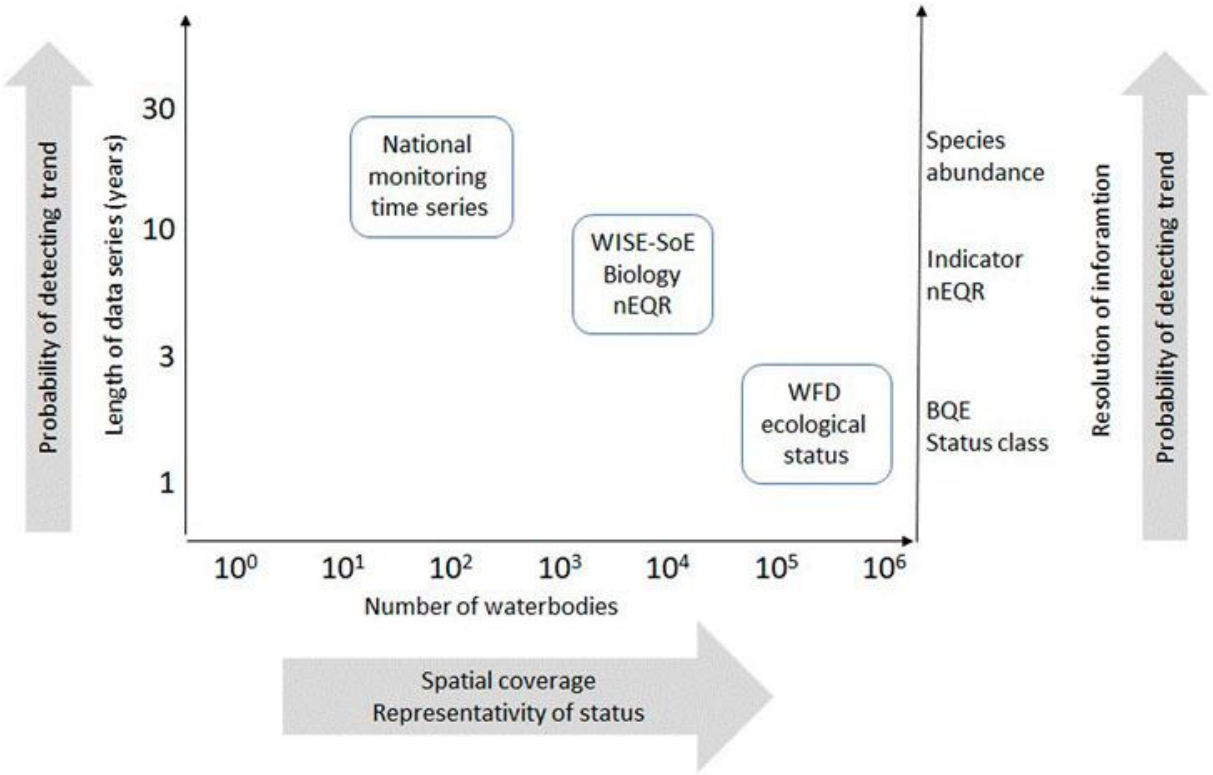


Figure 22. Schematic diagram of the spatial and temporal extent and resolution of WISE-SoE biology (WISE-2) data, compared to alternative monitoring data sources (national data and WISE-WFD data). Note: the position and extent of the text boxes do not represent exact values. After Moe et al. 2023.

Spatial resolution:
The spatial resolution of reporting is at monitoring site level (defined by point coordinates). Data processing such as temporal interpolation (gap-filling) and further assessments are done at the level of water body (defined by shape files), which is the lowest unit of spatial assessment in the WFD. All spatial information is obtained from the WFD database. (For countries not reporting to WFD, spatial information is obtained through EEA's WISE-5 data flow). For assessments by EEA/ETC, water bodies are typically aggregated to groups such as river basin district, country, geographic region, or water body type.

Temporal resolution:
The temporal resolution of the reported data is annually aggregated values. For phytoplankton in lakes, the annual means are based on monthly sampling during the growing season in the year of monitoring. The monitoring frequency for phytoplankton in lakes varies from every year in some lakes (operational monitoring) to once every 2-3 years (surveillance monitoring). For benthic invertebrates, the monitoring frequency is usually once or twice during the monitoring year, which is normally every 2-3 years. The timeseries are kept by interpolation between the observations reported for up to 3 year gaps.

Extent:
The EQR-EBV dataset covers the monitoring years 2009-2021 so far. Most of the data are reported after 2015, which is therefore used as the starting point for trend analyses.
The spatial extent covers most of Europe, but there is a gap in the middle, as France stopped reporting in 2013 and Germany has not reported EQR values at all due to internal dataflow challenges from the federal states to the national level. The German central UBA is now working to solve that problem. The maps shown in Figure 23 and Figure 24 gives an idea about the spatiotemporal extent for the period 2016-2021.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

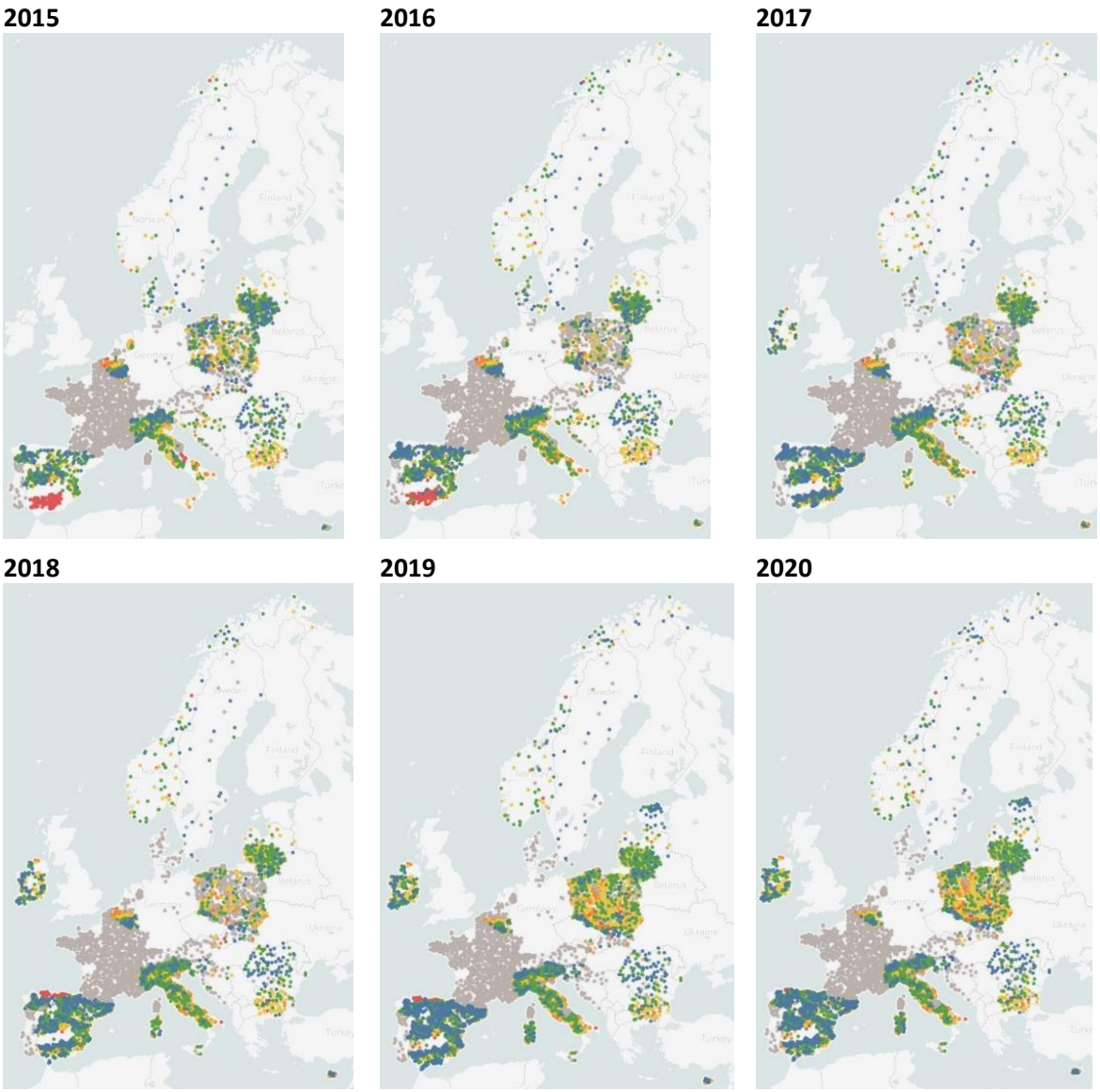


Figure 23. Spatiotemporal extent of the EQR data represented by status class for benthic invertebrates in rivers for the period 2015-2020. (https://public.tableau.com/app/profile/sophie.menztel/viz/Improved_BiologyDataByWaterBody_SOM/ClassValuebyWaterBody-Map?publish=yes) The online version of the maps is interactive and the user can select any year and any BQE and water category. Colour code for status classes: blue = high, green = good, yellow = moderate, orange = poor, red = bad. Grey dots mean not reported for the year shown, but reported in another year.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

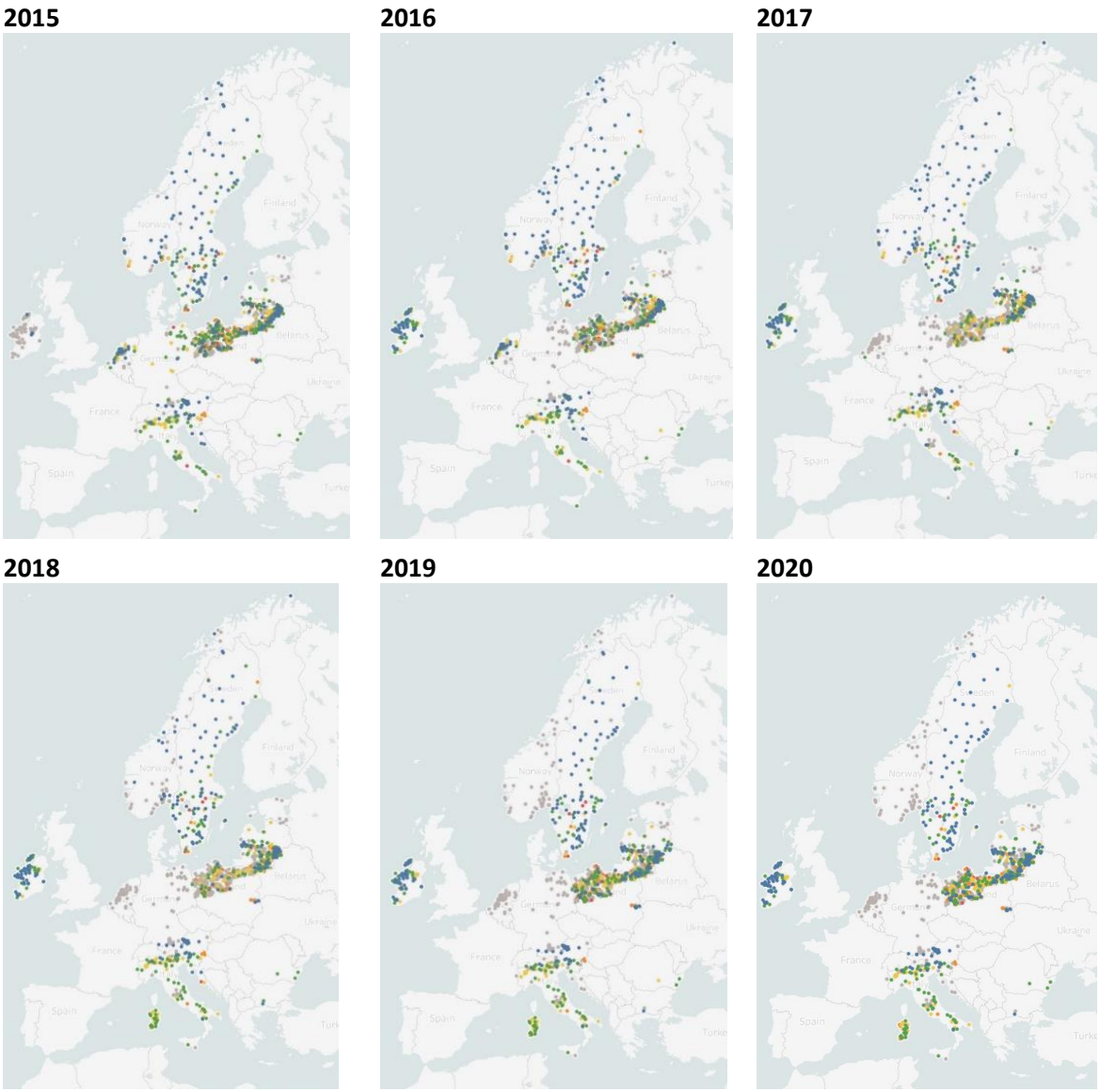


Figure 24. Spatiotemporal extent of the EQR data represented by status class for phytoplankton in lakes for the period 2015-2020. For more details see Figure 23.

5.2. Input biodiversity data

5.2.1. EBV-usable data: overall description of biodiversity data sources

The WISE-2 Biology dataflow was established by EEA around year 2010 to obtain a harmonised flow of biology data reported as yearly aggregated Ecological Quality Ratios (EQRs) from all surface water categories; rivers, lakes, transitional and coastal waters (https://cdr.eionet.europa.eu/help/WISE_SoE/wise2).

As mentioned elsewhere, the WISE-2 data flow does not include raw species abundance data, but information associated with the national datasets can potentially provide a link to the underlying raw data.

Biology data can be reported to WISE-2 as normalised EQR values (Figure 20) or as national EQR value together with information on the national classification system. The EEA WISE-2 database (Table 8) currently contains data from different phases of biology data reporting:

- Reporting years 2011 - 2013: biology in rivers and lakes reported to WISE Rivers and WISE Lakes
- Reporting years 2015 - 2019: biology in rivers and lakes reported to WISE-4 dataflow
- Reporting years 2019 - 2022: biology in all water categories reported to WISE-2 dataflow

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

(WISE-2 database does not contain biology data from transitional and coastal waters reported prior to 2019.)

The biology data can be reported either at the level of monitoring sites or at the level of water bodies (as shown in the data model; https://dd.eionet.europa.eu/datasets/latest/WISE-SoE_Biology#model). The majority of the data are reported for monitoring sites.

5.2.2. Data quality and assurance

The reported data undergo four steps of quality checking (QC) by EEA/ETC (see also Figure 25):
1st level: Automatic QC in Reportnet3, using a set of more than 100 rules in SQL. These include tests of data type, internal consistency and consistency with previously reported data, logical rules, etc. The QC rules can result in four different validation types (blocker, error, warning, info); blockers will prevent the dataset from being released (i.e. delivered).
2nd level: Algorithms and/or expert assessments resulting in final feedback from EEA to countries after the dataset release (focusing on errors and warnings), to allow for corrections and re-release of the dataset.
3rd level: Additional automised test run after the harvest of released datasets: statistical outliers etc.
4th level: Additional expert-based QC performed by content experts (visual checks, distributions etc., particularly during development of indicators and visualisations), resulting in quality statement flags for individual records published together with the data.

5.2.3. Data handling

The reported data are processed into so-called “statistics tables” (formerly called “indicator tables”) with the following main steps.

1) Aggregation of EQR values from monitoring sites to water body level (in SQL)
2) Interpolation of EQR values for missing years (gap-filling of up to three years) (in SQL)
3) Identification of consistent time series (no missing years) after interpolation, for given periods (e.g. 2015-2021)

The consistent time series are further used for analyses such as explorative analyses of aggregated time series (Figure 26) and testing of statistical trends per water body (Figure 28 and Figure 28).

This project receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101003553.

Table 8. Raw data availability and access

DATASET TITLE	Waterbase – Biology (Prod-ID: DAT-241-en Published 07 Jul 2023 Last modified 07 Jul 2023)
Raw data collection design	The WISE-2 datasets contain aggregated metric values (ecological quality ratio). Information on the underlying raw data is given in the EuropaBON database, under Project WISE-2 (where possible).
Monitoring programs	National monitoring programs based on the Water Framework Directive (WFD) and national classification systems for ecological status of water bodies. More details are given in the EuropaBON database, under Project WISE-2 (where possible).
Types of data access	Public access to published data through two portals: 1) Waterbase - Biology (for download): https://www.eea.europa.eu/data-and-maps/data/waterbase-biology-1/waterbase-biology-4-tables 2) WISE statistics – biology (for download or API): https://discodata.eea.europa.eu/
Data repositories	Data repository until 2023: EEA's Central Data Repository (CDR) – Eionet Reporting Obligation Database - Deliveries for WISE SoE - Biology data (WISE-2) (https://rod.eionet.europa.eu/obligations/630/deliveries) (From 2023, the data flow has moved to Reportnet 3: Reportnet 3 (europa.eu))
Persistent identifier(s)	Identifiers are given in the EEA Geospatial metadata catalogue: Waterbase - Biology (europa.eu)
Metadata description	Metadata catalogue: https://sdi.eea.europa.eu/catalogue/srv/eng/catalog.search#/metadata/7bd881f0-59ca-4f0a-a43c-7e0cd0c66d6c
Other provenance information	A processed version of Watebase Biology is published from 2023 onwards: European Environment Agency (2023). WISE statistics - biology (2009-2021), 2023. WISE statistics - biology (2009-2021), 2023 (europa.eu) This version includes temporal interpolation and other types of gap-filling.

5.3. The EBV model and workflow

5.3.1. Overview of the WISE-2 Biology data flow

The main steps of the WISE-2 data flow are illustrated in Figure 25. The steps most important in the context of biodiversity data are described in more detail below. Boxes with green text represent data processing carried out within EEA (harvesting of reported data, quality checking, further processing, etc.) Boxes with purple text represent publication of WISE-2 data in public databases available for download (Waterbase, see Table 8) or in extended data flows through an application programming interface (API; Discodata). The boxes with blue text represent data analyses and assessment carried out by ETC currently (solid outlines) or proposed for future tasks (dotted outlines).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

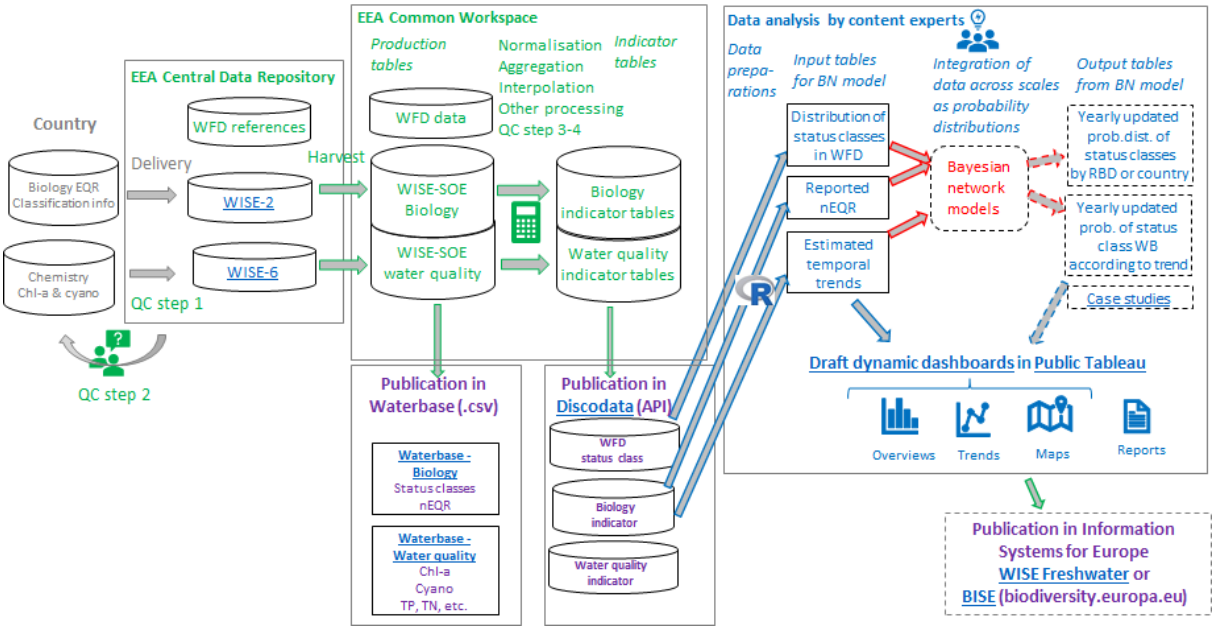


Figure 25. Main steps of the WISE-2 Biology data: flow from reporting of national datasets to analysis of status and trends at the European level. Boxes with dashed outline represent proposed but not yet performed activities.

5.3.2. Calculation and processing of normalised EQR values

All countries reporting to WFD have developed national classification systems for assessment of ecological status based on EQR values. However, the boundaries defined between status classes (e.g., Good/Moderate) in national EQR scale vary among countries, as well as among determinands and water body types (Birk et al., 2012). For example, the Good/Moderate boundary of a given BQE could be 0.62 in one country and 0.66 in another case; an EQR value of 0.64 would then mean Good status in the former case but Moderate status in the latter case. This means that national EQR values are not directly comparable between countries, without considering the distance to their respective class boundaries (Birk et al., 2013). Therefore, the national EQR values are being normalised to a scale of 0–1, in order to obtain a consistent scale across all countries (Figure 20). On the normalised scale, the status class boundaries are identical for all countries (high: 0.8–1.0, good: 0.6–0.8, moderate: 0.4–0.6, poor: 0.2–0.4, bad: 0.0–0.2). An EQR value identical to the boundary between two status classes belongs to the worse of the two status classes.

The conversion from national to normalised EQR requires the information on class boundaries (in EQR scale) which can be specific for the determinand and for the national water body type, as well as for natural vs. artificial and highly modified water bodies. The normalised EQR are calculated using the formula:

$$\begin{aligned}\text{normalisedEQR} &= (\text{EQR} - \text{LowerBoundaryEQR}) \times 0.2/(\text{UpperBoundaryEQR} - \text{LowerBoundaryEQR}) \\ &\quad + \text{LowerBoundaryNormEQRnormalisedEQR} \\ &= (\text{EQR} - \text{LowerBoundaryEQR}) \times 0.2/(\text{UpperBoundaryEQR} - \text{LowerBoundaryEQR}) \\ &\quad + \text{LowerBoundaryNormEQR}\end{aligned}$$

where *LowerBoundaryEQR* and *UpperBoundaryEQR* are the lower and upper status class boundaries in the national EQR scale, respectively, the factor 0.2 is the width of any status class at the normalised EQR scale, and LowerBoundaryNormEQR is the lower class boundary in the normalised EQR scale. The calculation can be performed by the WISE-2 reporters before reporting, or by EEA after reported if sufficient information is provided on the national classification system.

5.3.3. Interpolation

To maximise the number of so-called indicator values for aggregated time series plots and for trend analysis, missing yearly values were replaced with imputed values as far as possible. Gaps of up to 3

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

years within a data series have been interpolated as the average of the previous and following years. Likewise, gaps of up to 3 years at the beginning or end of a series were extrapolated to be identical as the first or last available value. This procedure follows the methodology used for the EEA indicator "Nutrients in Freshwater", as described in its Supplementary Material section (https://www.eea.europa.eu/ims/nutrients-in-freshwater-in-europe).

The interpolation procedure follows the routines described for EEA's freshwater indicators (https://www.eea.europa.eu/en/analysis/indicators/nutrients-in-freshwater-in-europe), under Supporting Information:
"For time series analyses, only complete series after inter/extrapolation are used. This is to ensure that the aggregated time series are consistent, i.e. including the same sites throughout. Inter/extrapolation of gaps up to three years are allowed, to increase the number of available time series. At the beginning or end of the data series missing values are replaced by the first or last value of the original data series, respectively. In the middle of the data series, missing values are linearly interpolated."

5.3.4. Spatial reference data

The data reported under the WISE-2 flow are linked to a monitoring site (or to a water body, for data aggregated by water body), using an existing spatial identifier. The location, identification and characteristics of the spatial objects must have been provided via the WFD Spatial data flow (or WISE-5 data flow, for countries that do not report under the Water Framework Directive). During the quality control procedures of the data reported under WISE-2, the consistency of these identifiers are checked against the spatial data already reported. The spatial data includes coordinates, shape files, water category, national water body type, and whether the water body is natural, artificial or heavily modified.

5.3.5. Covariates

The water quality data reported under WISE-6 (https://cdr.eionet.europa.eu/help/WISE_SoE/wise6/) can provide relevant covariates for the WISE-2 biology data. Relevant chemical variables include nutrients (phosphate, total phosphorus, and nitrate), which can be expected to influence primary producers, and ammonium biological oxygen demand, which can be expected to influence benthic invertebrates. A preliminary study of these covariates is described further in the Section Outstanding challenges and proposed solutions.

5.3.6. Trend analysis

Mann-Kendall trend test is used to evaluate time trends (magnitude of change in nEQR per year) and the distribution of trends for the different BQEs. This is a non-parametric test for a monotonic trend over time. The magnitude of change is estimated by the Sen's slope. Trends were calculated per water body, and both observed and inter-/extrapolated data were included, a method which is also used in other EEA indicators. Including inter- and extrapolated data reduces variance and inflates the p-values, but using only observed data would have reduced the number of usable time series and complicated the assessment of overall trends. For this analysis, the direction of the trends and magnitude of the Sen's slopes is most important, while the exact p-values are less important. In Figure 28, trends are defined as "significant" when $p < 0,05$ and as "marginally significant" when $p = 0.05\text{-}0.1$. The quantitative change of nEQR values are shown as histograms. A t-test was done to test whether the distribution of slopes was significantly different from 0, i.e. whether there on average was an increase or decrease in nEQR over time.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 23/05/2024. DOI: https://doi.org/10.3897/arphapreprints.e128109

5.4. Performance of EBV-derived policy indicators

5.4.1. Methods and outputs for producing EBV-derived state indicators and trends

Currently, annual data allowing trend analysis has been reported from 14 countries for benthic invertebrates in rivers and 10 countries for phytoplankton in lakes. To test for differences in trends between WBs with different ecological status, we grouped all WBs according to the nEQR value in the starting year (2015). WBs with nEQR ≤ 0.4 in 2015 was classified as "poor or bad", WBs with 0.4 < nEQR ≤ 0.6 as "moderate" and WBs with nEQR > 0.6 as "good or high" (Figure 26).

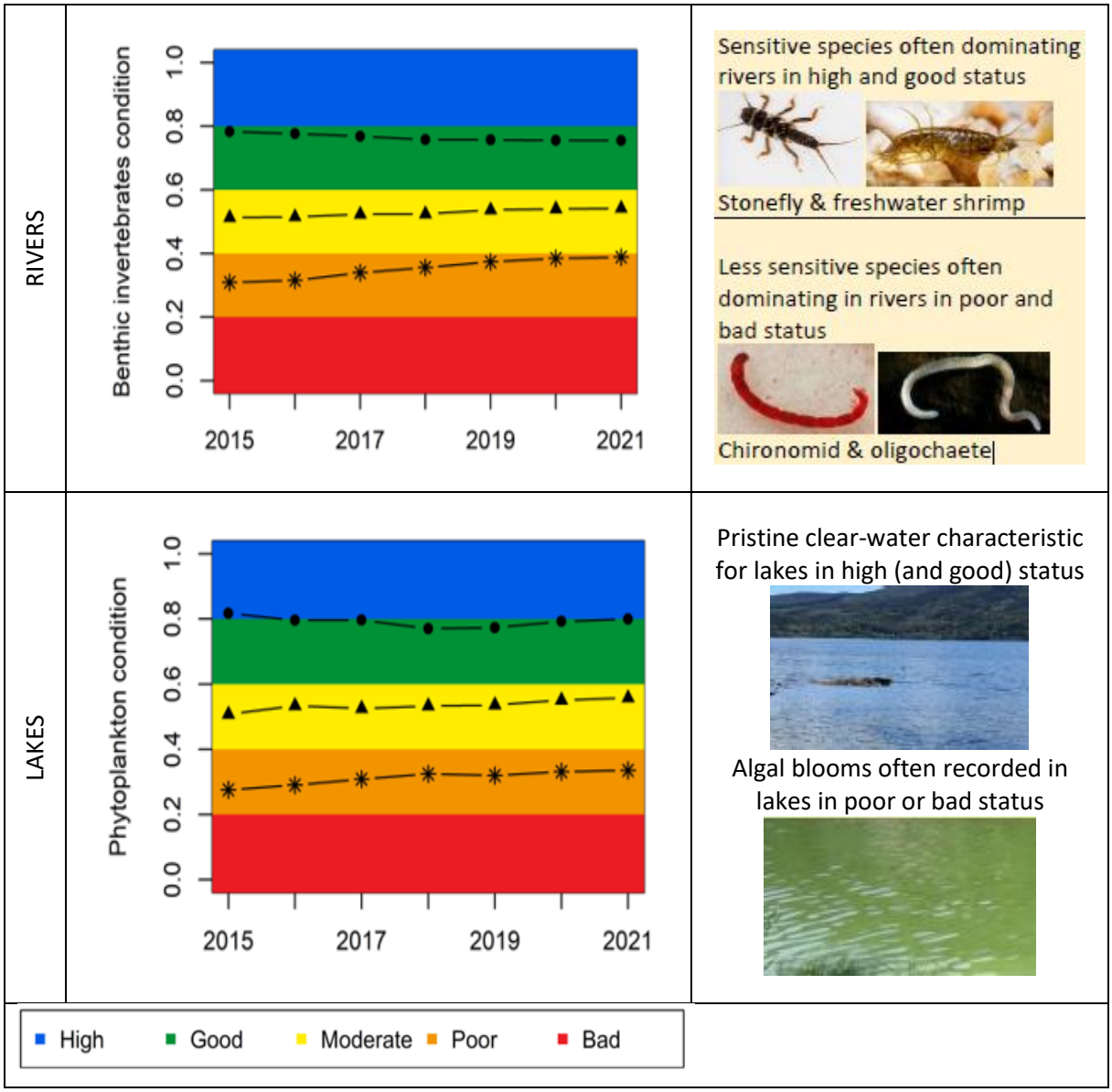


Figure 26. Time series plots: Development of ecological condition for benthic invertebrates in European rivers (upper plot) and phytoplankton in European lakes (lower plot) during the period 2015-2021. The lines with symbols show the mean value of the normalised ecological quality ratio (nEQR) for all consistent time series in three groups of water bodies with nEQR values within high or good condition (dots), moderate condition (triangles) and poor or bad condition (stars) in 2015.

Trends are summarized in Figure 27 for the water bodies in high or good condition in 2015 showing slightly more water bodies with deteriorating (negative) trends than with improving (positive) trends, while for water bodies in poor or bad condition in 2015 the trend results were opposite showing more water bodies with improving (positive) trends than with deteriorating (negative) trends.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

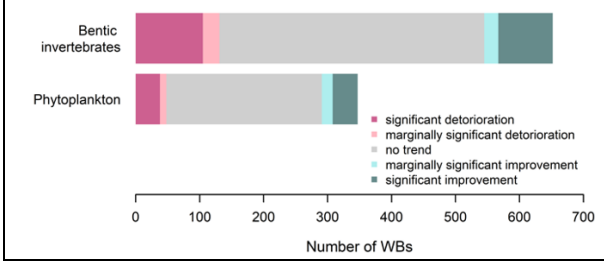
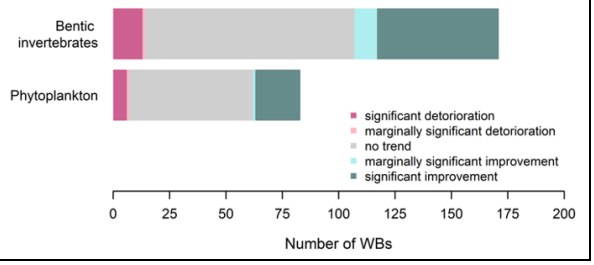
Trends for water bodies with complete time series and nEQR values corresponding to **high or good condition** in 2015	Trend for water bodies with complete time series and nEQR values corresponding to **poor or bad condition** in 2015
	

Figure 27. Trend analysis of the time-series data for EBVs based on EQR-data. Graphs with red, grey and green colours show the number of water bodies (x-axis) for benthic invertebrates in rivers (upper bar plots) and phytoplankton in lakes (lower bar plots) having significant, marginally significant or no trends based on Mann-Kendall trend analysis for each individual water body where consistent time series are identified. Statistical significance of the trends are defined as follows: $p < 0.05$ means significant; $0.05 \leq p < 0.1$ means marginally significant; $p \geq 0.1$ means not significant.

The amount of change in EQR-values during the period 2015-2021 can also be shown as frequency plots, using zero as no change, giving the number of water bodies for each change interval (corresponding to 0.02 EQR-units) (Figure 28). This figure illustrate that most water bodies change less than 0.05 EQR-units during this period, corresponding to one quarter of the width of one status class. So small changes would not be visible if only the status class were used, as very few water bodies move to another status class during this period.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

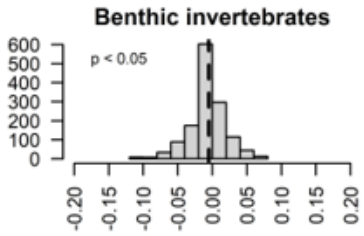
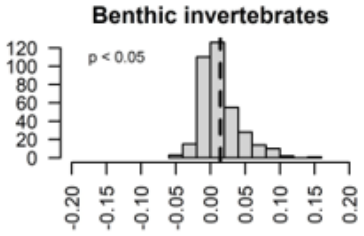
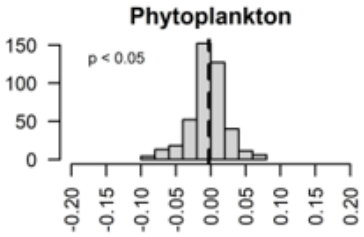
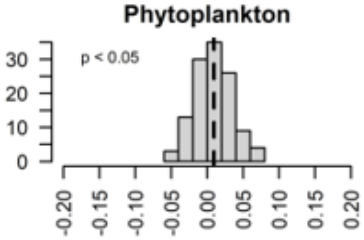
EBV	Water bodies with complete time series and nEQR values corresponding to **high or good condition** in 2015	Water bodies with complete time series and nEQR values corresponding to **poor or bad condition** in 2015
Benthic invertebrates in rivers		
Phytoplankton in lakes		

Figure 28. Frequency distribution of water bodies in intervals of change in nEQR values from 2015-2021. Negative values indicate deterioration of the conditions, while positive values indicate improvement of the conditions. Values of 0.20 (or -0.20) corresponds to a whole condition class change, while 0.10 (or -0.10) corresponds to a half condition class change. y-axis shows number of water bodies; x-axis shows change in nEQR values from 2015-2021. Dashed lines show median change for all the water bodies within the group included in each plot.

5.4.2. Relevance of the new EBV product for policy reporting as compared to customarily used data flows

Results indicate that most of the reported water bodies that were in poor or bad ecological condition in 2015 show improvements towards better condition during the period from 2015-2021 (Figure 26-Figure 28). Observed improvements can reflect effects of management measures, such as reduction of nutrients and organic pollution in European lakes and rivers. These improvements indicate that the natural biodiversity is slowly recovering in the water bodies with improving trends, although opportunistic or tolerant species may still dominate.

For water bodies that were initially in high or good condition in 2015, the results indicate on average a small deterioration for all these biological quality elements (Figure 26-Figure 28). However, longer time series are needed to assess whether or not these patterns will continue in the years to come.

Assessment of biological quality elements (BQEs) condition allows us to observe trends before a water body changes condition from one class to another. This is needed to guide regular management efforts and inform about the effectiveness of management. The observed small deterioration of areas in high and good condition can be an early warning to management that measures should also be considered to maintain high or good condition of water bodies under potential impacts of climate change.

5.4.3. Representation of the resulting EBV cube

The WISE statistics tables – Biology (EEA 2023) contain values representing community structure with spatiotemporal dimensions, and can therefore be considered a step towards developed EBV cubes.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 23/05/2024. DOI: https://doi.org/10.3897/arphapreprints.e128109

However, the spatial unit of EBV cubes should be raster cells, while the WISE-2 statistics data have water bodies (defined by shape files) as spatial unit. An additional step of spatial data processing is therefore needed to convert these data into proper EBV cubes. However, a grid with a defined grid cell size can be superimposed on the current water body status map, see Figure 29).

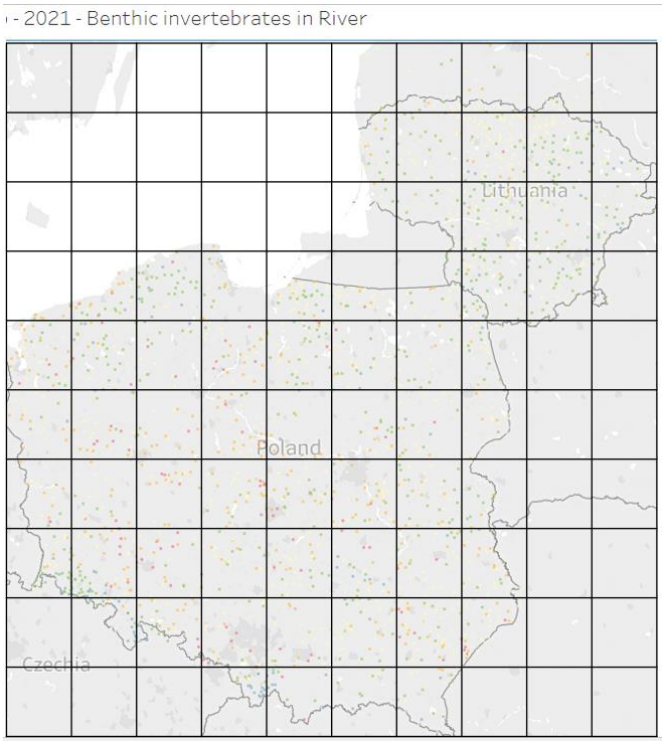


Figure 29. Map of ecological condition for benthic invertebrates in rivers based on EQR-values zoomed in on a data-rich region of Poland and Lithuania. The online version of this map is interactive (Tableau software). See maps in Figure 23 for all Europe overview. The grid is added on top of the non-gridded map just for illustration.

5.4.4. **Uncertainty assessment**

There has not yet been any evaluation of model performance related to WISE-2 data. So far there are two types of models that could have been evaluated: (1) the interpolation of nEQR values for missing years could have been evaluated if such values were obtained, (2) the estimated trends could be used to project nEQR values for future years, which could then be validated with later reportings. The spatial patterns of EQR values could also be compared against the reported WFD status classes, although a one-to-one relationship should not be expected, since the WFD status class can be based on more data than the EQR value. However, the representativeness is quite good for both the BQEs presented above (see section 5.4.5 Representativeness assessment).

5.4.5. **Representativeness assessment**

A preliminary assessment of representativity was done for the EBVs on benthic invertebrates in rivers and phytoplankton in lakes by comparing the status reported for each of the two BQEs in the WFD database for the 2nd RBMPs (Figure 30). Ideally, this comparison should have been done with the WFD data for the 3rd RBMPs, which cover the years shown for the timeseries and trends in (Figure 26-Figure 28). However, the BQE-level data are not yet available due to delayed reporting by many countries.
The comparison shows that the EQR-based status class data have a slightly higher proportion of water bodies in good or high status for both the BQEs. The difference can be due to spatial data gaps in the EQR-data, which is based on monitoring data and/or to different methods used for assessing WFD status per BQE, including not only monitored water bodies, but also grouping and expert judgement.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

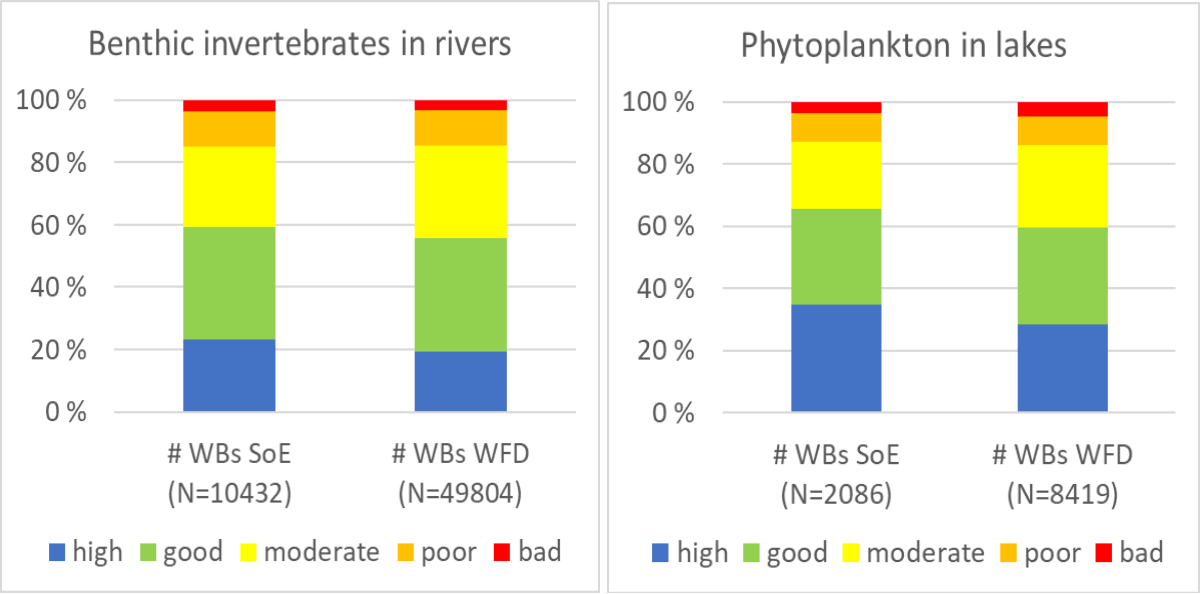


Figure 30. Comparison of the relative distribution of status classes (high, good, moderate, poor and bad) reported for water bodies to the WFD database for the 2nd RBMPs (WBs WFD) (right columns in both plots) and the EQR-based status class data reported to the WISE-2 SoE data, which is used for EBVs (WBs SoE) (left columns in both plots). The total number of water bodies reported for the two datasets are given in parentheses.

5.4.6. Impact of spatial and temporal biodiversity data gaps on the reliability of assessment

The spatial gaps of the EQR-EBVs are considerable, as some countries do not report data at all, e.g. DE, while others have stopped reporting after 2012, e.g. FR or after 2020, e.g. UK (due to Brexit). The interactive maps represented by pictures in Figure 23-Figure 24 also illustrate that the density of monitoring sites varies a lot between countries. This would suggest that the reliability of assessment in terms of geographic coverage is quite poor. However, when comparing the distribution of ecological status classes between those reported for the WFD mandatory reporting with those calculated from the voluntary reporting of EQR values, there are only minor differences between those two datasets, in spite of much less data reported with EQR values than for the WFD status class reporting (Figure 30). This would suggest that the EQR data are quite representative in terms of the overall distribution of ecological quality, and that the reliability is acceptable.

The temporal data gaps also considerable. Annual data reporting is encouraged by EEA, but since the WISE-SoE data flows are voluntary, not all countries prioritise WISE-2 reporting every year. EEA uses the expression “consistent time series” about data series (EQR values for individual BQEs) that are complete (no missing years) after temporal gap-filling (interpolation/extrapolation) and spatial aggregation (from monitoring site to water body level). Consistent time series are defined as either “short” (currently 7 years, monitoring years 2015-2021) or “long” (monitoring years 2011-2021); if not specified the “consistent time series” refer to the shorter version. The consistent time series therefore only represent a small fraction (3-4 %) of all water bodies in Europe, so far. Nevertheless, given the fact that these time series are complete(d) and are able to detect significant temporal trends in indicators of ecological quality (Figure 26-Figure 28), the temporal assessments can be considered reliable based on this selection of the WISE-2 dataset.

Regarding the further use of WISE-2 biology data, the most recent attempts within the ETC was exploratory analyses to link the BQE EQR values as responses to abiotic explanatory variables (an ETC BE deliverable in December 2023). The selected predictor variables for EQR values of a given water body were water quality represented by nutrients and organic pollutants (from WISE-6) as well as land cover in the catchment (from CORINE, provided by the ETC partner TC Vode) of the same water body.

This project receives funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024.* DOI: https://doi.org/10.3897/arphapreprints.e128109

Regression tree analysis was applied to identify the most significant predictor variables, as well as threshold values along gradients of the predictor variables that caused significant changes (branches) in the distribution of the response variable. An example is shown in Figure 31, where the EQR values of benthic invertebrates in rivers show significant responses to both agricultural and urban land cover. More specifically, this example suggests that declines in ecological quality of benthic invertebrates is caused primarily by agricultural land cover (lower EQR for water bodies with >34% agriculture in the catchment) and secondarily in urban land cover. In low-agriculture areas (<34%), urban land cover exceeding 1% results in ecological quality decreasing from High-Good to Good; while in high-agriculture areas (>34%), urban land cover exceeding 10% results in ecological quality decreasing from Good-Moderate to Moderate-Poor. (For the link from normalised EQR values to status classes, see Figure 20).

This approach illustrates that spatial gaps may be reduced by using land-use variables as a coarse predictor for ranges of EQR-values.

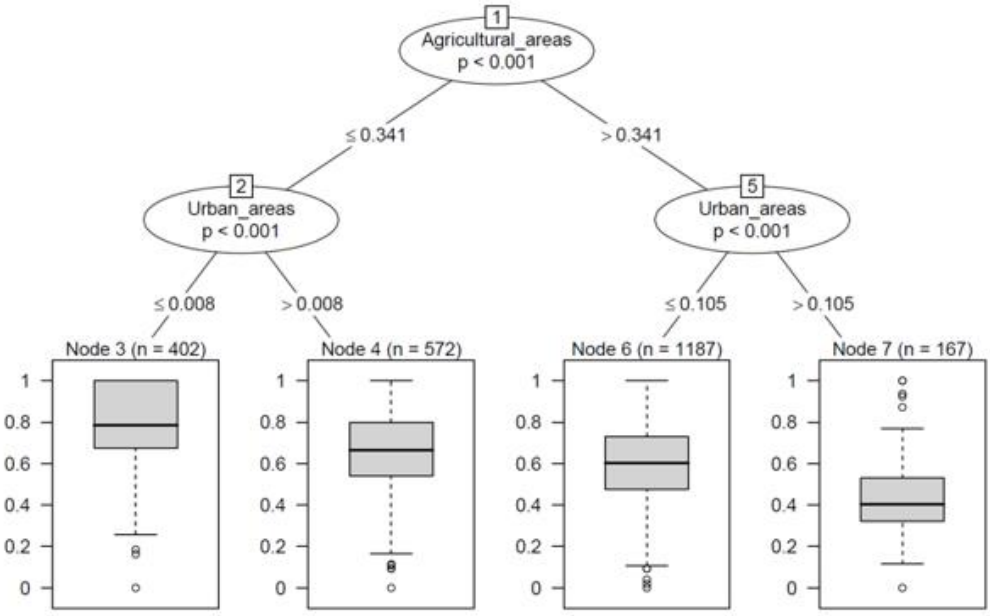


Figure 31. Results from regression tree analysis grouping the nEQR for benthic invertebrates in rivers (mean nEQR from 2015-2021, y-axis in boxplots) according to the fraction of agricultural areas and urban areas.

5.5. Discussion

5.5.1. Advantages and caveats of the EBV result

- **Advantages compared to mandatory WFD status class data**

  The community structure EBVs based on EQR-data offer a more precise estimation of the current condition for each of the WFD-BQEs as it is shown on a continuous scale from 0-1, in contrast with the coarser categorical WFD-BQE status class data. Another important advantage of the EBVs based on EQR-data is that these data are reported annually, while the WFD status class data are reported only once every 6 years (Figure 22). Both the continuous scale and the higher temporal resolution of the EBVs based on EQR-data allow trends to be assessed much earlier than what is possible with the WFD data, and they can also show the direction and amount of change for all water bodies with consistent time-series (Figure 26, Figure 27, Figure 28). These EBVs can therefore be more important for water managers than the WFD status class data, as they can quantify effects of costly measures to reduce human pressures on the water bodies also within a status class, and thereby motivate for further action to restore the biodiversity in rivers and lakes.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

- **Advantages compared to national species-level data:**
  The community structure EBVs based on EQR-data are more available than the national species-level data, as the WISE-2 dataflow is well established at the EU level (Figure 25). These EBVs also integrate single species data to show the current condition for the whole biological community. Moreover, the EBVs are mostly based on harmonised field sampling procedures and harmonised taxonomy, as well as agreed lists of indicator species responding to various human pressures thanks to a number of WFD-related projects (https://freshwaterecology.info/). Furthermore, the EQR-based EBVs represent community level indices with intercalibrated high/good and the good/moderate boundaries across countries within each region, thereby also showing comparable deviations from reference conditions (Buffagni et al., 2001; Phillips et al. 2013) . They thereby offer comparable data across countries, and can also be used to show results for different common types of water bodies (e.g. lowland, highland, large, small, siliceous, calcareous) based on their spatial location, whenever needed (Lyche Solheim et al. 2019).

- **Caveats for community structure EBVs based on EQR-values**
  The EQR EBVs only show the deviation from undisturbed communities, but cannot show the actual underlying species composition. These species level data are only stored in national or sub-national databases. However, there are numerous papers available to show how indicator species are identified, ranging from taxa that are highly sensitive to a human pressure to highly insensitive (tolerant or opportunistic) taxa for each of the BQEs, e.g. Penning et al. 2008, Phillips et al. 2013. Thus, EQR-values close to 1.0 will represent water bodies with most of the highly sensitive taxa but few of the tolerant taxa, while EQR-values < 0.4 will represent strongly degraded water bodies with mainly tolerant taxa.
- **Caveats related to geographic gaps in Europe**
  The main caveat of the current EQR EBVs are the geographical gaps, which is illustrated in Figure 23 and Figure 24. A recent test of the representativity of the EU-level EQR data reported to WISE-2 show, however, that the distribution of the main ecological status classes is quite similar in the WFD-dataset from the 2nd RBMPs and the WISE-2 dataset (Figure 30). However, getting the France and Germany on board the WISE-2 reporting is essential to get a better coverage across Europe for these EBVs in the coming years. Land-use variables can also be used to reduce these gaps, although the uncertainty in EQR-ranges predicted by these variables needs further considerations.
- **Caveats related to temporal interpolation**
  The interpolation of data gaps in the timeseries allow trends to be analysed for more water bodies, but at the same time introduce uncertainty. However, this interpolation is needed as most BQEs are not monitored annually, but only once every 2-3 years. This is also a general policy used by EEA for all their core set indicators, after evaluation of advantages and disadvantages.

5.5.2. Breakthroughs and lessons learned

While the experiences reported in this section are based on work done by NIVA within the ETC, the breakthroughs and lessons learned mentioned here are considered relevant for the synergies with EuropaBON and future work building upon EuropaBON.

The development of the aquatic biology data flow (now labelled WISE-2) has since the first preparations in 2007-2008 been tightly coupled to and harmonised with the abiotic water quality data flow (now labelled WISE-6), and subsequently with the WFD reporting. The WISE-SoE data flows are continuously being more automatised and efficient, as a means to handle increasing amounts of data with limited resources. This implies a trade-off for the biology data (Figure 22): to keep pace with the increasing harmonisation and automatization of data processing at EEA, it was necessary to sacrifice information content and resolution, handling of special cases, etc.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

A main strength of the WISE-2 Biology data flow is that it is fully integrated with the WISE-SoE data flows, and therefore included in all steps from national reporting to publication in Waterbase and Discodata (Figure 25). This means that the data reporting is very cost-efficient in terms of making use of other available information. For example, all spatial and related information on water bodies can be obtained from the WFD database, and will not require additional reporting or quality checking. Moreover, the automatised procedure for interpolation of up to three years means that reported data can be used more efficiently in trend analysis. For example, even if a water body is monitored only every two or three years, the resulting interpolated data series can still be included in the analysis of consistent time series, after the minimum number of actual observations has been obtained.

Eventually, this automatised workflow can also facilitate the further steps such as visualisation of data in online dynamic dashboards (currently in draft state) if these will be approved by EEA for publication. The dynamic dashboards can be used for two main purposes: (1) inspection of the reported data both by EEA/ETC and the national reporters (e.g. quality, quantity, temporal and spatial coverage), and (2) assessment of status and trends (e.g. geographic patterns of increasing or decreasing EQR values). Feedback from reporters suggest that these (draft) tools are appreciated and motivating for continued data reporting, and should be further developed.

A main drawback of WISE-2 is the constraints imposed on the data flow by the harmonisation to WFD spatial information. For example, all water bodies that are not reported to WFD (e.g. lakes with surface area below 0.5 km^2), will also be excluded from the WISE-2 reporting, even if biology data from such lakes have been reported previously. Moreover, there are many situations where the national classification system is more complex than the WISE-2 data dictionary (e.g. dependent on more variants of water body types than those reported to WFD). In such cases, it will not be possible for countries to report the information on classification systems required for EEA/ETC to calculate normalized EQR values from national EQR values. An alternative for such cases is for countries to report the normalised EQR directly without the classification system, but this option means that EEA/ETC get less possibility for quality checking.

As mentioned, WISE-2 Biology data flow has been developed within the frame of WISE-SoE through 15 years and compromised on details to obtain efficiency. This also means that some of the evolved structures are not optimal, but kept for historical reasons. For example, the biology data flow was for some years split into freshwater (rivers and lakes) and marine (transitional and coastal), and had slightly diverging developments before they were merged into the current WISE-2. During this period the freshwater biology determinant were further developed into impact-specific types (e.g. PhytoplanktonEQR_E for phytoplankton impacted by eutrophication), while the transitional/coastal determinands were unchanged (e.g. PhytoplanktonEQR). This difference is not based on biology or policy needs, but still maintained for reasons of consistency and traceability with the older data. Such structures are typically no longer visible after in the published biology data after the final processing, but they can be confusing for data reporters, and can require additional guidance and support to reporting.

NIVA has been responsible for thematic and technical support to EEA on the WISE-2 data flow since the initiation, and NIVA's knowledge of this background and development is therefore also crucial for the continued development and extended use of this data flow.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

5.5.3. Outstanding challenges and proposed solutions

The main outstanding challenges for the WISE-2 biology data flow can be grouped into three issues, considering the synergy with EuropaBON and projects focusing on biodiversity data, e.g. the upcoming EP pilot project for establishing the BMCC/EBOC.

1) Access to the underlying species abundance data.
Compilation of single species abundance or occurence data underlying the EQR values has never been and will most likely never be a goal for EEA. Instead, through the EuropaBON database, we have tried to provide metadata and other information for each individual dataset (Figure 21), including contact information for data reporters who might share more information on the underlying species data. Only a few of the reporters have given approval so far. However, this effort might be repeated in 2024 (if approved by EEA) and possibly result in contact information for more countries. A step towards facilitating reporting or providing access to the underlying species data is likely to be taken in the EP pilot project for establishing the BMCC/EBOC. In parallel, the data owners could be considered to report their underlying species data to GBIF. Access to such data would enable modelling of the status of biological communities using the approaches described in the chapter on macrophytes species modelling (chapter 4).

2) Recent changes in the reporting infrastructure.
The whole set of WISE-SoE data flows have been transferred from the old system (Central Data Repository) to the current system (Reportnet 3) during 2022-2023; with WISE-2 (Biology) and WISE-6 (Water quality) transferred only from August 2023. The change in reporting system has caused many technical challenges and additional work for both EEA/ETC and data reporters and may cause a (temporary) decline in the biology data reporting in 2023. However, the new reporting system should be more intuitive and will hopefully increase the reporting amount and efficiency in the longer term.

3) Further automatization of assessments.
While most steps of the biology data quality checking and processing have been automized within EEA's common workspace, there are still certain steps in the analysis and assessment carried out by ETC in other software. For example, temporal trend analysis is currently carried out in the statistical programming language R. This step precludes the inclusion of trend analyses in dynamic dashboards produced by EEA, whose work flow should be fully contained within the common workspace. Adaptation of such calculation and analyses to the common workspace (possibly in SQL) will be a next step towards further automatization and efficiency of the biology data flow, including the visualisation of trends in EQR values representing community-level EBVs.

4) Further use of the WISE-2 Biology data in environmental assessments
The analysis of WISE-2 Biology data in combination with environmental explanatory variables was only started in autumn 2023 within the ETC BE task 1.2.1.6 (Drivers of change: assessment of land use change in river catchments). The initial exploratory analysis showed some interesting patterns (exemplified in Figure 31), but no follow-up activities are currently planned within ETC. Nevertheless, we have identified a great potential for further analysis of the EQR values in response to environmental conditions. Such integrated assessments can primarily incorporate other data sources that are already harmonised with the WISE-2 data flow (WISE-6 data on water quality data and WFD data on ecological status; Figure 22), but can also be extended to incorporate for example other types of land-use information or climate information (cf. Moe et al. 2024). Probabilistic (Bayesian) modelling methods have been suggested for this purpose (see Figure 25) and have proved useful for integrated assessments in other field such as ecological risk assessment (Moe et al. 2022).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

6. Conclusions

6.1. Macrophytes species distribution modelling

Summary of the work done (8 person months):

- A novel georeferenced macrophyte database of 2079 Fennoscandian lakes with 109 species including aquatic vascular plants (excluding helophytes) and charophytes was compiled together with lake surface area, water quality (notably alkalinity, nutrients, water colour) and climate data (annual sum of degree days).
- Species distribution models using glm, gam and ensemble models (and their uncertainties) were produced from 1081 lakes with a full dataset for two species with contrasting niche reflecting different habitats (siliceous and calcareous lakes).
- A subset of reference lakes was used to derive lake specific reference conditions based on total phosphorus (TP), i.e. free of lake typology.
- Species status assessment was derived from species distribution projections (spatial EBV cube) comparing current with reference conditions. A general formula for species lake status assessment was also derived from species distribution projections under the species optimum, current and reference conditions (for a given pressure variable, here TP).
- Species distributions were projected at 3693 lakes in Fennoscandia using glm models applied to the 1995 north European lake water chemistry surveys.

Notable achievements for stakeholders and EU policies:

- The species distribution model approach allows to stack species (EBV cubes) to form predicted communities on which biodiversity indices may be derived (richness, turnover) for reference and current conditions as well as alternative scenarios, hindcasting and forecasting. EBV cubes can therefore be useful for assessing effects of restoration measures under the Nature Restoration Law and WFD.
- The use of species distribution models provides an alternative to the intercalibration approach of the WFD and integrate all conceptual approaches: Norway community indices based on species sensitivity (the slope in our models), Sweden index based on species phosphorus optima (the optima and tolerance in our models) and Finland index based on species turnover (can be derived from stacking our species distribution model)
- The availability of data and transparency of our approach (data and R scripts provided in Section 4.3) remove two major stumbling blocks for ecological status assessment of waterbodies, in sharp contrast to the WFD Macrophyte Northern GIG *intercalibration* exercise which did not report the data or the calculations and had no geographical display. Our approach could make it possible for anyone to compare their approach with the WFD intercalibration.
- The species distribution models (EBV cubes) can be used to characterise the health of key habitats (e.g. EU Habitats Directive, EUNIS habitats) from characteristic species and their behaviour under varying scenarios. Thereby EBV cubes can contribute to more confident and comparable assessments of conservation status for the Habitats Directive.

Further challenges remaining:

- The species distribution models are empirical models (based on the concept of realised niche) and cannot be applied beyond the spatial extent and environmental space. Scenarios of climate change may be applied to the northern species, but a more southward extent would be necessary to predict species response to warming in Fennoscandia. One alternative may be to develop species distribution models based on biological species traits to increase predictive abilities based on biological mechanisms, despite limitations in trait coverage and imputation methods.
- Lake shore plants (helophyte) and bryophyte species were available from Sweden and Finland. Taxonomic skills for bryophytes are currently limited and the identification of bryophytes remains

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

generally poor. Further identification based on DNA could improve recording effort and accuracy. This is particularly relevant for conservation as both helophytes and bryophytes have different elevational and latitudinal gradients compared to submerged vascular plants.

Breakthroughs and lessons learned:

- The first and most important breakthrough was to gather decades of monitoring efforts and learn from each other's across disciplines and generations. The curation and common understanding of the data are essential priors to any statistical analyses, if we are to make sense of the results.
- The cost effectiveness could certainly be improved. Data curation and harvesting and statistical analyses needs more resources but could be automated to a larger extent
- The stacked species distribution models (EBV cubes) provide a powerful, flexible and integrative tool for the status assessment of species, habitats and ecosystems using the WFD concept of reference conditions.
- The representation of complicated statistical analyses through custom made maps allows to easily communicate the results and interact with stakeholders.

6.2. Community structure EBVs

Summary of the work done (4 person months):

- The community structure EBVs have been identified and described in terms of the rationale and objective, the monitoring done to create the underlying datasets and the use of the datasets to calculate the EQRs
- The data from each country reporting EQR-values to EEA have been reported to the EuropaBON metadatabase, and the dataflow has been described from the country level to the EU level
- The use of the data to create maps and trends have been explained and illustrated with figures
- The link to land use data have been explored and show clear connections with negative relationships between the EQR-values and the share of agricultural and urban areas in the catchments of the water bodies.

Notable achievements for stakeholders and EU policies:

- The EQR-based EBVs allow trends to be shown for the development of the condition of each BQE required by the WFD. The trends can be shown within single ecological status classes, which quantifies effects of measures to reduce human pressures on Europe's waters.
- The dataflow is fully developed by EEA/ETC-BE, including a detailed data dictionary to support data providers with the annual reporting. EEA also organises annual webinars with data providers, in which the use of the data is also shown in order to motivate them for further reporting
- The EQR EBVs can also support assessments of the ecological status and function components of the conservation status under the Habitats Directive for habitat types that match the WFD broad types for rivers and lakes.

Further challenges remaining:

- Keep the annual dataflow running in the coming years, including the EEA and ETC-BE support staff and the annual webinars with data providers. In 2023, the EEA stopped the process of endorsement for the EQR-based indicators, due to a high-level decision to not establish any new indicators (at least for the freshwater and marine realms) mainly due to financial & staff capacity limitations. There is also a discussion in EEA that the frequency of processing and updating existing indicators could be changed from annual to once every two or three years to save resources. It is unclear whether they will still continue with their annual data request. That discussion is linked to the proposed revision of the WFD to report monitoring data used for status class assessment annually. This proposal is currently discussed in the EU Council, but it is unclear what is meant with

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

monitoring data, as this can be anything from the raw data (which is less likely to be agreed), the community structure metrics data, the EQR-values or only the status class for single BQEs and supporting QEs. Therefore, the future of the EQR-based EBVs is highly uncertain.

- Identify the main bottlenecks preventing Germany and France to report EQR-values to WISE-2, as solving those bottlenecks would close most of the geographical/spatial gaps in the current dataset. In Germany, there is a process initiated to renegotiate the agreement between the federal states and the national UBA to report EQR-data to EEA. These data are currently not accessible for the central UBA, which is the national focal point for Germany in EIONET.
- To demonstrate more widely (through the BMCC/EBOC) the usefulness of the EQR-based community structure EBVs to capture effects of costly measures to reduce pressures on Europe's waters and restore their biodiversity.

Breakthroughs and lessons learned:

- The community structure EBVs based on EQR-values have been demonstrated as powerful trend indicators, allowing early trend detection and quantification of change for each BQE in each single water body where EQR-values have been reported. The EBVs can therefore capture impacts of pressures and measures on the condition of freshwater biodiversity.
- The EBVs have also been successfully linked to land-use. The preliminary results of regression tree analysis show a clear decrease in EQR-values with increasing proportion of agricultural and urban areas in the catchments of the water bodies. This would allow further model development to close spatial gaps.
- The EBVs can also be used as indicators for ecosystem structure and function within the conservation status assessments of the Habitats Directive for all types of water bodies that match the habitat types listed for inland waters in the HD Annex 1.
- It is essential to communicate these results to stakeholders responsible for water and habitat management at national and regional level to motivate further data reporting in the coming years.

7. References

7.1. References for the main introductory chapters and for the community structure EBVs based on EQR-values

Birk, S., Chapman, D., Carvalho, L., Spears, B.M., Estrup Andersen, H., Argillier, C., Auer, S., Baattrup-Pedersen, A., Banin, L, Beklioğlu, M., Bondar-Kunze, E., Borja, A., Branco, P., Bucak, T., Buijse, A. D., Cardoso, A.C., Couture, R.M., Cremona, F., de Zwart, D., Feld, C. K., Ferreira, M. T., Feuchtmayr, H., Gessner, M.O., Gieswein, A., Globevnik, L., Graeber, D., Graf, W., Gutiérrez-Cánovas, C., Hanganu, J., Işkın, U., Järvinen, M., Jeppesen, E., Kotamäki, N., Kuijper, M., Lemm, J. U., Lu, S., Lyche Solheim, A., Mischke, U., Moe, S.J., Nõges, P., Nõges, T., Ormerod, S. J., Panagopoulos, Y., Phillips, G., Posthuma, L., Pouso, S., Prudhomme, C., Rankinen, K., Rasmussen, J. J., Richardson, J., Sagouis, A., Santos, J. M., Schäfer, R.B., Schinegger, R., Schmutz, S., Schneider, S.C., Schülting, L., Segurado, P, Stefanidis, K., Sures, B., Thackeray, S. J., Turunen, J., Uyarra, M. C., Venohr, M., von der Ohe, P. K., Willby, N. and Hering, D., 2020. Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems. Nature Ecol Evol.: https://doi.org/10.1038/s41559-020-1216-4

Birk, S., W. Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, A. Solimini, W. van de Bund, N. Zampoukas & D. Hering, 2012. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18: 31–41. https://doi.org/10.1016/j.ecolind.2011.10.009

Brucet, S., S. Poikane, A. Lyche Solheim & S. Birk, 2013. Biological assessment methods for European lakes: ecological rationale and human impacts. Freshwater Biology 58. 1106–1115, https://doi.org/10.1111/fwb.12111 .

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Buffagni, A., Kemp, J., Erba, S., Belfiore, C., Hering, D., & Moog, O. (2001). A Europe-wide system for assessing the quality of rivers using macroinvertebrates: The AQEM project and its importance for south-ern Europe (with special emphasis on Italy). Journal of Limnology, 60 (Suppl. 1), 39–48. https://doi.org/10.4081/jlimnol.2001.s1.39

Geijzendorffer, I.R., Regan, E.C., Pereira, H.M., Brotons, L., Brummitt, N., Gavish, Y., Haase, P., Martin, C.S., Mihoub, J.B., Secades, C., Schmeller, D.S., Stoll, S., Wetzel, F.T., Walters, M., 2016. Model-assisted monitoring of biodiversity: Bridging the gap between biodiversity data and policy reporting needs: An Essential Biodiversity Variables perspective. J. Appl. Ecol. 53: 1341-1350. https://doi.org/10.1111/1365-2664.12417

Junker, J., Pedro Beja, Lluís Brotons, Miguel Fernandez, Néstor Fernández, W. Daniel Kissling, Maria Lumbierres, Anne Lyche Solheim, Joachim Maes, Alejandra Morán-Ordóñez, Francisco Moreira, Martin Musche, Joana Santana, Jose Valdez, Henrique Pereira, 2023. D4.1. List and specifications of EBVs and EESVs for a European wide biodiversity observation network. https://doi.org/10.3897/arphapreprints.e102530

Kissling, W. D., Walls, R., Bowser, A., Jones, M.O., Kattge, J., Agosti, D., Amengual, J., Basset, A., van Bodegom, P. M., Cornelissen, J. H. C., Denny, E.G., Deudero, S., Egloff, W., Elmendorf, S.C., García, E.A., Jones, K.D., Jones, O.R., Lavorel, S., Lear, D., Navarro, L.M., Pawar,S., Pirzl, R., Rüger, N., Sal, S., … Guralnick, R.P., 2018. Towards global data products of Essential Biodiversity Variables on species traits. Nature Ecol. Evol. 2: 1531-1540. https://doi.org/10.1038/s41559-018-0667-3

Lyche Solheim, A., C.K. Feld, S. Birk, G. Phillips, L. Carvalho, G. Morabito, U. Mischke, N. Willby, M. Søndergaard, S. Hellsten, A. Kolada, M. Mjelde, J. Böhmer, O. Miler, M. T. Pusch, C. Argillier, E. Jeppesen, T.L. Lauridsen, S. Poikane, 2013. Ecological status assessment of European lakes: a comparison of metrics for phytoplankton, macrophytes, benthic invertebrates and fish. Hydrobiologia, 704, p. 57-74, https://doi.org/10.1007/s10750-012-1436-y .

Lyche Solheim, A., Persson, J., Austnes, K., Moe, J., Kampa, E., Stein, U., Feher, J., Poikane, S., Kristensen, P., 2015. Figure 30European Freshwater Ecosystem Assessment: Cross-walk between the Water Framework Directive and Habitats Directive types, status and pressures. ETC-ICM-report, Prague, 176 pp: http://icm.eionet.europa.eu/ETC_Reports/FreshwaterEcosystemAssessmentReport_201509/Freshwater_Ecosystem_Assessment_Report_for_publication_04_09_2015_final.pdf

Lyche Solheim, A., Globevnik, L., Austnes, K., Kristensen, P., Moe, J., Persson, J., Phillips, G., Poikane, S., van de Bund, W., Birk, S., 2019. A new broad typology for rivers and lakes in Europe: Development and application for large-scale environmental assessments. Science of the Total Environment, 697: https://doi.org/10.1016/j.scitotenv.2019.134043

Lyche Solheim, A., Austnes, K., Birk, S., 2020. ETC/ICM Report 4/2020: Comparison of ecological status between countries and river basin management plans (RBMP) cycles — Eionet Portal (europa.eu)

Moe, S.J., Benestad, R.E. and Landis, W.G. (2022), Robust risk assessments require probabilistic approaches. Integr Environ Assess Manag, 18: 1133-1134. https://doi.org/10.1002/ieam.4660

Moe, S.J., Mentzel, S., Welch, S.A., Lyche Solheim, A., 2023. From national monitoring to transnational indicators: reporting and processing of aquatic biology data under the European Environment Agency's State of the Environment data flow. Frontiers in Environmental Science 11: https://doi.org/10.3389/fenvs.2023.1057742

Moe, S.J., Brix, K.V., Landis, W.G., Stauber, J.L., Carriger, J.F., Hader, J.D., Kunimitsu, T., Mentzel, S., Nathan, R., Noyes, P.D., Oldenkamp, R., Rohr, J.R., van den Brink, P.J., Verheyen, J. and Benestad, R.E. (2024), Integrating climate model projections into environmental risk assessment: A probabilistic modelling approach. Integr Environ Assess Manag, 20: 367-383. https://doi.org/10.1002/ieam.4879

Morán-Ordóñez, A., Pino, D.M., Brotons, L., 2021. D3.1 Inventory of current European network for monitoring. Web-based database. Arpha preprints: https://doi.org/10.3897/arphapreprints.e109168

Morán-Ordóñez, A., Beja, P., Fraixedas, S., Herrando, S., Junker, J., Kissling, W.D., Lumbierres, M., Lyche Solheim, A., Miret, G., Moe, J., Moreira, F., Pereira, H., Santana, J., Villero, D., Brotons, L., 2023.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

D3.3 Identification of current monitoring workflows and Bottlenecks. Project Report. Arpha Preprints: https://doi.org/10.3897/arphapreprints.e103765

Phillips, G., Lyche-Solheim, A., Skjelbred, B., Mischke, U., Drakare S., Free G., Järvinen M., de Hoyos C., Morabito G., Poikane S., Carvalho L. 2013. A phytoplankton trophic index to assess the status of lakes for the Water Framework Directive. Hydrobiologia 704, p. 75-95. https://doi.org/10.1007/s10750-012-1390-8 .

Santana, J., Porto, M., Brotons, L, Junker, J., Kissling, W.D., Lumbierres, M., Moe, J., Morán-Ordóñez, A., Pereira, H., Lyche Solheim, A., Villero, D., Moreira, F., Beja, P., 2023. D3.2 Report on gaps and important new areas for monitoring in Europe. Arpha Preprints: https://doi.org/10.3897/arphapreprints.e103657

7.2. References for the Macrophytes species level EBV:

Andersson B. (1999) Vattenvegetation. Bedömningsgrunder för miljökvalitet. Sjöar och vattendrag. Bakgrundsrapport 2. Biologiska parametrar. Rapport SNV 4921.

Bjornstad O.N. (2022) ncf: Spatial Covariance Functions. R package version 1.3-2, https://CRAN.R-project.org/package=ncf.

Broennimann O., Di Cola V. & Guisan A. (2023) ecospat: Spatial Ecology Miscellaneous Methods_. R package version 4.0.0, https://CRAN.R-project.org/package=ecospat.

Cornes R.C., Van Der Schrier G., Van Den Besselaar E.J.M. & Jones P.D. (2018) An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research-Atmospheres,* **123,** 9391-9409.

Demars B.O.L. & Trémolières M. (2009) Aquatic macrophytes as bioindicators of carbon dioxide in groundwater fed rivers. *Science of the Total Environment,* **407,** 4752-4763.

Direktoratsgruppa Vanndirektivet. (2018) Veileder 02:2018. Klassifisering av miljøtilstand i vann. Økologisk og kjemisk klassifiseringssystem for kystvann, grunnvann, innsjøer og elver. www.vannportalen.no

Ecke F. (2007) Bedömningsgrunder för makrofyter i sjöar - Bakgrundsrapport [Water Quality Element of macrophytes in lakes - background report]. Luleå University of Technology. Research Report 2007:17. Luleå.

Ecke F., Hellsten S., Mjelde M., Kuoppala M. & Schlacke S. (2010) Potential conflicts between environmental legislation and conservation exemplified by aquatic macrophytes. *Hydrobiologia,* **656**, 107-115.

Ecke F. (2014) Vattenväxter i sjöar Utvärdering av det gemensamma delprogrammet [Aquatic plants in lakes - Evaluation of the joint sub-programme]. Länsstyrelsen Stockholm [Stockholm County Administrative Board]. Stockholm.

Ecke F. (2015) Makrofyter i sjöar [Macrophytes in lakes]. Havs- och vattenmyndigheten [Swedish Agency for Marine and Water Management]. Version 3:0, 2015-06-26. Göteborg.

Elith J., Ferrier S., Huettmann F. & Leathwick J. (2005) The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling,* **186,** 280-289.

Folster J., Johnson R.K., Futter M.N. & Wilander A. (2014) The Swedish monitoring of surface waters: 50 years of adaptive monitoring. *Ambio,* **43,** 3-18.

Guisan A., Thuiller W. & Zimmerman N.E. (2017) *Habitat Suitability and Distribution Models with Applications in R,* Cambridge University Press, Cambridge.

Hellsten S., Willby N., Ecke F., Mjelde M., Phillips G. & Tierney D. (2014) Northern lake macrophyte ecological assessment methods. European Commission, Joint Research Centre, Institute for Environmental Sustainability. Ispra.

Henriksen A., Skjelkvaale B.L., Mannio J., Wilander A., Harriman R., Curtis C., Jensen J.P., Fjeld E. & Moiseenko T. (1998) Northern European Lake Survey, 1995 - Finland, Norway, Sweden, Denmark, Russian Kola, Russian Karelia, Scotland and Wales. *Ambio,* **27,** 80-91.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Hijmans R. (2023) raster: Geographic Data Analysis and Modeling. R package version 3.6-26, https://CRAN.R-project.org/package=raster.

Hong D.Y. (1991) A biosystematic study on Ranunculus subgenus Batrachium in S Sweden. *Nordic Journal of Botany,* **11,** 41-59.

Irvine K. (2009) Harmonizing assessment of conservation with that of ecological quality: fitting a square peg into a round hole? *Aquatic Conservation-Marine and Freshwater Ecosystems*, **19**, 365-369.

Johnson R.K. & Toprak V. (2021) Local habitat is a strong determinant of spatial and temporal patterns of macrophyte diversity and composition in boreal lakes. *Freshwater Biology,* **66,** 1490-1501.

Jongman R.H.G., Ter Braak C.J.F. & Van Tongeren O.F.R. (1995) *Data analysis in community and landscape ecology,* Cambridge University Press, Cambridge.

Kassambara A. (2023) ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'. R package version 0.1.4.1, https://CRAN.R-project.org/package=ggcorrplot.

Kolada A., Willby N., Dudley B., Noges P., Sondergaard M., Hellsten S., Mjelde M., Penning E., Van Geest G., Bertrin V., Ecke F., Maeemets H. & Karus K. (2014) The applicability of macrophyte compositional metrics for assessing eutrophication in European lakes. *Ecological Indicators,* **45,** 407-415.

Kuoppala M., Hellsten S. & Kanninen A. (2008) Sisävesien vesikasviseurantojen laadunvarmennus [Quality assurance of aquatic plant monitoring in inland waters]. Suomen Ympäristökeskus [Finnish Environmental Centre]. Helsinki.

Legendre P. & Legendre L. (1998) *Numerical Ecology,* Elsevier Science, Amsterdam.

Lohammar G. (1938) Wasserchemie und höhere Vegetation schwedischer Seen. *Symbolae Botanicae Upsalienses,* **3,** 1-252.

Lundh A. (1951a) Studies on the vegetation and hydrochemistry of Scanian lakes. I. Higher aquatic vegetation. *Botaniska Notiser,* **Supplement 2,** 5-141.

Lundh A. (1951b) Studies on the vegetation and hydrochemistry of Scanian lakes. II. Ion determinations in lake waters. *Botaniska Notiser,* **Supplement 2,** 143-174.

Mjelde M., Thrane J.E. & Demars B.O.L. (2023) High aquatic macrophyte diversity in Norwegian lakes north of the Arctic Circle. *Freshwater Biology,* **68,** 509–522.

Pebesma E. (2018) Simple features for R: standardized support for spatial vector data. *The R Journal,* **10,** 439-446.

Pebesma E.J. (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences,* **30,** 683-691.

Perrin S.W., Van Der Veen B., Golding N. & Finstad A.G. (2022) Modelling temperature-driven changes in species associations across freshwater communities. *Global Change Biology,* **28,** 86-97.

R Core Team. (2023) *R: A language and environment for statistical computing*. http://www.R-project.org/, 31-10-2023

Roy D.B., Hill M.O., Rothery P. & Bunce R.G.H. (2000) Ecological indicator values of British species: an application of Gaussian logistic regression. *Annales Botanici Fennici,* **37,** 219-226.

Tammi J., Appelberg M., Beier U., Hesthagen T., Lappalainen A. & Rask M. (2003) Fish status survey of Nordic lakes: Effects of acidification, eutrophication and stocking activity on present fish species composition. *Ambio,* **32,** 98-105.

Ter Braak C.J.F. & Barendregt A. (1986) Weighted averaging of species indicator values: its efficiency in environmental calibration. *Mathematical Bioscience,* **78,** 57-72.

Thuiller W., Georges D., Gueguen M., Engler R., Breiner F., Lafourcade B. & Patin R. (2023) biomod2: Ensemble Platform for Species Distribution Modeling. R package version 4.2-4, https://CRAN.R-project.org/package=biomod2.

Thuiller W., Guéguen M., Renaud J., Karger D.N. & Zimmermann N.E. (2019) Uncertainty in ensembles of global biodiversity scenarios. *Nature communications,* **10,** 1446.

Valavi R., Guillera-Arroita G., Lahoz-Monfort J.J. & Elith J. (2022) Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs,* **92,** e01486.

Wickham H. (2016) *ggplot2: Elegant Graphics for Data Analysis,* New York, Springer-Verlag.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Wood S.N. (2003) Thin plate regression splines. *Journal of the Royal Statistical Society Series B-Statistical Methodology,* **65,** 95-114.

Wood S.N. (2017) *Generalized Additive Models: An Introduction with R,* Chapman and Hall/CRC, Boca Raton, Florida.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 23/05/2024*. DOI: https://doi.org/10.3897/arphapreprints.e128109

Appendix 1. Selected list of aquatic plant species in the Fennoscandian database (2079 lakes)

Code	Occurrence (# lakes)	Species names
BALD_REP	1	*Baldellia repens* (Lam.) Ooststr. Ex Lawalrèe
CALA_GLO	6	*Calamistrum globuliferum* (L.) Kuntze
CALL_COP	46	*Callitriche cophocarpa* Sendtn. ex Hegelm.
CALL_HAM	226	*Callitriche hamulata* Kütz. ex W.D.J.Koch
CALL_HER	103	*Callitriche hermaphroditica* L.
CALL_PAL	300	*Callitriche palustris* L.
CALL_STA	23	*Callitriche stagnalis* Scop.
CERA_DEM	239	*Ceratophyllum demersum* L.
CHAR_ACU	2	*Chara aculeolata* Kützing, 1832
CHAR_ASP	64	*Chara aspera* Willdenow
CHAR_BRA	7	*Chara braunii* C.C. Gmelin
CHAR_CAN	1	*Chara canescens* Desvaux & Loiseleur
CHAR_CON	71	*Chara contraria* A. Braun ex Kützing
CHAR_FIL	1	*Chara filiformis* Hertzsch
CHAR_GLO	185	*Chara globularis* Thuiller
CHAR_HIS	22	*Chara hispida* Linnaeus
CHAR_PAP	19	*Chara papillosa* Kützing, 1834
CHAR_STR	37	*Chara strigosa* A. Braun
CHAR_SUB	25	*Chara subspinosa* Ruprecht, 1846
CHAR_TOM	26	*Chara tomentosa* Linnaeus
CHAR_VIR	133	*Chara virgata* Kützing
CRAS_AQU	41	*Crassula aquatica* (L.) Schönl.
ELAT_HEX	23	*Elatine hexandra* (Lapierre) DC.
ELAT_HYD	145	*Elatine hydropiper* L.
ELAT_ORT	31	*Elatine orthosperma* Düben
ELAT_TRI	74	*Elatine triandra* Schkuhr
ELEO_ACI	736	*Eleocharis acicularis* (L.) Roem. & Schult.
ELOD_CAN	314	*Elodea canadensis* Michx.
ELOD_NUT	39	*Elodea nuttallii* (Planch.) H. St. John
HIPP_VUL	380	*Hippuris vulgaris* L.
HOTT_PAL	35	*Hottonia palustris* L.
HYDR_MOR	193	*Hydrocharis morsus-ranae* L.
ISOE_ECH	767	*Isoëtes echinospora* Durieu
ISOE_LAC	851	*Isoëtes lacustris* L.
JUNC_BUL	535	*Juncus bulbosus* L.
LEMN_GIB	2	*Lemna gibba* L.
LEMN_MIN	444	*Lemna minor* L.
LEMN_TRI	117	*Lemna trisulca* L.
LIMO_AQU	42	*Limosella aquatica* L.
LITT_UNI	314	*Littorella uniflora* (L.) Asch.
LOBE_DOR	717	*Lobelia dortmanna* L.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

LURO_NAT	1	*Luronium natans* (L.) Raf.
LYTH_POR	21	*Lythrum portula* (L.) D.A.Webb
MYRI_ALT	977	*Myriophyllum alterniflorum* DC.
MYRI_SIB	75	*Myriophyllum sibiricum* Kom.
MYRI_SPI	141	*Myriophyllum spicatum* L.
MYRI_VER	129	*Myriophyllum verticillatum* L.
NAJA_FLE	13	*Najas flexilis* (Willd.) Rostk. & W.L.E.Schmidt
NAJA_MAR	4	*Najas marina* L.
NAJA_TEN	6	*Najas tenuissima* (A. Braun) Magnus
NITE_CON	7	*Nitella confervacea* (Brébisson) A. Braun ex Leonhardi
NITE_FLE	NA	*Nitella flexilis* (Linnaeus) C. Agardh
NITE_GRA	2	*Nitella gracilis* (Smith) Agardh
NITE_MUC	29	*Nitella mucronata* (A. Braun) Miquel
NITE_O_F	450	*Nitella opaca / Nitella flexilis*
NITE_OBT	5	*Nitellopsis obtusa* (N.A. Desvaux) J. Groves
NITE_OPA	NA	*Nitella opaca* (C. Agardh ex Bruzelius) C. Agardh
NITE_TRA	2	*Nitella translucens* (Persoon) C. Agardh
NITE_WAH	24	*Nitella wahlbergiana* Wallman
NUPH_LUT	1364	*Nuphar lutea* (L.) Sm.
NUPH_PUM	241	*Nuphar pumila* (Timm) DC.
NYMP_ALB	1010	*Nymphaea alba* L.
NYMP_PEL	4	*Nymphoides peltata* (S.S.Gmel) Kuntze
NYMP_TET	201	*Nymphaea tetragona* Georgi
PERS_AMP	384	*Persicaria amphibia* (L.) Delarbre
PERS_FOL	3	*Persicaria foliosa* (H. Lindb.) Kitag.
POTA_ALP	415	*Potamogeton alpinus* Balb.
POTA_BER	458	*Potamogeton berchtoldii* Fieber
POTA_COM	63	*Potamogeton compressus* L.
POTA_CRI	116	*Potamogeton crispus* L.
POTA_FRI	85	*Potamogeton friesii* Rupr.
POTA_GRA	532	*Potamogeton gramineus* L.
POTA_LUC	105	*Potamogeton lucens* L.
POTA_NAT	1208	*Potamogeton natans* L.
POTA_OBT	304	*Potamogeton obtusifolius* Mert. & W.D.J.Koch
POTA_PER	823	*Potamogeton perfoliatus* L.
POTA_POL	43	*Potamogeton polygonifolius* Pourr.
POTA_PRA	296	*Potamogeton praelongus* Wulfen
POTA_PUS	27	*Potamogeton pusillus* L.
POTA_RUT	26	*Potamogeton rutilus* Wolfg.
POTA_TRI	1	*Potamogeton trichoides* Cham. & Schltdl.
RANU_AQU	44	*Ranunculus aquatilis* L.
RANU_CIR	55	*Ranunculus circinatus* Sibth.
RANU_CON	112	*Ranunculus confervoides* (Fr.) Fr.
RANU_PEL	328	*Ranunculus peltatus* Schrank
RANU_RPT	804	*Ranunculus reptans* L.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

RANU_TRI	24	*Ranunculus trichophyllus* Chaix
SAGI_NAT	100	*Sagittaria natans* Pall.
SAGI_SFO	217	*Sagittaria sagittifolia L.*
SPAR_ANG	630	*Sparganium angustifolium* Michx.
SPAR_EME	523	*Sparganium emersum* Rehmann
SPAR_GLO	2	Sparganium glomeratum (Laest. ex Beurl.) Beurl.
SPAR_GRA	382	*Sparganium gramineum* Georgi
SPAR_HYP	98	*Sparganium hyperboreum* Læst. ex Beurl.
SPAR_NAT	149	*Sparganium natans* L.
SPIR_POL	109	*Spirodela polyrhiza* (L.) Schleid.
STRA_ALO	103	*Stratiotes aloides* L.
STUC_FIL	181	*Stuckenia filiformis* (Pers.) Börner
STUC_PEC	77	*Stuckenia pectinata* (L.) Börner
STUC_VAG	4	*Stuckenia vaginata* (Turcz.) Holub
SUBU_AQU	492	*Subularia aquatica* L.
TOLY_CAN	12	*Tolypella canadensis* Sawa
UTRI_AUS	NA	*Utricularia australis* R.Br.
UTRI_INT	250	*Utricularia intermedia* Hayne
UTRI_MIN	270	*Utricularia minor L.*
UTRI_O_S	85	*Utricularia ochroleuca / stygia*
UTRI_OCH	NA	*Utricularia ochroleuca* R.W.Hartm.
UTRI_STY	NA	*Utricularia stygia* G.Thor
UTRI_V_A	679	*Utricularia vulgaris / australis*
UTRI_VUL	NA	*Utricularia vulgaris* L.
ZANN_PAL	24	*Zannichellia palustris* L.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.