SELINA

SCIENCE FOR EVIDENCE-BASED AND SUSTAINABLE DECISIONS ABOUT NATURAL CAPITAL

D3.3 Report on definition of reference conditions that describe good ecosystem condition

v2.2

26/06/2025

Lead Author/s: Bálint Czúcz, Meri Lappalainen, Susanna Jernberg, Samuli Korpinen

Contributing Author/s: Chiara Cortinovis, Anders Kolstad, Sabine Lange, Paula Rendón, Graciela Rusch, Hanno Sandvik, Fernando Santos-Martín, Eszter Tanács, Sara Vallecillo, Fátima Arrogante-Funes, Emily Bank, David Barton, Filipe Bernardo, Nicolas Grondard, Paulo Pereira, Adrián Regos, Vanya Stoycheva, Thomas Strasser, Christos Zoumides, Iwona Zwierzchowska

Funded by the European Union

SELINA receives funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101060415. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

Prepared under contract from the European Commission
Grant agreement No. 101060415
EU Horizon 2020 Research and Innovation Actions

Project acronym:	**SELINA**
Project full title:	**Science for Evidence-based and sustainabLe decIsions about NAtural capital**
Project duration:	01.07.2022 – 30.06.2027 (60 months)
Project coordinator:	Prof Dr Benjamin Burkhard, Gottfried Wilhelm Leibniz University Hannover
Call:	HORIZON-CL6-2021-BIODIV-01
Deliverable title:	Report on definition of reference condition that describe good EC
Deliverable n°:	D3.3
WP responsible:	WP3
Nature of the deliverable:	Report
Dissemination level:	Public
Lead beneficiary:	NINA/Syke
Citation:	Czúcz B, Lappalainen M, Jernberg S, Korpinen S (2025). SELINA D3.3: Report on definition of reference conditions that describe good ecosystem condition.
Due date of deliverable:	Month n°36
Actual submission date:	Month n°36

Deliverable status:

Version	**Status**	**Date**	**Author(s)**	**Reviewer**
1.1	Draft	09/05/2023	Bálint Czúcz, Anders Kolstad (NINA)	SELINA Task 3.3 participants
1.2	Draft	01/10/2023	Bálint Czúcz (NINA)	SELINA WP3 participants
1.3	Draft	12/06/2024	Bálint Czúcz (NINA)	SELINA WP3 participants
2.0	Draft	05/05/2025	Bálint Czúcz (NINA), Meri Lappalainen, Samuli Korpinen, Susanna Jernberg (SYKE), SELINA Task 3.3 participants	SELINA Task 3.3 participants
2.1	Draft	02/06/2025	Bálint Czúcz (NINA), Meri Lappalainen, Samuli Korpinen (SYKE)	Paula Rendón, Fernando Santos Martín (URJC)
2.2	Final	26/06/2025	Bálint Czúcz (NINA), Meri Lappalainen (SYKE)	Davide Geneletti (UniTrento), Lars Hein (WUR), Fernando Santos Martín (URJC)

Table of Contents

1 Preface

The importance of biodiversity, natural capital and ecosystems condition and the services they supply has increasingly been acknowledged in diverse policy initiatives such as the EU Nature Restoration Regulation, EU Biodiversity Strategy to 2030, Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES), UN's System for Environmental Economic Accounting – Ecosystem Accounting (SEEA EA), Intergovernmental Panel on Climate Change (IPCC) and the General Biodiversity Framework under the Convention on Biological Diversity (CBD GBF).

The EU Horizon Research and Innovation Action "Science for evidence-based and sustainable decisions about natural capital" (SELINA) aims to provide robust information and guidance that can be harnessed by different stakeholder groups to support transformative change in the EU, to halt biodiversity decline, to support ecosystem restoration and to secure the sustainable supply and use of essential ecosystem services (ES) in the EU by 2030.

SELINA builds upon the Mapping and Assessment of Ecosystems and their Services (MAES) initiative that has provided the conceptual, methodological, data and knowledge base for comprehensive assessments on different spatial scales, including the EU ecosystem assessment (Maes et al., 2020) and assessments in EU member states. As knowledge and data for different ecosystem types are increasingly available, methodological development can readily provide operational products.

The overall objective of Work Package (WP) 3 "Ecosystem type, biodiversity & condition mapping and assessment" is to develop and test a methodology to map and assess the condition of terrestrial and aquatic ecosystems to support the EU implementation on ecosystem accounting, the legally binding restoration targets in BDS, and thus a better integration of ecosystem condition in public and private decision-making on different levels. One of the tasks in WP3 was to develop the concept of reference condition across all the ecosystem types and associate this with the indicator development.

Deliverable 3.3 "Definition of reference conditions that describe good ecosystem condition" presents the general concept of reference condition and how it can be operationalized to numeric reference levels in an indicator. Already looking forward to next SELINA tasks, this deliverable also paves the way for ecosystem assessments by providing an indicator concept that can be readily integrated with other indicators.

2 Summary

SELINA Deliverable 3.3 presents a methodology for the definition and application of reference conditions and levels as well as the indicator rescaling process in the context of ecosystem condition assessments.

Chapter 4 establishes the importance of references in providing meaning to ecological measurements and the need for international comparability in ecosystem condition assessments. It highlights the inconsistencies in the terminology and methods found amongst existing frameworks like SEEA EA, the Water Framework Directive (WFD), and the Marine Strategy Framework Directive (MSFD). The SELINA reference framework aims to provide a "common denominator" across these frameworks for harmonized reference setting and rescaling.

Chapter 5 outlines the process of ecosystem condition assessments in accordance with the SEEA EA, focusing on the rescaling of ecosystem condition variables into EC indicators on a 0-1 scale. It highlights differences and similarities between the concepts of reference condition (RC) and reference levels (RL) and establishes the use of continuous vs. discrete scales and the mathematical role of reference levels in rescaling functions.

Chapter 6 establishes the process for setting references, starting from setting the appropriate reference condition for the ecosystem type, establishing the reference levels (starting from the “upper” RLs) and moving on to rescaling. It defines the differences and uses of natural, seminatural, and anthropogenic RCs and presents a list of methods that can be used for setting references in different contexts.

Chapter 7 provides practical examples of how the previously established concepts and methods have been applied in previous policy frameworks, assessments and studies. Section 7.1 provides a collection of examples from SELINA partners, including SELINA Demonstration Projects and Test Sites. Section 7.2 and 7.3 present in-depth analyses of the similarities and differences between the SELINA reference framework and the EU Water Framework Directive (WFD) and the Marine Strategy Framework Directive (MSFD).

Chapter 8 presents recommendations on the best practices for the entire rescaling process, emphasizing the need for transparency, consistency, and clear documentation.

Annex 1 contains a glossary of the key terms.

Overall, this report provides a comprehensive framework for ecosystem condition assessment, with the ambitious aim to harmonize approaches across different policies, frameworks and scientific disciplines.

3 List of abbreviations

BDS	Biodiversity Strategy
CBD	Convention on Biological Diversity
CFP	Common Fisheries Policy
DP	Demonstration Project
EC	European Commission
EEA	European Environment Agency
ES	Ecosystem Services
ET	Ecosystem Type
ETC/BD	European Topic Centre on Biological Diversity
EU	European Union
EUWI	European Union Water Initiative
FAD	Forest Area Density
FCI	Forest Condition Index
GES	Good Environmental Status
HELCOM	Helsinki Commission (Baltic Marine Environment Protection Commission)
IBECA	Index-Based Ecological Condition Assessment
INES	INtegrated assessment and mapping of water-related Ecosystem Services for nature-based solutions in river basin management
IPBES	Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services
IPCC	Intergovernmental Panel on Climate Change
MAES	Mapping and Assessment of Ecosystems and their Services
MCDA	Multi-Criteria Decision Analysis
MSFD	Marine Strategy Framework Directive
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
RC	Reference Condition
RL	Reference Level
SEEA	System of Environmental Economic Accounting
SEEA EA	System of Environmental Economic Accounting - Ecosystem Accounting
SOC	Soil Organic Carbon
TS	Test Site
UN	United Nations
WFD	Water Framework Directive
WP	Work Package

4. Introduction

Measurements are important tools for environmental policy. Numbers are often considered to convey objective messages, which can effectively support evidence-based governance and decision making. Nevertheless, there is a question that pops up very frequently in some form: *How can one decide if a measured number indicates a 'good' or a 'bad' situation*? The simple answer is that one needs a *reference* for this (Fig. 4.1). References are the tools that can give the numbers a meaning that goes beyond the specific measurement context. Consequently, they are also needed for meaningful comparisons across the different contexts. Well-chosen references can help policy to get a higher-level overview and show the big picture behind the data. Poorly chosen references can, on the other hand, lead to invalid comparisons and biased aggregations, resulting in misleading artefacts. As ecosystem accounting hinges around comparisons and comparability, references have a central role in accounting, in particular in ecosystem condition accounts that indicate the state or health of an ecosystem. Condition accounts include different indicators, and a meaningful comparison or integration of these indicators can only be done when they are each compared to their specific reference level.

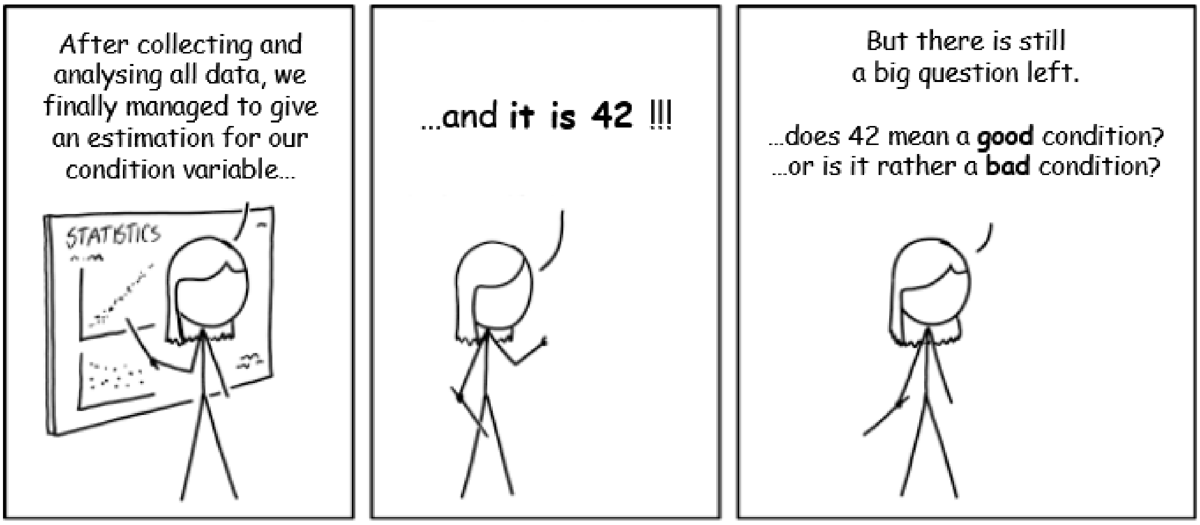


Figure 4.1. Why do we need references? (source: https://xkcd.com/2560/, modified)

While in everyday language the words "baseline", "benchmark", and "reference" are often used interchangeably, there is a slight difference in their meaning, which is also reflected in the scientific use of these terms. A *baseline* is typically associated with an initial point in time, representing a state against which future changes are measured. A *benchmark* is traditionally something that is "best in its class", which is used in comparisons for this reason[1]. While baselines and benchmarks are typically specific to a concrete context (time series, class of objects), the term *reference* typically denotes a point of comparison that goes beyond the current context, and is connected to something external, theoretical, or universal. A reference can be, for example, a set of values, a model, or criteria that can guide assessments and comparisons. In assessments, references often have an important role in providing context, guiding evaluations, and ensuring consistency. For example, in the context of climate change

[1] This term, originally coming from the military & business sphere, has actually an interesting history: see e.g., https://en.wikipedia.org/wiki/Benchmarking.

impact assessments, baselines refer to current or pre-intervention states and references generally refer to standard scenarios or models (Field et al., 2014). Similarly, in clinical trials and healthcare research, the term baseline is often used to denote the state before treatment begins, while “reference standards” are generally used for evaluating outcomes (Friedman et al., 2015)[2].

International comparability is critically important for statistical accounting,[3] like the System of Economic Environmental Accounts developed by the United Nations (SEEA EA) (UN et al., 2024). References are particularly important for *ecosystem condition* accounts, which need to establish international comparability for a very broad range of biophysical variables measured in diverse units following various protocols. References in SEEA EA ecosystem condition accounts are defined on a theoretical basis (anchored to the concept of ecological integrity, see e.g., Andreasen et al., 2001, Stoddard et al., 2006), in order to enable meaningful comparisons across a broad range of geographical, environmental, and sectoral contexts. SEEA EA also made major efforts towards establishing a simple and standardised workflow for setting references in a meaningful and harmonised way across biomes and disciplines. This work was then continued in the EU-wide methodology to map and assess ecosystem condition (Vallecillo et al., 2022).

Both SEEA EA and the EU-wide methodology are relatively recent frameworks, which aim to provide a universal solution for measuring condition that is applicable in all types of ecosystems in any part of the world. Nevertheless, in Europe there exist several long-standing policy frameworks supported by dedicated scientific communities, who have elaborated tailored solutions for measuring the condition of specific ecosystem types during the past decades. This includes, for example, the Water Framework Directive (WFD – for freshwater and coastal marine ecosystems), the Marine Strategy Framework Directive (MSFD – for marine ecosystems)[4], and there are also other frameworks at the country level like the Index-Based Ecological Condition Assessment framework (IBECA – for several terrestrial ecosystem types, Jakobsson et al., 2020, 2021) in Norway. While in theory, the principles and techniques applied by these condition frameworks are similar to the ones proposed by the SEEA EA, there is also a lack of coherence in the way how these principles and methods are used. For instance, the different frameworks often apply different terms with the same meaning (e.g., to describe the same process in a different biome), while in other cases the same terms are used with different meanings. This inconsistency of terminology masks the (much smaller) inconsistencies in methods and workflows, making intersectoral / interdisciplinary collaboration and the harmonisation of efforts challenging.

---

[2] In contrast with *baselines* and *references*, for *benchmarks* there is a slight difference between the original meaning of the term and its current use in biodiversity science, where *benchmarks* often denote targets or goals for future conditions, typically based on management objectives or regulatory requirements. In this sense *benchmarks* may be seen as a synonym/subtype of references (where the external point of comparison is a normative consideration).

[3] “*To provide an integrated, complete system of accounts enabling international comparisons*” is the main goal of the UN System of National Accounts, which is the model system for SEEA (https://en.wikipedia.org/wiki/System_of_National_Accounts).

[4] This report makes the point that the concept of ecological/environmental “status” (as applied in the WFD (2000/60/EC) and the MSFD (2008/56/EC)) is eq uivalent with the concept of ecosystem condition as used by the SEEA EA.

Table 4.1. The major condition frameworks and external rescaling frameworks analysed in this study.

	Name of the framework / embedding policy	Major ETs covered	Key references	Comments
SEEA EA	System of Economic Environmental Accounts, Ecosystem Accounting	All	UN et al., 2024; Keith et al., 2020	main framework of reference
EU methodology	EU-wide methodology to map and assess ecosystem condition	All	Vallecillo et al., 2022	based on SEEA EA
WFD	Water Framework Directive	Freshwater, (coastal) marine	https://environment.ec.europa.eu/topics/water/surface-water_en	
MSFD	Marine Strategic Framework Directive	Marine	https://environment.ec.europa.eu/topics/marine-environment_en; EU, 2017	
CFP	Common Fisheries Policy	Marine	https://oceans-and-fisheries.ec.europa.eu/policy/common-fisheries-policy-cfp_en	(no rescaling, just reference levels)
EU forest condition accounts	Ecosystem condition accounts for forests in Europe	Forests	Maes et al., 2023	(SEEA test case at a continental scale)
IBECA	Index-Based Ecological Condition Assessment framework of Norway	Terrestrial (semi) natural	Jakobsson et al., 2020; 2021; Framstad et al., 2022	
MCDA	Methods used for rescaling variables in the context of Multi-criteria Decision Analysis	–	https://en.wikipedia.org/wiki/Multiple-criteria_decision_analysis	external framework
Composite socio-economic indicators	Methods used for rescaling socio-economic indicators to produce composite indicators (CI)	–	Nardo et al., 2005	external framework

The main focus of this report is the use of references (including *reference levels* and *reference condition*, see Chapter 5.1) across these frameworks. Starting from the terminology and principles of SEEA EA and the EU methodology, we aim at establishing a "common denominator" across the different frameworks, which can describe the reference-setting and rescaling approaches followed by the *studied condition frameworks* (Table 4.1). To achieve this, we follow an interpretative approach, linking the terms Furthermore, we also looked at several further "*external frameworks*" from other disciplines (beyond environmental policy) where "rescaling techniques" (that require a reference level in support of scaling) are applied to bring a heterogeneous set of raw variables to a harmonised scale in a context oriented at policy/decision support. The outcome of this process is a harmonised framework, the *SELINA reference framework for condition accounting*, that can be applied to any type of condition variable in a robust, meaningful, and operative way.

We think that the SELINA reference framework can provide significant support to the implementation of EU policies related to ecosystem condition, enabling them to set references and rescale variables in a practical, efficient yet scientifically solid way, which is also consistent with SEEA EA. Furthermore, the SELINA reference framework can hopefully also initiate and support the convergence of the diverse methods and terminologies currently applied in the different EU policies. Finally, the application of these harmonised concepts and methods in SELINA *Demonstration Projects* (DP) and *Test Sites* (TS) can also help to test and fine-tune this framework, providing vital feedback for its future development, and a stock of use cases for its potential future users.

The remaining chapters of the report present the details of this SELINA reference framework, richly illustrated with examples and explanations. Chapter 5.1 presents two key SEEA EA concepts, *reference condition* (RC), and *reference levels* (RL). Chapter 5.2 explains SEEA EAs approach for rescaling condition variables into condition indicators, and 5.3 presents practical mathematical solutions for implementing this rescaling with the help of reference levels. Then Chapter 6 discusses the methods that can be used for identifying a meaningful reference condition and reference levels in diverse ecological and policy contexts. Finally, Chapter 7 gives practical examples how these can be applied in different ecosystem types, and provides a detailed analysis of the use of references in two major EU environmental policies.

5 The SELINA reference framework

SEEA EA defines ecosystem condition as the quality of an ecosystem measured in terms of its abiotic and biotic characteristics (UN et al., 2024; Keith et al., 2020). To make this abstract concept measurable and comparable across countries and ecosystems, SEEA EA proposes a stepwise procedure consisting of several consecutive stages[5]:

- **Stage A**: identification / delineation of **ecosystem types** (ETs);[6]
- **Stage B**: identification of *ecosystem condition* (EC) **variables** (for each ET);[7]
- **Stage C**: *rescaling* the variables into EC **indicators** (for each variable);
- **Stage D**: *aggregation* of EC indicators into EC **indices** (for each ECT / ET, optional)[8].

From this list, *Stage C* is the main topic of this report: this is the stage where references get a central role. Nevertheless, all of the stages share a common final goal, so they need to form a coherent workflow, with the earlier stages preparing the scene for the subsequent ones. Therefore, this report also aims to ensure that Stages A and B support Stage C and that the framework presented here will also support Stage D.

Accordingly, the identification / definition of ETs and EC variables (*Stages A-B*) sets the boundaries of the rescaling. For instance, if the ETs are too broad and heterogeneous, or if a variable has too much variance "*at the reference condition*", then it will be challenging (or even impossible) to identify RLs *against which it is meaningful to compare other values of the variable* (SEEA EA 5.65). This is especially problematic in urban and agricultural ecosystems, since they have been very intensively managed for millennia. Similarly, a coherent rescaling (as well as a representative selection of variables) is a prerequisite for the aggregation of indicators into meaningful condition indices. Accordingly, the four *main stages* (and the corresponding four SELINA WP3 Tasks) are parts of a single larger assessment framework, which has to be performed in a coordinated way. To help with this coordination we will also give recommendations (see Chapter 8).

Looking more concretely at *Stage C*, this task can also be described as a series of four consecutive steps:

- **Step 1**: identification of a suitable ***reference condition*** (**RC,** one for each ET),
- **Step 2**: identification of ***reference levels*** (**RL**) **linked to the RC** (one RL per variable),
- **Step 3**: identification of **further RLs along the range of values of the variable**,
- **Step 4**: constructing a **rescaling function** (one per variable) and doing the rescaling.

This order of steps is rooted in SEEA EA and its supporting literature (Keith et al., 2020), but the same steps can also be identified in most of the studied condition frameworks, too (e.g., HELCOM, 2013; EC, 2011). In the following section we will explain the different concepts and

[5] This is a slightly extended version of the three stages discussed by Keith et al., (2020).

[6] Stage A is covered in detail by SELINA Deliverable 3.1 (Rendon et al., 2023)

[7] Stage B is covered in detail by SELINA Deliverable 3.2 (Nicholson Thomas et al., 2025)

[8] The classes of the SEEA EA Ecosystem Condition Typology (ECT, SEEA EA's major reporting categories for condition accounts) are presented in SEEA EA and Czúcz et al., 2021a; Stage D will be covered in SELINA Deliverable 3.4.

methods that the studied frameworks use to implement these steps. Section 5.1 explains the key elements (concepts) behind the selection and use of references in a rescaling transformation (Steps 1-4). Section 5.2 then explores the "practical side" of using RLs, by discussing the different "roles" they can play in the rescaling (Step 4). This is followed by an overview of the "normative considerations" for reference levels, represented by the RCs (Step 1, Section 5.3.1), and detailed descriptions of the different RL-setting methods that can be applied in the different cases (Steps 2-3, Sections 5.3.2-5). Finally, we will also discuss "alternative methods" for rescaling, which circumvent the concept of references, and are thus not fully compatible with SEEA EA principles.

5.1 Reference condition and reference levels

SEEA EA provides the first international operationalization of metrics of ecological condition based on a series of indicators integrated into an index that can inform about spatio-temporal changes of all ecosystems. It measures the condition of an ecosystem "*in terms of the distance between its current condition and a reference condition*", where the *reference condition* is a state of the ecosystem "*against which it is meaningful to compare*" all of its other possible states (SEEA EA 5.69). This definition follows a long-standing logic (Andreasen et al., 2001) and a broad range of existing practices in several ecosystem types and policy domains (Birk et al., 2013; Wurtzebach & Schultz, 2016). To make the measurement of this "distance" operational, SEEA EA introduced a second important concept: *reference levels* (RL).[9] Accordingly, the term reference condition (RC) denotes a *multidimensional state* of the entire (eco)system, while the term reference level (RL) always refers to a *concrete value* of a single variable (SEEA EA 5.65).[10]

SEEA EA also distinguishes two types of quantitative metrics: *variables* and *indicators*. These two terms are often used interchangeably in ordinary language, but for the sake of operationalization, SEEA EA makes a useful distinction. *Variables* in SEEA EA are "raw" biophysical quantities measured in their natural units, while *indicators* are transformed (*rescaled*) versions of these variables, which are brought to dimensionless indicator scale comparing the actual value to the reference levels (SEEA EA 5.63). In this report, we have followed this definition even in cases where the original texts use them differently (e.g., in Chapter 6 HELCOM indicators are here called HELCOM variables).

SEEA EA further standardises this common dimensionless scale (henceforward called the *indicator scale*) in the following way:

- the scale ranges from 0 to 1 (i.e., condition indicators can take up only values within this interval);
- a value of 1 should correspond to the RC (i.e., the selection of condition variables and their rescaling to indicators should ensure that if the ecosystem is in RC, then the value of all indicators is 1);

[9]

[10] This also implies that variables that cannot tell if we are "close to" or "far from the RC" are not so useful as condition variables. This an important criterion for condition variables (labelled as "directional meaning" in SEEA EA Annex 5.1, and Czúcz et al., 2021b) which is covered in SELINA under Deliverable 3.2 (Nicholson Thomas et al., 2025).

- a value of 0 should correspond to a "maximally degraded" condition (this should lead to meaningful comparisons but does not need to correspond to an existing condition – c.f. SEEA EA 5.65-66).

Accordingly, the "raw" condition *variables* provide a relatively technical description of an ecosystem, while the rescaled *indicators* allow integration of multiple indicators regardless of the variable scales and units, but still answer to the "central question" of measuring ecosystem condition: "*How far are we from the reference condition*?".[11] This means that condition indicators are significantly easier to interpret & communicate across diverse scientific and policy contexts than the raw variables (Jakobsson et al., 2021) – as long as the rescaling is implemented in a scientifically *valid* and *transparent* way. Furthermore, rescaled indicators also enable the integration of variables across different ecosystem properties, including biotic structures and function, and chemical and physical properties (e.g., Nygård et al., 2016). This can have a wide range of policy applications (Czúcz et al., 2025, Ramon et al., 2025), including the development of *condition indices* in Stage D of condition accounts (Keith et al., 2020). Accordingly, both RC and RL are *instruments* to support the rescaling between variables and indicators. A consistent and transparent application of RCs and RLs can help to maintain the integrity of the condition accounts (Table 5.1).

Table 5.1. Differences and similarities between the concepts of *reference condition* (RC) and *reference levels* (RL) in the context of SEEA EA.

	Reference level (RL)	Reference condition (RC)
What is it? (*Ontology*)	A *concrete value* of a *specific condition variable* (to which it makes sense to compare other values)	A *state* (or range of states) of the *entire (eco)system* in a natural state (defined theoretically or via prototypes)
How many of them are there? (*Cardinality*)	Each condition variable requires a set of RLs: - A minimum of 2 RLs associated with *scale endpoints* (see Chapter 5.2) - Other RLs may correspond with *class boundaries* (see Chapter 5.2)	Each ecosystem type (ET) requires a single well-defined RC linked to the *upper endpoint* of the indicator scale (see Chapter 5.2), associated with the "optimal" state of the ET (e.g., the pristine state in a natural ET)
Where can they be used? (*Scope, domain*)	Each RL has a spatial & thematic *domain of validity*, which is typically just a subset of the ET (e.g., a specific ecosystem subtype (sub-ET), or a "homogeneous ecosystem area" (see Section 5.3)	Typically, the same RC is valid for the whole ET at a specific level of the ET classification (all occurrences/*assets* of the same ET are compared to the same reference state)
What are they used for? (*Purpose, function*)	To *operationalise* the rescaling (anchor specific values of the condition variable to specific points of the common indicator scale)	To make the rescaling *coherent* across indicators (anchor the *upper endpoint* of the scale to the same (theoretical) state)
Definition in the SEEA EA Glossary	The value of a variable at the reference condition,[12] against which it is meaningful to compare past, present or	The condition against which past, present and future ecosystem condition is compared in order to measure relative change over

[11] See also SEEA EA 5.69 and the opening sentence of Section 2.1. Nevertheless, the actual scale applied in condition indicators is rather a *similarity* than a *distance* scale (1 means "maximum closeness" to RC, whereas 0 means "maximum distance"), so the underlying question could be better phrased as "How *close* are we to the RC on a 0-1 scale?" (see more in Section 2.1 and SEEA EA 5.60-62).

[12] The text of SEEA EA Chapter 5 clarifies that there exist also other types of reference levels, which are not linked to the RC (e.g., see SEEA EA 5.66).

future measured values of the variable time

5.2 Rescaling variables into indicators

As discussed above, in SEEA EA all ecosystem condition indicators are measured on a *continuous scale* ranging from 0 to 1, with 1 denoting the reference condition[13]. This scale is not unique to SEEA EA: several existing frameworks for ecological condition (e.g., the EU Water Framework Directive, and the marine assessments in European Environment Agency and HELCOM) use similar scales, and they are also routinely applied in socio-economic indicators (Nardo et al., 2005), cumulative effects assessments (e.g., Korpinen et al., 2021) or in Multicriteria Decision Analysis (MCDA) studies (Bana e Costa et al., 1999; Bigaret et al., 2017). Nevertheless, this continuous scale is not the only possible standardised scale that can be used for assessing ecosystem condition. Policy users often have a preference for *discrete scales* that can be used for "grading" ecosystems according to their condition. This preference may be related to a general perceptual tendency for imposing categorical distinction over continuous data, often termed as *binary bias* in psychology (Fisher & Keil, 2018), but the use of discrete categories can also reflect real ecological discontinuities (López et al., 2011). Discrete categories can also be made after a variable has been normalized between 0 and 1 (e.g., EU WFD).

While the dimensionless continuous 0-1 scale is used practically with the same meaning in a broad diversity of various scientific and policy contexts,[14] there are many different discrete scales used by different communities in specific contexts. For example, the WFD applies a 5-grade scale to specify "*status classes*" for the "biological quality elements" (condition indicators), and these classes split the range of the continuous indicator into five equal intervals (Fig. 5.1a). The Common Fisheries Policy (CFP) applies three status classes following a "traffic light" system (green, yellow, and red). The MSFD only distinguishes two main *classes* of "environmental status" ("good" and "not good") (Fig. 5.1b). Status classes are separated from each other by *class boundaries*. In well-established indicator systems *class boundaries* usually also have their own "names". The two boundaries in CFP are called "threshold reference point" and "limit reference point",[15] and the single good/not-good boundary in the MSFD is called "threshold value". Such class boundaries often also have an important role in the rescaling transformation in their respective policy systems, as intuitive "anchoring points" for reference levels (see more in section 5.3).[16]

[13] In measurement theory this type of scale is often called an *absolute* scale (Ellis, 1966, Brunsdon, 2018).

[14] The only notable diversity is that some policies prefer to express the 0-1 values in percentages (0-100%), but from a mathematical perspective, both the absolute values and the percentages describe the same (dimensionless) scale. This report also uses both absolute values and % notation to denote indicator values.

[15] See https://harveststrategies.org/what-are-harvest-strategies/reference-points/

[16] And in rare cases the *class centres* can also be used for anchoring RLs (e.g., the *class centre* of the WFD "moderate" class is 0.5).

The transformation of data from a continuous to a discrete scale always implies some loss of information. While this may be considered as favourable in some contexts (e.g., if a binary decision needs to be taken), storing and handling primary information in qualitative/discrete scales can result in a substantial loss of precision, as well as significant threshold biases (Gorrod et al., 2013). Continuous data are also much better for statistical inference than their discretized versions, even for complex decisions that eventually require qualitative judgements (King & Powell, 2008). For instance, the European Environment Agency's WISE-Freshwater portal includes a reporting system for WFD indicator results in a continuous scale,[17] even though the official WFD reporting is on the discrete scale. Similarly, ecosystem condition accounts should always keep a copy of the original (continuous) indicator (or variable) data, even if a transformation to discrete scales is desired by policy.[18] This way the "less processed" format (continuous indicators) will always be available if further data processing or statistical analysis is desired.

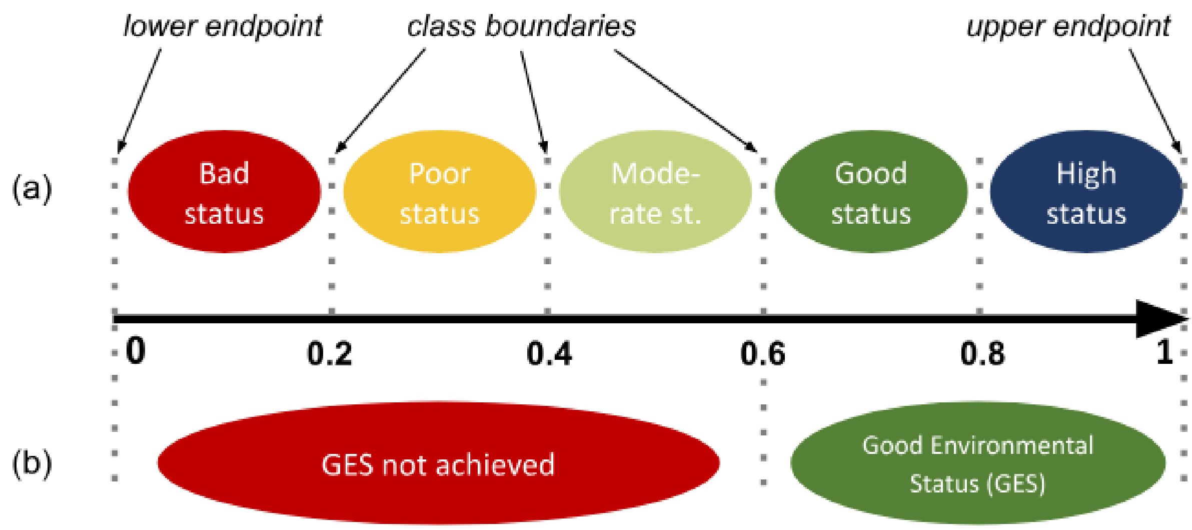


Figure 5.1. The continuous indicator scale (in the middle) matched against (a) a five-grade discrete scale applied by the EU WFD for biological quality elements, and (b) a binary scale applied by the MSFD. Both of these scales are explicitly "anchored" to the continuous indicator scale via class boundaries.

In rare cases the primary data for an ecosystem condition variable/indicator data may already arrive on a qualitative (discrete, ordinal) scale. According to social science, if there is insufficient quantitative information, it is generally a good practice to seek additional qualitative information to improve the assessment (King & Powell, 2008). Also, the HELCOM integrated biodiversity assessment involves ordinal indicators for biodiversity elements with

[17] https://water.europa.eu/freshwater

[18] On a side note, SEEA EA prescribes that the untransformed variables are always stored along with the (transformed) continuous indicators (SEEA EA 5.53, 5.63), which means that each data point needs to be stored in two different "formats" (raw, rescaled). For policies that also need a graded scale, an inventory of discrete/qualitative "status classes" can be considered as an auxiliary third "data format", which can be recorded and updated along with the two other "formats" prescribed by SEEA EA (=the accounts for variables and indicators). Alternatively, the discretisation of the indicators can also be applied on a "needs basis" at the "end of the data processing pipe", i.e., for only those outputs that are intended for direct policy consumption.

no quantitative data where expert inputs were considered as the "best available knowledge". [19] This was technically transformed as numbers on the 0-1 scale in order to support the indicator integration. In such cases, structured expert elicitation methods (O'Hagan, 2019, Courtney Jones et al., 2023) can help to rate the status of the biodiversity element for each ecosystem asset lacking data in a systematic way. Nevertheless, expert data are often viewed as subjective, and their use in accounting is controversial, which is reflected in the fact that SEEA EA recommends caution with the use of such data (SEEA EA A5.11).

5.3 The mathematical role of reference levels

The rescaling of an EC variable from its original (biophysical) scale to a new indicator scale can be seen and described as a mathematical operation (a *function*). This mathematical operation gives the context in which the RLs are practically applied. There are many subtle methodological choices inherent to this operation, and a poor understanding of these choices may easily lead to suboptimal results. Accordingly, beyond the ground truth (the biophysical reality of the ecosystems) and the policy goal (the intended meaning) an appropriate rescaling framework also needs to focus on the "mathematics" happening in the background.

In the context of this mathematical operation, the RLs are *parameters*[20] of the *rescaling function.*[21] In general, there are several different options for the mathematical function, and each of them relies on more than one parameter. A common approach is a specific "family" of simple parametric functions for rescaling that offers a good trade-off between flexibility and simplicity/transparency: i.e., *piecewise linear functions*. Such piecewise linear functions are used for rescaling, for example, in the context of WFD and MSFD (see more in Section 4), but similar functions are also routinely applied to standardise socio-economic indicators (Nardo et al., 2005), and in MCDA decision support frameworks [22].

SEEA EA also recommends the use of this function family for rescaling variables into indicators. The function described in SEEA EA (5.61) is the simplest and most straightforward of all possible (piecewise) linear rescaling functions:

$$Y = \begin{cases} \ldots\text{if } X < X_0 & \Big|\ 0 \\ \ldots\text{if } X_0 < X < X_{100} & \Big|\ (X - X_0) / (X_{100} - X_0) \\ \ldots\text{if } X_{100} < X & \Big|\ 1 \end{cases} \qquad \text{Eq. (1)}$$

[19] As applied in the recent HOLAS assessment (https://helcom.fi/baltic-sea-trends/holistic-assessments/state-of-the-baltic-sea-2023/)

[20] Here we use this term in a mathematical sense (e.g., https://www.britannica.com/topic/parameter). Some policies (e.g., WFD) use the word "parameter" in a different meaning (~"raw/elementary data").

[21] Several other disciplinary areas apply mathematically similar solutions for rescaling; however they use a different terminology to describe what they do. For example, Multicriteria Decision Analysis (MCDA) calls rescaling functions as *value functions*, and they apply a broad range of different names for RLs (e.g., upper/lower limits, thresholds, breakpoints; see e.g., Bigaret et al., 2017). In the context of *composite socioeconomic indicators*, scaling functions are often called *achievement functions*, but RLs are (fortunately) also called *reference levels*.

[22] https://en.wikipedia.org/wiki/Multiple-criteria_decision_analysis

In this function X is the original variable, Y is its rescaled version (i.e., the *indicator*), and x_{100} and x_0 are two RLs, X_{100} pertaining to the *upper endpoint* of the indicator scale (Y=1, the *reference condition*), and X_0 to the *lower endpoint* (Y=0, a "maximally degraded" condition) respectively (Fig. 5.2). Outside the range of these two RLs the value of the indicator is *truncated*, so the indicator will not distinguish variable values which are exactly equal to the RLs, and values beyond.[23] Eq. (1) can actually be used for both variables from which "more is better" (*positive directionality* sensu Czúcz et al., 2021b), as well as "less is better" variables (*negative directionality*) that often quantify negative properties of the studied ET (e.g., pollution or disturbances; SEEA EA 5:60).

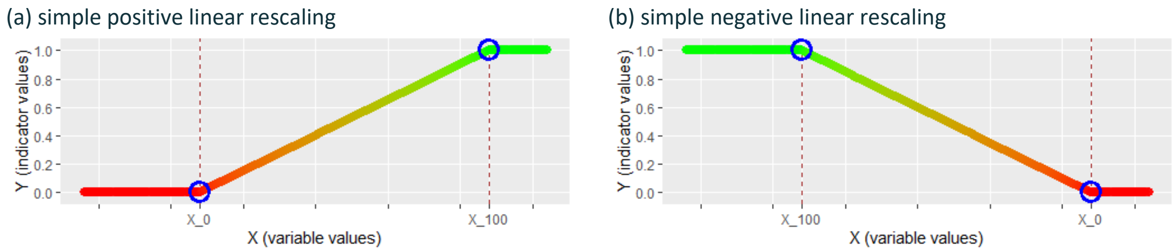


Figure 5.2. The linear rescaling function described by SEEA EA applied to (a) a variable with "positive directionality", and (b) a variable with "negative directionality"(e.g., pollutants, invasive species). The two reference levels (X_{100}: "upper RL", X_0: "lower RL", highlighted with blue circles) correspond to the two endpoints of the transformed scale. Note that X_{100} always denotes the RL corresponding to Y=1 (i.e. for negative directionality variables the "lower RL" will be higher than the "upper RL").

In addition to the two *endpoints*, a *piecewise linear* rescaling function can also have one or more *breakpoints* (Fig. 5.3), which can make the rescaling significantly more flexible. Each breakpoint consists of a reference level (RL, i.e., a concrete value of the original variable, measured in its original (biophysical) units) and an associated indicator value (i.e., a concrete number between 0 and 1, dimensionless). Several policies connect such breakpoints to *class boundaries*, which assigns an intuitive meaning to each RL for users familiar with the "status classes" (Fig. 5.1). For example, the MSFD, WFD, and several other condition frameworks (e.g., the Norwegian IBECA; Jakobsson et al., 2020) apply a *mid-range RL* anchored to the indicator value 0.6 to separate the domain of "good" condition from that of "not good" condition (Fig. 5.4a,b). Chapter 6 presents approaches for setting this mid-range RL on a variable scale, with an ecologically meaningful definition (e.g., HELCOM, 2013).

[23] This actually gives an important question which can help to check that the "optimal RL" (X_{100}) is not set "too low": *can an additional increase (above X_{100}) be still seen as an improvement?* If yes, then X_{100} is probably set too low. In principle, the upper scale endpoint should be set to a variable value which the variable cannot normally reach, and/or above which an additional increase is not a real improvement anymore (and thus e.g., it cannot compensate for a small decrease elsewhere).

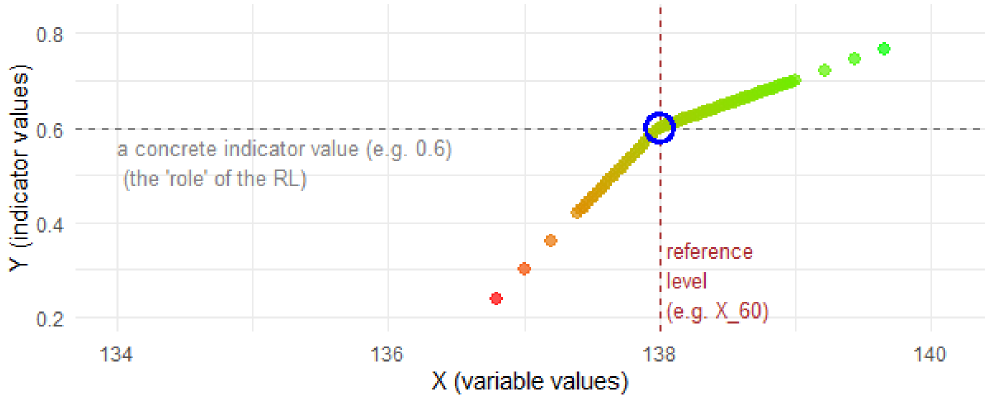


Figure 5.3. Breakpoints in a piecewise linear rescaling function are determined by a reference level (RL) and the associated indicator value.

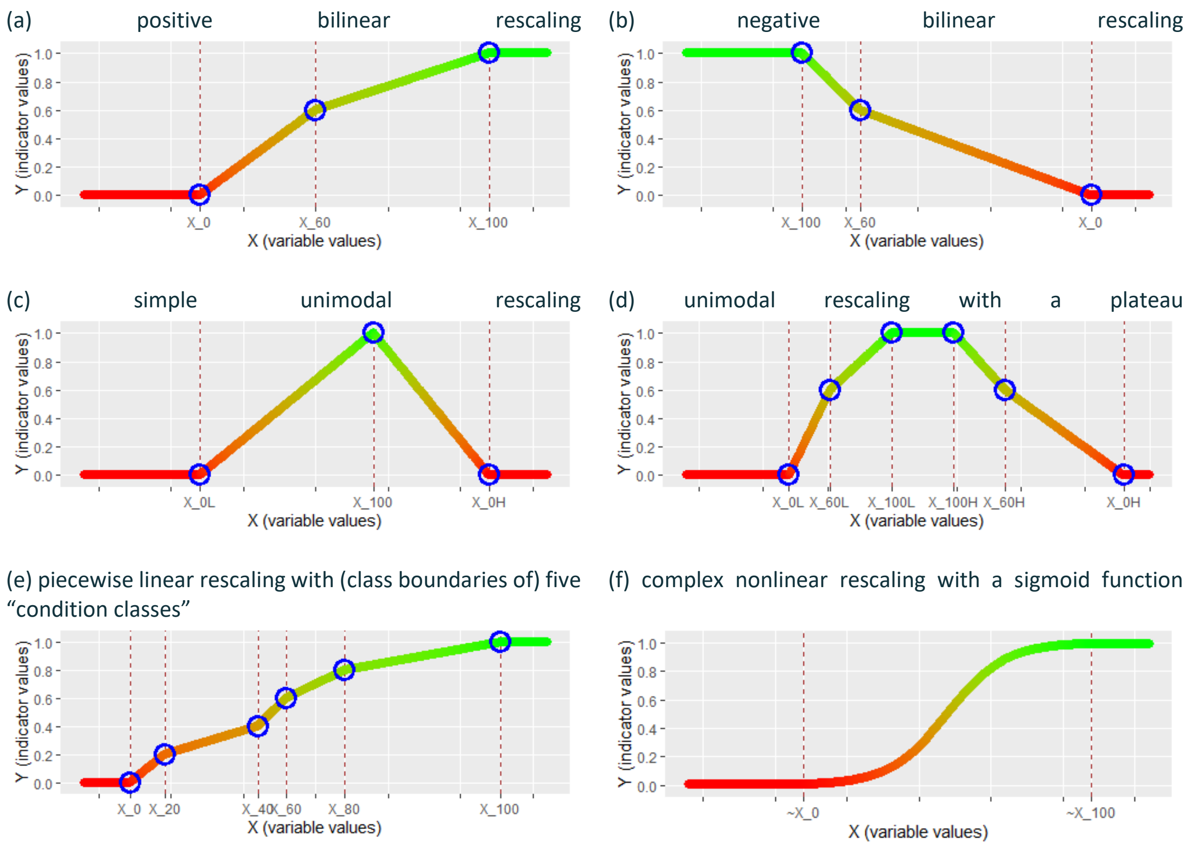


Figure 5.4. Further examples for piecewise linear rescaling functions: a, e: positive monotonic functions (with increasing complexity); b: a negative monotonic function; c, d: unimodal functions; f: a sigmoid function.

Not all variables show a monotonous directionality across their entire range. "Hump-shaped", or *unimodal* relationships (showing an *optimum* in the middle) (Fig. 5.4c,d) are relatively common in ecology (e.g., VanderMeulen et al., 2001).[24, 25] In principle, such variables can also be rescaled with piecewise linear functions, possibly applying an *optimum plateau* where a desired state is defined by two reference levels[26]. While SEEA EA itself does not discuss the use of unimodal variables, they are covered by some of the practical rescaling frameworks due to their prevalence in ecology (Keith et al., 2020, Jakobsson et al., 2020). A practical solution is to limit the variable scale on either side of the function (e.g., in a eutrophication variable define the best value on top of the unimodal function and the worst value on the right side slope at the maximal nutrient concentration). The use of unimodal variables in ecosystem condition accounts is, however, a contested topic, because they are particularly prone to aggregation bias (Rastetter et al., 1992, Pedersen & Skarpaas, 2012).[27] This means that if a condition variable is spatially aggregated over ecosystem assets that are "too low" and "too high", then the resulting average value might give a much better condition rating than the condition of any of the assets individually. This bias may be (partially) overcome by splitting a single unimodal variable into two condition variables: one describing how much we are below the optimum, and the other one describing how much we are above it (Jakobsson et al., 2021).[28] If the unimodality appears only in a larger ecoregion (over a gradient), one can also split the variable into ET subtypes.

To emphasise the link between a RL and the indicator value (endpoint or class boundary) to which it is associated, in this report we will use the following notation to distinguish the different RLs for variable X:

- X_{100} denotes a RL linked to the *upper endpoint* of the indicator scale (corresponding to the RC and an indicator value of 1), which we will henceforward simply call "upper RL".[29]
- X_0 is the RL linked to the lower scale endpoint (corresponding to an indicator value of 0, and a "maximally degraded condition", which is typically defined by some sort of

[24] In fact, many of the variables that policy currently largely considers "one-sided" (e.g., deadwood in forests, population numbers of large predator species) are actually unimodal variables where human activities have shifted the balance to just one side of the optimum (Jakobsson & Töpper, 2019).

[25] Referring to an old English fairy tale (https://en.wikipedia.org/wiki/Goldilocks_and_the_Three_Bears), such variables are sometimes also called "Goldilocks variables".

[26] In this case, the expression on 0-1 scale may require a conditional formula, where the maximal value 1 is achieved on the top of positive unimodal function and the lower values on both sides of the slopes (and vice versa in negative unimodal function).

[27] In fact, this bias affects all variables with a (strongly) non-linear relationship, not only the unimodal ones. This can be mitigated during the selection/development of EC variables by applying a transformation which ensures the relationship between the values of the variable and the (perceived) changes condition should be as linear as possible.

[28] Technically this (and other relationship-linearizing transformations performed on the variables) should be parts of Step B (the identification of condition variables). There are also several possible downsides of this solution (one of the resulting condition variables will always show an optimal state, the duplication of a variable can cause double counting during thematic aggregation (in Step D), etc.).

[29] It is important to note that for negative directionality variables (like pollutant concentrations, invasive species abundances, etc.) the suggested SELINA terminology implies that their *"upper RL" will be lower than their "lower RL"*(!) While it may seem counter-intuitive, this terminology still follows the logic of SEEA EA (see e.g., SEEA EA 5:62, which explains that "*for variables where an increase in variable value reflects a lower condition score, V_H will be lower than V_L*", and note that V_H ("high RL") and V_L ("low RL") correspond to what we call X_{100} (the "upper (endpoint) RL") and X_0 (the "lower (endpoint) RL") respectively.)

"natural" (physical, biological, mathematical) limit for the variable, e.g., a zero value – see more in Section 2.3.3).

- For any i (where 0 < i < 100), X_i denotes a RL that is anchored at the indicator value *i*/100 (e.g., X_{60} is anchored at 0.6). In practice, such *mid-range RLs* are typically linked to *class boundaries* (e.g., the "G/M boundary" in WFD, i.e., the threshold separating the status classes "good" and "medium" will be denoted here as X_{60}).
- In the case of unimodal functions there may be two RLs pertaining to the same indicator value: in this case the lower and the higher RLs are distinguished using an additional index (e.g., X_{0L} & X_{0H}, see e.g., Figs. 4.4c,d).

This notation emphasises that each RL has a specific *role*, and this role is set by the indicator value to which the RL is associated (Fig. 5.3). A clear recognition and documentation of the different roles can help to ensure accounting transparency and to avoid potential confusions and category mistakes (Wallace & Jago, 2017). Statements about RLs that don't specify *which* RL one is referring to (like "*we used 42% as the reference level*") are relatively common in environmental policy – yet they are potentially misunderstandable and therefore lack transparency. A failure to properly recognise and distinguish the different contexts and roles in a deliberation about references can lead to "*many hours of confusing discussion*" (Stoddard et al., 2006: p.1268) and may eventually derail the entire process.

While piecewise linear functions can offer a relatively good approximation for any functional form, there is also an infinite choice of more complex and sophisticated nonlinear rescaling functions (e.g., *polynomial* or *sigmoid* functions, Fig. 5.4f). Especially, the former typically doesn't have clear breakpoints (the latter may have if the mid-slope is steep, i.e., there is a discontinuity), and they also make smooth turns at the "endpoints" of the scale. But lack of clear endpoints and breakpoints also means that the function parameters will lose their clear "meaning", which makes these options less transparent / intuitive in practice. In line with the principle of parsimony, it is better to use a simpler and more intuitive rescaling model (i.e., a linear or piecewise linear rescaling), unless there is a good reason to go for a more complex one (e.g., theory suggests a nonlinear relationship with a specific functional form).

6 Methods for setting references

In the SELINA reference framework presented in Chapter 5, each (piecewise) linear rescaling function relies on *at least* two RLs, and specific methods/policies may prefer to use additional mid-range RLs (set at class boundaries). Similarly, SEEA EA, and all the EU policies discussed so far, rely on the use of several RLs. The values of these RLs will need to be determined (quantified) before any rescaling can be done. One of these RLs actually needs to be linked to the RC (specific to the studied ET), and the choice of the RC and the corresponding RL will have major influence on the resulting assessment. The different methods discussed in this Chapter can also be linked to the main steps of the workflow presented in Section 5.1. The first three steps of this workflow are connected to the task of identifying the suitable reference levels, and the rescaling itself only happens in the last step (Fig. 6.1). The different steps of the workflow are connected to different types of decisions, which require different types of participants and competences (Fig. 6.1).

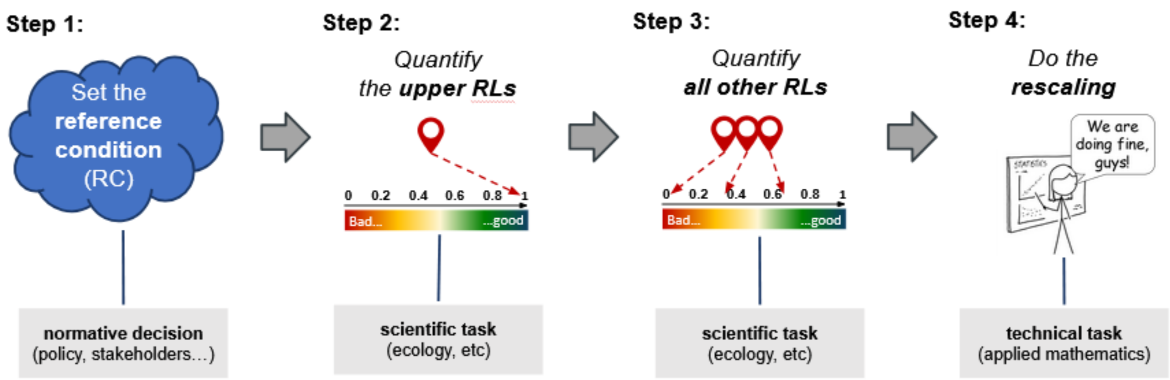


Figure 6.1. The four-step workflow for setting references for the SELINA references framework, and the main types of decisions and competencies connected to each step. (cartoon source: https://xkcd.com/2560/, modified)

There are many different methods that can be used to quantify individual RLs for an ecosystem type (ET). These methods differ considerably in their range of applicability. Depending on the type of the ecosystem, its overall condition in the studied area (and beyond), the properties of the condition variable, and the role of the RL to set, there are different methods that can be used, and there is no single method that could be used in all cases. Typically, RLs for the different variables of an ET need to be identified one by one, and the method will not be the same even for the different RLs of the same condition variable. The different methods can be divided into two main groups: (1) methods designed to address RLs that are "*close to the RC*" (henceforward called "*upper RLs*"), and (2) methods, that can address any RL roles (including *lower scale endpoints* and *status class boundaries*). The most important types of methods are listed in Table 6.1.

In the following sections we will first explore what a *reference condition (RC)* can mean for the different ETs and discuss how this RC can be "defined" from the perspective of policy and practice. Then, in Chapters 6.2 and 6.3 we will present and discuss the different families of methods listed in Table 6.1, that can be used to set the individual RLs of the diverse condition variables. Finally, Chapter 6.4 will present and discuss an additional important group of *data-driven* methods, which are highly popular in scientific studies, even though they do not create easily transferable reference levels and are thus rarely applied in major policy frameworks.

Table 6.1 An overview of the main methods used for setting reference levels (RL) in the main EU policies and scientific studies. See Chapter 7 for examples.

RL setting method	Brief description	Examples of application domains
*Methods for the upper reference level(s)**		
Reference sites	"Minimally disturbed" reference sites are often considered as the best way to "set the reference condition" for an ecosystem type.	Scientific studies, WFD, MSFD, EU forest accounts, etc.
Pressure removal	A modelling approach that links EC variables with disturbance gradients and extrapolates the RC	Scientific studies, WFD
Paleo-environ-mental data	RC can sometimes be inferred based on already disappeared historical "reference sites" (using i.a., sediment cores, museum collections, paintings, manuscripts...)	Scientific studies, WFD
Methods for any reference levels		
Special values	Special values of a variable (e.g., natural zero values, policy thresholds/targets) can often be (re)used as a RL.	Scientific studies, WFD, MSFD, EU forest accounts, etc.
RL transfer	Transferring an existing RL between two variables that are (largely) determined by the same pressure	WFD
Inter- & extrapolation	When at least 2 RLs are determined for a variable, interpolation (or extrapolation) can be used to set further RLs	WFD, IBECA

* possibly also including mid-range RLs "very close" to the upper endpoint of the condition scale (e.g., X_{80}).

6.1 The importance of the reference condition

According to the workflow presented in Section 5.1 and Fig 6.1, the identification of a RC (for an ET) involves societal value choices which specify a conceptual state. The ecological literature is relatively rich in studies which aim to set a "reference condition" for some condition variable, and this generally also involves the precise identification (description) and justification of a state that is considered to be special enough to be used as a "reference".[30] Policy frameworks which involve a reference condition also involve providing a "*normative definition*"[31] for the RC. This normative definition is critically important for guiding the identification of reference sites or the quantification of reference levels in Step 2, and in a sense it can be seen as the reference condition itself.

[30] And then the "translation" of this conceptual state into concrete numbers (=estimated RLs for concrete variables) is the task of the next step. Unfortunately, even the few studies that make a distinction between RCs and RLs (e.g., Pedersen et al., 2016; Jakobsson et al., 2020; Keith et al., 2020, and SEEA EA Annex 5.2) present the methods for "setting" them in a conflated way.

[31] This is what the WFD calls the guiding principles (concepts) associated to each class boundary in their framework, which should guide the identification of the concrete RLs.

According to SEEA EA, the RC needs to be defined independently for each ET. In other words, this means that the RC for *any ecosystem asset* needs to belong to *the same ET* as the asset itself. For example, a forest should never be presented as the RC for an agroecosystem or a city. In fact, it could be very difficult (or impossible) to compare condition variables across two different ETs, because there may be little overlap between the condition variables defined and measured for them.[32]

In this report we identify three *main types of RC* that can be used in the different ecosystem types. The two first types are connected to natural and seminatural ETs:

- A ***natural reference condition*** (Keith et al., 2020; synonyms: *undisturbed condition* in SEEA EA, *RC for biotic integrity* by Stoddard et al., 2006) is the natural or intact state of the studied ecosystem type. This is probably the most typical choice for "natural ETs" that are "native" in the study region, where natural dynamics is still more important than human activities in shaping structure, composition, and function. The RC for such an ecosystem (=*natural RC*), is a self-sustaining *natural* or *intact* state mostly governed by natural ecological and evolutionary processes, including food chains, species populations, nutrient and hydrological cycles, self-regeneration and involving dynamic equilibria in response to natural disturbance regimes (Stoddard et al., 2006; Pedersen et al., 2016; Keith et al., 2020). It may also make sense to adjust the definition of *"natural"* to include *long-term and large-scale environmental changes* (e.g., climate change) that would have occurred in the ecosystems of the study region even without the concrete local human interventions of the past.[33] The Norwegian Nature index, for example, follows this approach by conceptualising the natural reference as "*the condition in which an intact ecosystem would have been found in the climatic normal period 1961–1990*", assuming spontaneous immigration and extinction to be natural, with further specifications and details for each broad ET distinguished (Pedersen et al., 2016). Nevertheless, from a practical perspective, any definition of RC needs to balance theoretical considerations with practical ones.[34]
- A ***seminatural reference condition*** (Pedersen et al., 2016; Vrasdonk et al., 2019) is a historically stable and resilient state of a *seminatural* ET, together with its *traditional human management regime* (e.g., traditional grazing or fodder harvest). In such ecosystems the abandonment of the traditional use would negatively affect biodiversity (Pedersen et al., 2016) such as grasslands. However, it can be argued that some traditional human activities have evolved (e.g., the reindeer herding) and this may influence the choice of calling the ET seminatural. Nonetheless, the seminatural

[32] Some studies (e.g., Vrasdonk et al., 2019) also discuss an additional type of RC, called "re-naturalisation RC". This RC is defined as the hypothetical "future state" which would characterise the studied region if (suddenly) all human interventions had stopped. Nevertheless, such definition would frequently and easily lead to an "RC" that is of different ET than the original type of the ecosystem – which therefore does not comply with the principles of SEEA EA.

[33] Vrasdonk et al., (2019) calls this "natural counterfactual".

[34] Some studies use a concrete year when extensive anthropogenic changes started to happen in the given region to specify the RC applied in the study (e.g., Australian studies may say that they use 1750 as their reference condition; Stoddard et al., 2006; Keith et al., 2020). While this may look like a *baseline year* approach (Section 2.3.6), studies specifying a "*preindustrial era baseline*" are actually different (SEEA EA A5.9), and a preindustrial "baseline" is practically the same as a natural (or seminatural) *reference*.

ETs can be best described (and conserved) as socio-ecological systems, where the traditional human activities as well as the human communities performing them are integral parts of the system. Similarly to a natural RC, it may also make sense to include the effect of large-scale and long-term environmental changes into the definition of a seminatural RC. This type of RC should, nevertheless, always be defined and documented carefully, making all assumptions explicit in relation to the purpose of the ecosystem condition accounts (Keith et al., 2020).

For *natural ETs* SEEA EA clearly recommends the use of a *natural RC* (e.g., SEEA EA 5.70), whereas for *seminatural ETs* it recommends the option described above as *seminatural RC.*[35] Nevertheless, for highly *anthropogenic ETs*, like croplands or urban ecosystems, SEEA EA does not provide a particularly clear description about the type of RC to be used or the method to be followed to identify it. In such ETs, there are typically no "ideal prototypes" that could be used to quantify (or demonstrate) a suitable RC, and even to imagine such an RC is a hard exercise. Accordingly, it may be better to "skip" fruitless attempts at defining fictive (and possibly controversial and erroneous) ecosystem states, and proceed directly to identifying RLs for each variable. Fortunately, anthropogenic ETs are often well equipped with "special values" (e.g., management targets, regulation thresholds, etc. – see Section 6.3.1), which can possibly yield meaningful and policy-relevant RLs for the most important variables (Kervinio et al., 2023), and it is also possible to use extrapolation to get a suitable upper RL, when two other RLs have been identified (see Section 6.3.3).[36] If one manages to identify a meaningful "upper" RL for each of the condition variables, then the ensemble of these upper RLs defines a fictive "state" of the studied ET (the state with all of the variables at their respective RLs at the same time). While this state may not exist in practice, it can still be considered as a meaningful RC from the perspective of SEEA EA (Keith et al., 2020). In the following we will consider this fictive state to be a third type of RC, called ***anthropogenic reference condition***.[37] Accordingly, an anthropogenic RC can be defined as the (fictive) state of heavily transformed anthropogenic ETs (e.g., agricultural, urban) to which it is meaningful to compare other (existing) states.[38]

Finally, no matter which RC type is selected, the assumptions "defining" the reference should be clearly and explicitly documented (Keith et al., 2020). A detailed specification of the RC is vital for the transparency and integrity of the accounts, and it also helps the quantification of the reference in the next step.

---

[35] This option is described as "historical condition" in SEEA EA Annex 5.2 Table 5.8 (cf. also SEEA EA A5.1).

[36] In urban ecosystems such targets and thresholds are often set at the local (municipality) level, making them difficult to use for national or EU-level assessments. And the anchoring of the target / threshold values to the indicator scale may also be challenging.

[37] This is closely related to the concept of "*best attainable condition*" discussed in SEEA EA and the EU methodology (Vallecillo et al., 2022). Nevertheless, the emphasis on *attainability* can be subjective and prone to conflicting interpretations, so we propose to use "anthropogenic RC" instead.

[38] As a matter of fact, an *anthropogenic RC* reverts the "normal" direction of inference (here the RC is derived from the RLs, and not the other way round, which should happen normally), and this also questions how much such an RC will be able to fulfil its original "purpose" (i.e., aligning the upper scale endpoints of the different indicators – see e.g., Table 2). Nevertheless, an *anthropogenic RC* built on relevant policy targets / thresholds may still provide a more meaningful basis for comparisons than a completely fictitious state.

6.2 Sample-based methods for setting upper reference levels

Due to its uniqueness and its clear meaning, the RC of an ET has a special position among any other possible "references". Accordingly, for most of the possible condition variables, the RL(s) corresponding to the RC are also particularly important: they anchor the different variables to the same state of the ecosystem, thus harmonising the "meaning" of the individual condition indicators with each other (SEEA EA 5.69). These RLs actually operationalise the (often abstract and textual) definition of the RC, connecting the "ideal", "pristine", or "historical" (etc.) state in the definition to concrete numeric values of concrete condition variables. Accordingly, these *upper RLs* anchor the *upper endpoint* (X_{100})[39,40] of all condition indicators to the *same "ideal" state* of the studied ET: the reference condition.

In line with their importance, there are many methods that are specifically designed to address upper RLs. Most of the methods proposed in the ecological literature for quantifying the "reference condition" of an ecosystem are actually methods for setting upper RLs. For most ETs and condition variables, these methods offer the best way for setting the upper RLs. Nevertheless, there are a few notable exceptions from this rule, including (1) the anthropogenic ETs (for which such methods are not useful), and variables with a negative directionality (for which X_{100} can often simply be set at a zero value, see Section 6.3).

The most important such method is based on a set of sites (i.e. a "sample") that are in (or close to) the RC, where the value of the different condition variables can be directly observed. In some cases, it may be possible to use statistical modelling to infer (estimate) the value of an upper reference level from the sample, even if most of the measurements represent disturbed states (i.e., relatively far from the RC). This technique is often called "pressure removal". Finally, in some special cases it may be possible to infer upper reference levels from a sample of historical or paleoenvironmental records. In the next three sections we will present these "sample-based" methods in detail, and then in Section 6.2.4 we present a critical issue that is common to all methods discussed in this section.

6.2.1 Direct measurements in reference sites

In principle, the most straightforward way to assign numbers to a reference condition is to "measure" it. For this, one needs a set of sites that are "*in reference condition*", where the

---

[39] Following SEEA EA, here we use the convention that the term "upper" RL always denotes X_{100}, even for negative directionality variables (like pollutant concentrations, invasive species abundances, etc). This counterintuitively implies that on the original untransformed scale the "upper RL" these variables will be lower than their "lower RL" (cf. UN et. al. 2024: 5.62), which underlines why a notation like X_{100} and X_0 is much less confusing for specifying reference levels than any textual terminology involving words like "upper", "lower", "optimal", etc.

[40] Although in practice such methods are also often used for setting other RLs that are "just" close to the upper endpoint (e.g., X_{80}).

value of the different condition variables can be (or had been) measured (SEEA EA A5.5).[41] Ideally, the set of reference sites needs to be a *statistically representative sample* of the ET over the study area where we want to use the RLs. It should also be representative for all relevant subtypes of the studied ET, for which separate RLs may need to be calculated (see Section 6.2.4). While this is a scientifically robust approach, there are two major challenges with it:

- Due to *long-standing human degradation*, it is often difficult or even impossible to find such reference sites (Stoddard et al., 2006, Jakobsson et al., 2020). Possible reference sites are also often clustered geographically and environmentally, making it difficult (or often impossible) to create a sample that is statistically representative for the ET as a whole (Chen et al., 2019a).
- And even if such a sample can be established, it will still contain some *natural environmental variability*, which will affect the EC variables, making it difficult to link the RC to concrete RL values.

If the RC still prevails over the study area, covering all (combinations of the) major environmental gradients and all major subtypes of the ET, then it might be possible to establish a representative set of reference sites (e.g., SEEA EA A5.5; Stoddard et al., 2006). The same set of reference sites can be used to measure the values of several condition variables. The measured values for each variable will form a statistical sample, and the RLs is then determined as a statistical parameter of this sample. More concretely, this is typically done by selecting a specific mean, quantile (percentile) or extreme value (e.g., minimum), and anchoring it to the upper endpoint of the indicator scale (X_{100}), or somewhere close to this point (e.g., X_{80}). The different national WFD implementations, for example, often (but not always) use both the median as X_{100}, and the 10th percentile as X_{80} (EUWI, 2020).[42] The EU forest condition accounts, on the other hand, used the maximum values of the variables in the reference sites to establish the upper RLs (X_{100}; Maes et al., 2023). Considering that inference on extreme values (e.g., maximum, minimum) is always less robust (e.g., more sensitive to small sample sizes), and that the best (most natural, pristine) sites are almost always underrepresented in a sample, a strategy going for a lower percentile/confidence level (e.g., the 20% or 10%) and anchoring it *a little below* the upper endpoint (e.g., to X_{80} or X_{90}) can also make sense from a statistical perspective. In addition, if (some of) the sites are already in a (slightly) disturbed condition, then it is better to avoid using central tendencies (e.g., mean, median) as X_{100}. Such slightly disturbed reference sites are also described as "minimally disturbed condition" in the literature (Stoddard et al., 2006, Keith et al., 2020).

Nevertheless, no matter what percentiles and indicator values are picked, unhandled environmental variability may completely spoil/mislead the estimation. Therefore, if this

[41] This is the main reason why this method cannot be applied in anthropogenic ETs, where meaningful RCs typically do not exist in practice.

[42] The practice of anchoring the median of "minimally disturbed" reference sites to the upper endpoint of the scale (as applied in some WFD indicators) may not be the best choice: e.g., this leads to EQR (=indicator) values above the median will not get distinguished any more – while an additional improvement near a minimally disturbed median will certainly still have some ecological meaning… (i.e., the truncation seems to happen at too low values, see also the footnote in Section 5.1)

method is applied, it is of vital importance to carefully check and then minimise/mitigate the amount of environmental variability in the sample using one of the approaches discussed in Section 6.2.4.

Unfortunately, due to the lack of suitable reference sites this method is often unavailable, but even if the reference condition of the ET is still available, collecting data from them is often considered to be expensive. Reference sites can, however, also be important components of long-term ecological monitoring networks and nature reserves, and a good network of reference sites (combined with a suitable monitoring programme) can provide many further benefits for conservation policy and ecological science beyond just setting RLs. And even if there are just a few "minimally disturbed sites", a systematic monitoring system can help to derive meaningful upper RLs, e.g., by pressure removal.

6.2.2 Pressure removal

This approach, which is sometimes called *pressure removal*, relies on empirical models for linking EC variables to human disturbance gradients, and to predict (i.e., extrapolate, "hindcast") the values of the variables to the RC (Stoddard et al., 2006; Kilgour & Stanfield, 2006; Soranno et al., 2008; EC, 2011). In cases when there are too few (or no) sites "*in the RC*", it is often still possible to find sites that are *relatively close to the RC*, i.e., which are often also described as being in a "minimally disturbed condition" (Stoddard et al., 2006, Keith et al., 2020). If enough data can be collected from such sites (e.g., from an ecological monitoring network) then it might be possible to estimate the undisturbed values of condition variables using modelling techniques. For this the dataset needs to go beyond the relevant condition variables, and it also needs to cover all pressure variables that are considered relevant for the given ET in the studied region, from e.g., a representative monitoring network. Furthermore, the sites also need to be representative for the ET, including all of its relevant subtypes, and they also need to cover adequately long segments of each pressure gradient.

In such situations, it will be possible to fit statistical models to the data, which describes the relationship between the condition and the pressure variable(s). The exact type of the model and the modelling procedure should be carefully adapted to the specific situation (Soranno et al., 2011). The model(s) can then provide a predicted value for the variable(s) for a hypothetical "pressure-free" situation, together with confidence intervals. These confidence intervals can then be used for setting RLs in the same way as the percentiles in the previous method (see below).[43]

Statistical modelling can also be a useful method for quantifying a *seminatural RC*. In this case the sample should (also) cover the intensity gradient of the main historical human land use(s) of the ET in the study region, and then the models can be used to predict confidence intervals for the variables for a selected (historical) level of this historical land use.

[43] SEEA EA A5.6 also mentions process-based models (e.g., potential vegetation models) in this context, but such models rarely include the right variables in their outputs (as also mentioned there). The use of process-based models for quantifying a (semi)natural RC definitely deserves further scientific exploration.

There is however a critical weakness, which undermines the usefulness of pressure removal in many contemporary situations. As all statistical models, pressure removal models also rely on assumptions: i.e., that the condition strongly (and exclusively) depends on the current(ly available data on) pressures and land use, which may not be always true. Typically, the current condition of ecosystems is more determined by their historical past (long term effects of historical pressures and uses), which are unfortunately as difficult (impossible) to assess as the historical values of the condition variables themselves.

6.2.3 Paleo-environmental data

In the case of some ETs and EC variables it might also be possible to collect data from historical, already disappeared "reference sites", e.g., with the help of sediment cores (Bennion et al., 2011; Burge et al., 2023) or other historical records. Careful interpretation of such records can provide limited insight into the pristine conditions of the ecosystem type (Stoddard et al., 2006). Like in the other cases, the sample (of e.g., sediment cores) should be representative of the whole study region and all relevant "subtypes" of the studied ET, which may be challenging to ensure/verify in practice. Information about historical ecological conditions can also be inferred from "ex situ" sources, like museum collections, paintings, manuscripts (SEEA EA A5.8), in which case the spatial attribution of sample can already be challenging. This approach is also restricted to specific variables (e.g., community composition, nutrient concentrations, or disturbance frequencies) and indirect measurements (proxies), while for other types of variables paleo-environmental data may be completely unavailable.

6.2.4. Addressing the impact of natural variability

Any network of reference sites representing a natural or semi-natural RC will inevitably contain some "*natural variability*" in most of its characteristics and ecosystem condition variables. This variability is caused jointly by the environmental diversity of possible sites across the ecosystem type being assessed, as well as the internal dynamics of the RC state (due to e.g., natural patchiness or an internal disturbance-regeneration cycle). To be able to give information about the degree of departure from the RC, condition variables need to distinguish *anthropogenic impacts* from this *natural variability*.[44] This may be particularly challenging in large and highly heterogeneous ecosystem accounting areas, with long environmental gradients (Stoddard et al., 2008; Miller et al., 2016). There are several methods for "minimising" the impacts of natural variability on RL estimations:

- *Identification of* ***ecosystem subtypes***: A common approach to reduce the amount of "natural variability" that can be associated to the RC is to divide the ET (in the study region) into a number of smaller, more homogeneous *ecosystem subtypes*. This may be performed using a hierarchical ET typology (e.g., EUNIS), the lower levels of which

[44] This can also be seen as a "signal to noise ratio" that characterises the particular variable. If the natural variability of a variable (the "noise") is larger than the variability caused by anthropogenic degradation (the "signal"), then that variable will not be able to give any useful information to policy. Such variables should be either "improved" or discarded (cf. the criterion of *directionality* in Czúcz et al., 2021b).

correspond to ecologically more homogeneous units, for which it is often significantly easier to identify meaningful RLs. The drawback of this option is that there may not be enough data for (some of) the subtypes (the minimally disturbed subsample is too small or missing), and that the information on ET subtype may be missing for some parts of the study region (i.e., there is no ET map at the subtype level). If this approach is chosen, then the same set of subtypes should be consistently applied across all of the EC variables. Furthermore, all subtypes should have their own "natural states", i.e., "subtypes" defined as *degraded states* of their parent ET[45] (e.g., "degraded grasslands", "non-indigenous tree plantations") cannot be used for this purpose (= for reducing the impacts of *natural* environmental variability).

- *Identification of* ***ecoregions***: A similar approach to reduce the "natural variability" of the RC is to divide the study region (accounting area) into a number of smaller, environmentally more homogeneous *subregions* (e.g., "*ecoregions*": Stoddard et al., 2008; Pereira et al., 2016; or "*homogeneous ecosystem areas"*: Vallecillo et al., 2022). In several parts of the world there are ready-made general purpose "ecoregion maps" available, which are typically created based on a biogeographic analysis of species distributions (e.g., Metzger et al., 2005; ETC/BD, 2006; Morrone, 2018). Nevertheless, it is also possible to make new ones for a particular ecosystem condition assessment/accounting study. This approach is fundamentally similar to the previous approach (i.e., the ecoregions are used to outline ecosystem subtypes by spatial overlay), and the success of both approaches hinges on the ecological relevance of the classification, as well as the "accuracy" of the data (how precisely and consistently the different ecosystem units can be linked to the resulting categories).
- ***Redefining the condition variables ("back to square one")***: sometimes the best option for reducing the environmental heterogeneity is changing the variable itself. Practically this means going back to the step of *variable selection* and making changes in the way the variable is defined (i.e., how it is calculated from the data at hand).[46] Sometimes there are simple changes possible which make the variable much more "homogeneous" over a larger geographic range. For example, for a variable representing the population density of a territorial species for which the size of territories has a clear north-south gradient, it may be better to express the variable as population density relative to the territory size rather than density over unit area. It is also possible to use statistical models to remove the effect of major environmental gradients from an EC variable (sometimes called as ***gradient removal***). This approach also needs a set of *reference sites* for fitting a *statistical model* with the relevant natural environmental gradient(s) as predictors, and this model is then used to transform the variable itself (e.g., Pereira et al., 2016; Chen et al., 2019b).[47] There are, however, also several possible issues with changing the variables, e.g., the adjusted

[45] More exactly, it should be avoided to define *ET subtypes* based on condition variables, or any other variables that are closely correlated with condition variables (e.g., pressures, protection…).

[46] This step is therefore sometimes also called "variable *development*" (e.g., Czúcz et al., 2021b, Allain et al., 2018), which emphasises that the task is more than picking items from a list of available alternatives. In a SELINA context this step is covered by Task 3.2.

[47] More concretely: the model is used to make predictions for the sites/locations where the EC variable was measured, and these predicted values are then subtracted from the variable, which results in an adjusted EC variable with reduced natural variability.

variables are more difficult to interpret, and the adjustment may "destroy" other meaningful RLs. SEEA EA condition accounts also have a general preference for variables that are closer to the raw data over highly refined (modelled) variables (cf. "simplicity" in Czúcz et al., 2021b). Accordingly, there is a trade-off between the need for removing gradients/heterogeneity, and the need for "staying simple" – and in some cases the optimal balance of these two conflicting principles might be difficult to find. In any case, there are relatively few scientific studies that explore (and properly document) the application of variable transformations on EC variables for variance reduction, which makes it difficult to provide a detailed guidance on this topic.[48]

6.3 Methods for any reference levels

In Chapter 6.2 we discussed the methods most suitable for assessing the "upper" RLs for a condition variable. Nevertheless, the upper RLs are not the only ones that need to be set, and even for them there are some situations in which they cannot be applied (e.g., anthropogenic ETs). In this chapter we discuss methods that can be used for any types of RLs, including upper ones.

6.3.1 Prescribed levels (special values)

Prescribed levels, as a method, is based on the idea of reusing known *special values* of a concrete EC variable as a RL (SEEA EA A5.10). In this context, a "special value" is a value with a concrete well-defined meaning to which it makes sense to compare other values to (SEEA EA 5.65). Such special values might be connected to absolute mathematical or biophysical thresholds (e.g., natural zero values), thresholds recognised in the scientific literature (e.g., regime shift thresholds), or by policy (e.g., policy targets). *Such special values can be used as "thresholds" for good condition, sustainable levels of human activities or other targets.*

A particular challenge related to the use of special values is the identification of the "right role" for the resulting RLs (i.e., the indicator value, endpoint, class boundary, etc., where these RLs can be meaningfully anchored). In some cases, this is easy or even trivial, e.g., for *zero values* (or other "*natural scale endpoints*") this anchoring should always happen at one of the endpoints of the indicator scale, i.e., X_0 or X_{100}). Nevertheless, for many other special values (e.g., policy thresholds, socio-ecological thresholds finding the right anchoring might be far from trivial.

Here we summarise the most important types of special values that can be used as RLs for an EC variable:

- ***Natural scale endpoints***: As discussed in Section 5.2, the continuous indicator scale has two endpoints, and the values of the condition indicator are constrained by these endpoints (i.e., they cannot go above 1 or go below 0). Many of the original (unscaled)

[48] This also means that SELINA case studies can have a major role here.

EC variables also have one such endpoint, beyond which the value of the variable cannot go. This is most commonly a ***natural zero***[49] value, which sets the lower endpoint of the scale of the original variable, indicating typically the lack of something (e.g., a very silent spring with zero birds: the number of the birds cannot go below zero, or zero concentration of a substance). In some cases, there may be two natural scale endpoints on the scale of the same variable, a lower and an upper endpoint (e.g., variables measured in % often have endpoints at both 0% and 100%).[50] Note that the order of RLs may, however, be reverted, depending on the directionality of the variable (Fig 5.2).

- ***Policy thresholds***: Policy often formulates expectations for specific variables in specific contexts in the form of a concrete value (e.g., thresholds, targets, standards….). Such policy thresholds (incl. targets, standards) can also yield meaningful RLs, which should be used if they exist (Kervinio et al., 2023). Such thresholds may include, for example, *safety* thresholds/standards (e.g., for pollutants), an acceptable extinction *risk* of a population (in population viability analysis), *regulation* thresholds/standards (e.g., for urban green / impervious surfaces), policy *targets* (e.g., for the share of landscape features in agricultural landscapes), etc. Appropriate *anchoring* and *validity* may be two critical questions for the application of policy thresholds as RLs. In many cases the detailed text, context, or the purpose may give hints about a meaningful anchoring (e.g., for a threshold that separates "good condition" from "not good" X_{60} is the default choice in several existing condition frameworks, like WFD or IBECA, Jakobson et al., 2020), but the "right" anchor (the meaning of the RL) can also be unclear, or ambiguous. Nevertheless, the typical goal of policy thresholds is to divide the "cases at hand" into two groups (the ones below thresholds from the ones above threshold), which means that such thresholds should always be used as mid-range RLs (and never as X_0 or X_{100}). Another critical question for policy thresholds is their spatial and thematic *domain of validity* (i.e., the list of ecosystem subtypes and spatial regions for which they can be meaningfully applied as RL), which can sometimes also be inferred from the context.
- ***Scientific thresholds***: In few cases specific thresholds for the studied variable might be found in scientific analysis, models or literature. This might include sustainability thresholds (e.g., fish stocks), minimum viable population sizes, regime shift thresholds (tipping points), etc. For some condition variables there are readily available scientific thresholds, for example, for population sizes of larger mammals and commercially exploited fish stocks, which can be used in the same way as policy thresholds.[51] In

---

[49] Such a scale endpoint is "natural" in the sense that it is an inherent & essential part of the variable itself, but this value does not have to manifest in nature. A zero value can also be "non-natural" (i.e., arbitrary), e.g., in the case of the Celsius or Fahrenheit temperature scales.

[50] Variables measured in % (a.k.a. "absolute scale" in the literature; cf. Ellis, 1966, Brunsdon, 2018) are often also dimensionless. In the very rare cases when both endpoints of such a variable make sense as lower and upper RLs, the variable may not need any rescaling (i.e., the rescaling function will be either the identity function or the $Y = 1 - X$ function).

[51] Policy thresholds also often build on previous scientific evidence, so the demarcation between these two types of thresholds may not be absolutely clear. A new scientific threshold with broad support in the scientific community will typically also get endorsed by policy in a short time. If a scientific threshold is not documented clearly (or it is documented for something else than what we want to use it for, e.g., just a specific subtype of

addition, the two caveats (anchoring, validity domain) that were discussed above for policy thresholds, are similarly large challenges for scientific thresholds, too.

- ***Expert thresholds***: In some cases, there might be some evidence that a given variable exhibits some special values, which have not been properly documented yet. Such values may be related, for example, to precautionary thresholds, or suspected nonlinearities in the dynamics of the studied system that can be connected to the concrete variable by expert judgement (Jakobsson et al., 2020; EC, 2011: Annex IV). Nevertheless, the use of expert judgments for RLs is no less controversial than the use of expert judgements for EC variables (discussed in Section 5.1) in a SEEA EA context, as both forms of expert inputs are often considered too subjective for the purposes of formal accounts (SEEA EA A5.11). Well-structured survey methods however exist and are successfully used by professionals (O'Hagan, 2019).

6.3.2 Reference level transfer

In some cases, it may be possible to "transfer" a well-defined RL (e.g., a policy threshold) from one condition variable to another one using statistical methods. This requires a relatively large amount of data, with both variables measured on a representative set of occurrences of the studied ET. With such data, it is, in principle, possible to fit a statistical model capturing the relationship between the two variables. If the relationship is strong enough,[52] it is possible to make a statistical "prediction" to transfer the existing RL from its variable to the other one.

This approach is heavily used in the WFD, especially in major intercalibration exercises, where Member States sharing a biogeographical region use it to compare and harmonise the different variables developed in their national WFD implementations. During the intercalibration the same set of water bodies are assessed using all of the different variables, and then the reference levels of the matching pairs of variables are compared using RLs transfer (Birk et al., 2013).

A similar potentially important future use case for RL transfer methods is connected to the future updates of existing condition variables (e.g., because new data sources become available). In such cases a careful RL transfer from the old variable to the new variable may be the best way to avoid unnecessary artefacts in the accounts.

Nevertheless, intercalibration and indicator updates are relatively special situations where the compared condition variables are expected to be highly correlated. In a general case between two arbitrary condition variables this cannot be expected. If suitable data are available, the significance of the statistical relationship between two arbitrary variables can, in principle, be improved by including further covariates, e.g., pressures and environmental variables. The same approach can be also used for transferring a threshold (e.g., a policy target or a safety standard) from a pressure variable to a condition variable. This can be

---

the ET, or a different ET, or another geographic region, etc) then it should rather be considered as an expert threshold.

[52] According to the selection criteria presented in SEEA EA Annex 5.1 (and Czúcz et al., 2021b), there should be little correlation between any pair of condition variables.

relevant because it is often easier to find (policy) thresholds for pressure variables than for proper condition variables. Nevertheless, to our knowledge, there have not been any scientific studies yet that would have explored the applicability of RL transfer as a general method for identifying reference levels in condition variables. This might thus be an interesting topic for future scientific studies.

6.3.3 Inter- and extrapolation methods

Established policies and frameworks (e.g., WFD, MSFD, the Norwegian IBECA method), often prescribe which RLs need to be used for an indicator. For example, the WFD typically prescribes the use of *five RLs* (X_0, X_{20}, X_{40}, X_{60}, X_{80}, and X_{100}: see more in Section 7.2), while the Norwegian IBECA method prescribes *three RL roles* (X_0, X_{60}, and X_{100}: Jakobsson et al., 2021). In such policy frameworks it is a common practice[53] to recommend a sequential approach for identifying the RLs:

- anchor the scale near the *two endpoints* (e.g., at the RC and a suitable zero value),
- look at the *other RL roles* and try to find suitable *special values* for them,
- and finally, quantify the *remaining RLs* using *linear interpolation*.

Mathematically speaking, two points already determine a line in a space, and any linear function ($Y = f(X)$) can also be unambiguously determined by specifying two of its points (say, $Ya = f(X_a)$ and $Y_b = f(X_b)$). This linear function will then assign a Y value to any specific X value in between (say X_c), and the Y value assessed this way ($Y_c = f(X_c)$) can be seen as a meaningful "guess" for what Y "should" be at that point, as long as there is no further information available (Fig. 6.1a). This approach can be used for any *mid-range RLs*, and it is widely used, e.g., in the WFD, where it is the default "fallback option" to set the missing class boundaries after other options have failed (Fig. 6.1b).[54]

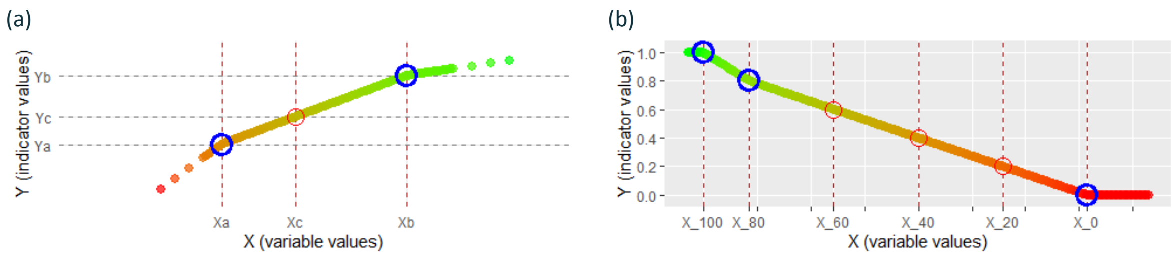


Figure 6.1. Interpolating estimated values of "missing" mid-range RLs (red circles) from previously known RLs (blue circles): (a) a general illustration, (b) a WFD-inspired example

[53] The WFD and the Norwegian IBECA frameworks largely follow this approach. (The WFD process is described in detail e.g., in EC (2011), and more briefly in Section 7.2; and the details of the IBECA approach are described in Jakobsson et al., (2020, 2021))

[54] See e.g., "Step 8" in Annex IV of EC (2011) (https://op.europa.eu/s/zJRT)

Nevertheless, it is not only mid-range RLs that can be estimated using this technique. With the help of two previously known RLs a meaningful estimation can also be given for a missing *endpoint RL* (i.e., X_0 or X_{100}) using linear *extrapolation*. This technique can be very useful if one of the endpoints can be matched to a natural zero, and at least one mid-range RL (e.g., X_{60}) is known. For example, for a variable with a natural zero corresponding to "maximally degraded condition" (X_0) it is often possible to find a suitable value for a missing X_{100} (i.e., the RC) starting out from a known mid-range RL (e.g., an X_{60}) by extrapolating the linear segment to $Y = 1$ (Fig. 6.2a). Similarly, in the case of a pollutant (or any other "*negative variable*") where the natural zero corresponds to the RC (X_{100}), it is often possible to extrapolate a suitable "maximally degraded value" (X_0) based on and a known mid-range RL (e.g., an X_{60} corresponding to a safety threshold separating "good" and "not good" cases; Fig. 6.2b). This can be helped by investigating maximum values in time series data and using expert elicitation.

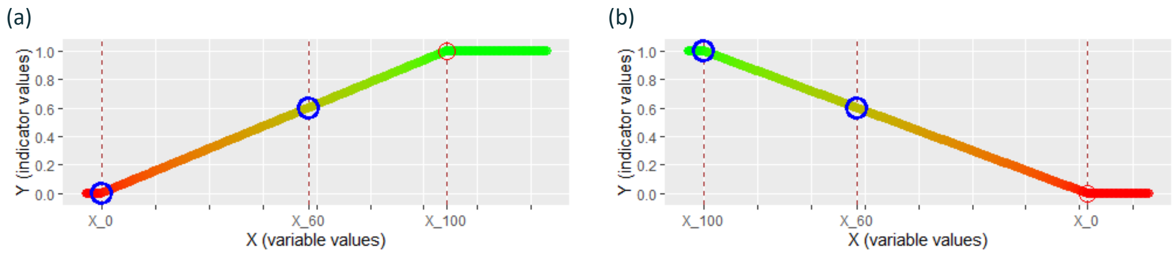


Figure 6.2. Extrapolating suitable estimations for "missing" endpoint RLs (red circles) from previously known RLs (blue circles): (a) deriving value that can suitably characterise the RC (X_{100}) for a variable with *positive* directionality; (b) deriving a suitable "maximally degraded value" (X_0) for a variable with *negative* directionality (e.g., a pollutant).

6.4 Data-driven methods

Condition assessments need to rely on external (theoretical) considerations that reach beyond the data themselves, and anchor them to something absolute, which goes beyond the data of the current study. These "external" references are then able to establish comparability across different studies performed in different regions over different time periods by different people. From the perspective of condition assessments this is a fundamental principle.

Nevertheless, there are several popular rescaling techniques, which perform a "relative rescaling" based on the data themselves, thus avoiding the use of external *references*. Instead, these *data-driven methods*[55] operate with the starting values (~*baselines*) or the maximum/minimum values (~*benchmarks*) of the data cube under analysis. The most

[55] Of course, any quantitative scientific method needs data. Here we use the term "data-driven" to label those methods which *exclusively* rely on the data cube for determining RLs (like e.g., an unsupervised clustering method). All of the previously discussed methods impose some sort of "theory" (information) on the data (e.g., via the selection of reference sites). Accordingly, these previously discussed methods are "supervised" by the external information (similar to, e.g., a supervised clustering method), so in a technical sense they are *not* data-driven.

important such method is *range-standardisation* (a.k.a. the “min-max” method, see e.g. Mazziotta & Pareto, 2022). This method compares the inner values of a data cube to the minimum and maximum values taken up in the entire data cube (using an equation very similar to Eq. (1) to perform the rescaling). Other variants of this method use other statistical parameters (such as 5% or 95% percentiles) as “reference levels”.[56] Nevertheless, if the dimensions of the data cube change (e.g. data for a new “accounting period” become available), the minima and maxima can change, which potentially ruins the rescaling (breaks the compatibility with SEEA EA, creates artefacts, and/or demands ad hoc coping strategies).

A similar group of data-driven methods is often used for rescaling time series. In these methods each data point is compared to its “own past” starting from a(n arbitrary) *baseline* year, using geometric means (see e.g. Rowland et al., 2021). While this solution has sophisticated mathematical foundations, it only offers a limited spatial comparability (each time series is “an island”, and the same indicator value (e.g., Y=1) can have a different meaning in the different “islands” depending on the starting value (i.e. the local baseline) and possibly other ad hoc circumstances. Furthermore, the rescaled values produced by such methods can easily go above 1, and there is no straightforward way to adjust the range to 0-1, as required by SEEA EA. Consequently, the comprehensive integration of such indicators into SEEA EA still demands substantial exploration and methods development.

Data-driven methods are relatively easy (and cheap) to implement, because they can be prepared from a “cookbook” following a “recipe”. They can simply just be applied to a data cube, sparing all the hassle of looking at the specificities of the concrete ecosystem types, variables, etc. They also avoid the hassle with the RC, and they jump directly to the “RLs” instead. This simplicity comes, however, at a price. The numbers produced by data-driven methods will have limited comparability across contexts (studies, regions, and often also timelines). Consequently, they are prone to *artefacts* whenever the “boundaries” of the data cube change.[57] So, while data-driven methods are relatively easy to implement, they are more suitable for single *one-time* scientific studies, than for long-term international policies, including accounting.

[56] Note that these min-max values are not real *references*, we only use the term *reference level (RL)* here, so that we could compare these methods with the previously discussed ones.

[57] This can easily happen if, e.g., an assessment is spatially or temporally expanded, “shrunk” (subsetted), or e.g., if the spatial scale is changed (smaller pixels are aggregated into larger pixels). In general, ecosystem accounting expects data cubes to be flexible and extensible (e.g., with each new accounting period).

7 Examples from existing approaches

In this chapter we provide examples for the framework outlined in Chapters 5–6 based on existing approaches from different ecosystem types. A general overview of the approaches was harvested through a survey among the SELINA partners representing (almost) all of the ecosystem types of the project (forests, grasslands, heathlands, wetlands, freshwater ecosystems, marine environments, agroecosystems and urban environments). We did not include examples from coastal terrestrial habitats due to their great overlap with other ecosystem groups (examples from e.g., the grasslands and wetlands chapters may be applicable also to coastal terrestrial ecosystems) and fluctuating definition under different contexts. In sections 7.2 and 7.3 we present the Water Framework Directive and the Marine Strategy Framework Directive in more detail, as the two major currently existing ecosystem condition frameworks in EU policy.

7.1 Existing approaches to defining ecosystem condition

The rescaling framework presented in this report is highly theoretical and conceptual, however, its constituent elements are already well used in many different contexts in Europe and the wider world. In particular, there is an extensive amount of experience about defining ecosystem condition in Europe. A survey among SELINA partners indicated that all the approaches presented in this report are already applied at least in one of the included ecosystem types and there are well-established practices in classifying ecosystem condition according to different environmental policy objectives. Table 7.1 shows a summary of the SELINA survey responses, and the most applicable examples are highlighted in sections 7.1.1-7.1.8 for further context.

Table 7.1 shows that eight of the nine broad ecosystem types already have operational classification schemes in some parts of Europe. The number of condition classes varies in terms of different variables, being highest in the forest ecosystem (10), relatively low in grasslands and heathlands and typically only two (or three) for the marine ecosystems where the main policy (Marine Strategy Framework Directive, MSFD) requires only an assessment of good / not good status. The condition variables in the wetland examples do not have a classification but use the continuous 0-1 index from “worst” to “best”.

We collected several examples of both natural and seminatural reference conditions from most ecosystem types (Table 7.1). There were even some examples of anthropogenic reference conditions (despite the difficulty of even defining what this means - see section 6.1) from urban, agricultural and even marine / freshwater ecosystems (e.g., heavily transformed waterbodies like coastal industrial or port waterbodies).

The most common approaches to defining reference levels are the use of reference sites and the expert-based approach. Historical and paleoenvironmental data are also used often. Natural scale endpoints were mentioned for five ecosystem types, but this may be an underestimate in this survey as this method is typically used to define the worst (e.g., zero population left) or the best conditions (no presence of pollutants or anthropogenic pressures). Scientific thresholds and policy thresholds were also mentioned for five ecosystem types. The interpolation and extrapolation approach was applied in two ecosystem types. The reference level transfer was mentioned only for the marine ecosystems, though it is often also used in WFD intercalibration.

Data-driven methods were mentioned three times. This approach is often the least reliable as it only uses the whatever data set as its basis and typically selects a statistical level (e.g., a percentile) or a baseline year. The method can be used in certain cases if a good justification is presented or it can be combined with other approaches such as the historical baseline, where a statistical level is given for a historic data of a suitable time period.

Table 7.1. Classification schemes of ecosystem variables in different ecosystem types. Reference condition type N=Natural, SN = seminatural, A=anthropogenic.

	Agroeco-systems	Grass-lands	Heath-lands	Forests	Wet-lands	Fresh-waters	Marine	Urban
RC type	SN, A	N, SN	N, SN	N	N	N, A	N, A	A
Sample-based methods								
Reference sites[1]	✓	✓	✓	✓	✓	✓	✓	✓
Paleoenvironmental / historical baseline[2]	✓	✓		✓	✓	✓	✓	
Other RL methods								
Natural scale endpoints3	✓		✓		✓	✓	✓	
Policy thresholds[4]	✓			✓		✓	✓	✓
Scientific thresholds[5]	✓	✓		✓		✓	✓	✓
Interpolation & extrapolation[6]							✓	✓
Reference level transfer[7]							✓	
Expert thresholds[8]	✓	✓	✓	✓	✓	✓	✓	✓
Data-driven methods[9]	✓						✓	✓
Number of RC classes	3-5	2-4	2-3	5-10	0	3-5	2-6	3-5

[1] **Reference site(s)**: the use of measured data from selected sites that are assumed to be in (or close to) the RC to set the RLs. Not available for heavily modified (anthropogenic / urban) ecosystems and certain variables (e.g., pressure metrics).

[2] **Paleoenvironmental / historical baselines**: estimating RC based on data from already disappeared "reference sites", e.g., with the help of sediment cores, historical records, or sources like museum collections, paintings, manuscripts.

[3] **Natural scale endpoints**: values corresponding to either 1 or 0 on the indicator scale. This is often a "natural zero" value (e.g., a "very silent spring" with zero birds, or lack of a pollutant).

[4] **Policy threshold** values usually denote policy/management targets or regulation levels**.**

[5] **Scientific thresholds** relate to thresholds recognised in scientific literature (e.g., sustainability thresholds, minimum viable population sizes, regime shift thresholds).

[6] **Interpolation & extrapolation** methods are often used in established policies / frameworks that prescribe a specific number of RLs to be used for each indicator (e.g., WFD status classes). If at least two RLs are defined, the rest of the boundaries can be anchored to the existing RLs**.**

[7] **RL transfer** refers to cases where an existing RL for one variable can be "transferred" to another variable due to their close correlation.

[8] **Expert thresholds**: Thresholds identified by expert judgment that are used as RLs for rescaling. This method can relate e.g., to precautionary thresholds, or suspected nonlinearities in the dynamics of the studied system that can be connected to the concrete variable by expert judgement.

[9] **Data-driven methods**: Methods using statistical parameters of the "data cube" (e.g., minima/maxima, percentiles, starting values) for rescaling, instead of proper "theory-based" RLs.

7.1.1 Forest ecosystems

Reference condition

In the case of forests, reference sites representing undisturbed or minimally disturbed forest ecosystems are often considered to be the best way to define and quantify natural reference conditions. The experimental national forest account of Spain (based on the SEEA EA framework) used this approach to identify the reference conditions for the six major forest types of Spain based on reference sites in protected areas and forests that have not experienced land cover changes since 1970 (corresponding to the beginning of Spanish carbon inventories) (Arets et al., 2019, Bruzón et al., 2023). The protected sites were determined to be in a natural or near-natural condition, as they are classified under International Union for Conservation of Nature (IUCN) categories I (strict nature reserve and wilderness area) and II (national park). Such areas are specifically designated for biodiversity protection and the conservation of species and habitats (Dudley, N. 2008).

Setting reference levels using sample-based methods

Table 7.1.1.1: Examples of approaches using reference sites or historical / paleoenvironmental records to set reference levels for condition variables in European forests

Approach	Condition variable	RC type	RL role	Context	Area
Reference sites: primary forest + IUCN category I and II protected areas + no land cover changes since 1970	NDWI, NDVI, GPP, Species richness Species richness (birds / flora), SOC, FAD, Tree cover, Naturalness Index	Natural	X_{100}	DP 01 (Spanish SEEA experimental accounts)	Spain
Historical records of pre-industrial ozone/nitrogen levels	AOT40f (Ozone), nitrogen depositions	Natural	X_{100}	DP 01 (Spanish SEEA EA experimental accounts)	Spain
Natural RC sites chosen based on forest biodiversity, productivity, and ecological integrity	NDWI, soil organic carbon, threatened bird species diversity, above ground biomass, net primary productivity, forest connectivity percentage	Natural	X_{100}	TS 09 (LIFE IP AZORES NATURA)	São Miguel, Azores, Portugal
Protected forests	Tree Cover Density, NDVI	Natural	X_{100}	German SEEA EA experimental accounts	Germany

The experimental national forest account of Spain derived the upper RLs (X_{100}) for variables like Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index

(NDVI), Soil Organic Carbon, Tree Cover, Forest Area Density, Naturalness Index and Gross Primary Productivity (GPP) using values gathered from reference sites in primary forests, IUCN category I and II protected areas or forests with no land cover changes since 1970, representing minimally disturbed forest condition (Bruzón et al., 2023). For tree cover, NDVI and GPP the upper RL was set from the 90-95th percentile of the reference values, and for species richness (forest birds and flora) the maximum and minimum values derived from the reference sites were used for indicator calculations. Additionally, historical records of pre-industrial levels of ozone and nitrogen depositions were used to set the upper RL (X_{100}) for these pressure variables (Bruzón et al., 2023).

In the Forest Condition Index (FCI) developed for the island of São Miguel in the Azores, the RC sites were selected to represent forests in a desired or ideal state for the transformed condition variables, which relate primarily to forest biodiversity, productivity, and ecological integrity. The sites were delineated based on habitat class, species composition, and structural relevance. Areas strongly influenced by disturbances like strong winds, landslides, or tree cutting were excluded. Due to the limits of the available data and scarcity of certain forest types, some criteria had to be relaxed and the FCI calculation could not take into account the presence of invasive species. Each condition variable was then rescaled on a continuous scale from 0 to 1, where higher values indicate better forest condition, and the upper RL (X_{100}) was defined as the 98th percentile of the variable's distribution within natural RC sites (Bruelheide et al., 2024), for the earliest year for which data are available across all variables, in order to ensure temporal consistency in the index.[58]

In the German ecosystem condition accounting, baseline optimal reference levels (X_{100}) for tree cover density and NDVI, intended to reflect the natural or near natural state of the ecosystem, were derived by comparing forest ecosystems to protected forests in each landscape region (Statistisches Bundesamt, 2023).

Other methods to set reference levels

Table 7.1.1.2: Examples for other methods used for setting reference levels for condition variables in European forests

Approach	Condition variable	Method	RL role	Context	Area
Policy threshold	Tree cover	10% Forest Cover Spain Policy / 20% Forest Cover Europe Policy	X_{60}	DP 01 (Spanish SEEA EA experimental accounts)	Spain
Scientific threshold	AOT40f (Ozone)	Ecological impact studies on forests	X_{60}	DP 01 (Spanish SEEA EA experimental accounts)	Spain

[58] For some ET subtypes there were very few potential reference sites (e.g., T1: broadleaved deciduous forest). In such cases the criteria for the ref. site selection was relaxed to include any existing areas from the island-wide forest inventory with no visible signs of abiotic disturbance.

Approach	Condition variable	Method	RL role	Context	Area
Expert thresholds	Proportions of native / invasive tree species, number of age and size groups	Expert thresholds based on a field study	X_{100}	TS 28 (TERMERD, MAES-HU)	Hungary
Expert thresholds	Proportion of non-native tree species	Expert thresholds based on a field study	X_{40}, X_{100}	TS 28 (TERMERD, MAES-HU)	Hungary

The Spanish experimental forest account used the Spanish national forest cover policy (10 %) and European forest cover policy (20 %) to set thresholds (X_{60}) for area covered by trees (Jennings et al., 1999; Bruzón et al., 2023). The threshold for ozone levels was determined based on ozone ecological impact studies in Europe (Skärby et al., 1998; Bruzón et al., 2023).

In the Hungarian MAES-HU condition mapping, composite indicators were developed for most ecosystem types (Tanács et al., 2022). In the case of the forests, one or two reference levels were defined separately for all the sub-indicators mainly by experts based on the experiences of an earlier national-scale field study of thousands of forest stands (Bartha et al., 2006). The results were checked against detailed field data in three mountain regions and some suggestions made for improvement (Zoltán et al., 2023).

Condition scale(s)

Table 7.1.1.3: Examples for discrete scales used for condition variables in European forests. (N: Number of condition/status classes distinguished)

N	Class names	Condition variable	Context	Area
6	Natural, Near-natural, Secondary, Transitional, Artificial, Plantation forests	Forest naturalness	Hungarian Forestry Law	Hungary
5	(5) most favourable… (1) least favourable	Composite forest condition indicator	TS 28 (MAES-HU)	Hungary
10	(10) Virginal natural system, (9) Natural system, (8) Sub-natural system, (7) Almost natural system, (6) Semi-natural system, (5) Cultural auto-sustainable system, (4) Assisted cultural system, (3) Highly intervened system, (2) Semi-transformed system, (1) Transformed system	Vegetation Naturality Index	TS 09 (LIFE IP AZORES NATURA)	São Miguel, Azores, Portugal

Deforestation and degradation are recognised as the major threats to forest ecosystems in the world (Li et al., 2018, Song et al., 2018). Following this reality, the examples we have collected of forest-specific ecosystem condition scales all describe the degree to which the ecosystem has been transformed from its natural state by human intervention.

The Hungarian Forestry Law features a discrete ecosystem condition scale of “forest naturalness”, which is used in national management decisions and legal constraints (Act XXXVII of 2009). On the other hand, the composite forest condition indicator established in MAES-HU for assessing Hungarian ecosystem condition on a national level takes into account more ecological aspects and presents a more simplified categorization from (5) most favourable to (1) least favourable (Tanács et al., 2022).

The “Vegetation naturality index”, developed for the vegetation and peat bogs habitats of the Azores islands, in the context of the LIFE IP AZORES NATURA project (LIFE17 IPE/PT/00010), has as many as ten discrete condition classes ranging from virginal natural systems to semi-natural cultural systems and highly pressured transformed systems (Dias et al., 2022). This index is particularly relevant as a descriptor of habitat type and condition, since the map legend units follow the typologies of the Habitats Directive, associated with their respective degree of naturalness, rather than the syntaxonomy of the vegetation unit.

7.1.2 Agroecosystems

Reference condition

Agricultural ecosystems are mostly, by definition, heavily transformed and systemically controlled for the cultivation of crops or raising animals.[59] Trying to agree on a suitable textual definition for an “agricultural” reference condition, or to “prototype” such RCs using samples of existing croplands or plantations may not be a meaningful or even feasible approach (see section 6.1). Instead, it may be better to begin with identifying suitable reference levels for each individual condition variable. Often, there are plenty of suitable “special values” thresholds available for agricultural ecosystems: management targets, policy / regulation thresholds, natural scale endpoints, and even reference levels, and the policy use of such thresholds ensures that they can be used for meaningful comparisons (Kervinio et al., 2023).

In some cases, it may be possible to use extrapolation or proxies to define a suitable upper RL (“best attainable” condition). For example, in MAES-HU (which had the ambitious goal to map and evaluate the status of all Hungarian ecosystems), the reference condition for agricultural habitats was defined using the standardized relative richness of the ecosystem’s characteristic bird species, with the logic that bird diversity would reflect the state of the related ecosystems (Tanács et al., 2024).

[59] While many semi-natural grasslands are also traditionally grazed by livestock, we have included these examples in the grasslands section (6.1.4)

Setting reference levels

Table 7.1.2.1: Examples of approaches for setting reference levels for condition variables in European croplands.

Approach	Condition variable	Method	RL role	Context	Area
Natural scale endpoints	Share of semi-natural (high-diversity) landscape features in agricultural land (%)[60]	Refers to a theoretical state where high-diversity ("non-productive") features are at 0%	X_0	SELINA TS 27	Lower Saxony, Germany
	Soil biodiversity	A situation where no indicator species are present in the soil	X_0	ENVASSO Project	EU
	Soil cover (%)	Refers to the (theoretical) state where soil cover is 100%	X_{100}	TS 27	Lower Saxony, Germany
Policy thresholds	Soil cover (%)	Refers to the requirement of min. 80% soil cover in the most sensitive season in the CAP 2023-2027 German Strategic Plan	X_{60}	TS 27 (CAP)	Lower Saxony, Germany
	Crop species diversity/richness	Requirement of at least 2 different crop species on arable land per farm (of a specific size) in the Regulation 1307/2013 (CAP)	X_{60}	TS 27 (CAP)	Lower Saxony, Germany
	Share of semi-natural (high-diversity) landscape features in agricultural land (%)[1]	Requirement of at least 4% of non-productive areas and elements in the CAP 2023-2027 German Strategic Plan	X_{60}	TS 27 (CAP)	Lower Saxony, Germany
Scientific thresholds	Soil bulk density (g/cm³)	The EEA Report No 08/2022 classifies soil bulk density from very loose to very impermeable with respective threshold values	X_{60} (and possibly also X_{20}, X_{80}, etc.)	TS 27 (EEA)	Lower Saxony, Germany

[60] For this variable there are several "natural scale endpoints" and "policy thresholds", which can be used for various RL roles (see discussion below).

Approach	Condition variable	Method	RL role	Context	Area
	Soil organic carbon (%)	Below certain SOC thresholds, soil stability, nutrient availability and available water capacity are limited	X_{20}, X_{40}	ENVASSO-project	EU
Expert thresholds	Proportions of areas under different uses (semi-natural / high value / fallow areas, proportions of areas cultivating maize or alfalfa, etc.; %)	Refers to corresponding literature of the interrelations between agro-ecosystem characteristics and biodiversity, according to which a minimum of 2% of semi-natural areas has a positive impact on biodiversity (X_{40}) and from 20% of semi-natural habitats, the condition is (very) good (X_{80})	X_{40}, X_{80}	TS 28 (MAES-HU)	Hungary

There is considerable environmental variability and heterogeneity across Europe (i.a., climate, soil, and topography). When defining reference conditions (RC) and reference levels (RL) for agroecosystem indicators, it is essential to ensure they are context- and location-specific[61]. This approach ensures that reference levels are ecologically meaningful and comparable, reflecting the natural variation in agroecosystem characteristics across Europe and improving the accuracy of condition assessments (see Section 6.2.4).

The natural scale endpoint approach is used in the contexts of the ENVASSO project (aiding the implementation of the EU Soil Framework Directive) and the SELINA Test Site located in Lower Saxony. The ENVASSO project makes use of the "natural zero" value of the indicator of "soil biodiversity", where a state of total absence of the target species corresponds to the lower endpoint X_0 (taking into account specific cases where the absence occurs due to natural soil characteristics, like the lack of earthworms in very acidic soils) (JRC, 2008). Additionally, the SELINA Test Site 27 tests the use of natural scale endpoints for several other variables. For the variable "share of semi-natural (high-diversity) landscape features in agricultural land" (measured in %), there are possible two natural scale endpoints (at 0% and 100%). The zero value could be used as the lower endpoint (X_0), representing "maximally degraded condition". However, the upper limit is not so useful – in agroecosystems, a 100% share of diverse areas would transform the agroecosystem into a different ecosystem type (and defeat the

[61] The Environmental Stratification of Europe, which combines six climatic and seven physiographic clusters, offers a useful framework for establishing RC and RL at the European scale (Metzger et al., 2005; 2018). However, for assessments at national or regional levels, more detailed environmental stratifications should be used where available. For instance, in Germany, the "Wuchsgebiete" classification provides a finer-scale combination of landscape and climate characteristics that can inform the setting of regionally specific reference levels (Gauer & Kroiher, 2012).

purpose). For the variable “soil cover” (i.e., the share of surface covered by vegetation during the most sensitive winter period), on the other hand, 100% represents the optimal situation that could serve as the upper endpoint (X_{100}), but using 0% as X_0 (reflecting a maximally degraded condition) is useful primarily for mathematical rescaling purposes. The threshold RL (X_{60}) could be set through national implementations of EU policies (BMEL, 2024), but it should be relatively conservative as even relatively small uncovered areas can cause severe negative impacts (e.g., through soil erosion). These examples illustrate how different methods for setting reference levels can be integrated for a single condition variable.

Policy thresholds in general are typically used to define a mid-range reference level, often X_{60}, as they usually aim at a tolerable or sustainable, but not the optimal, condition. However, the use of policy-derived thresholds can yield different reference levels depending on the underlying policy objectives and priorities. For example, there are two potential policy thresholds that could potentially be used for the RL role X_{60} for the variable "share of semi-natural (high-diversity) landscape features in agricultural land". For this variable the 2023–2027 German CAP Strategic Plan applies a main threshold of 4%, whereas the EU Biodiversity Strategy and the proposed Nature Restoration have set a target of 10%. While the 4% threshold reflects a baseline requirement intended to preserve essential ecological functions within productive agricultural systems, the 10% target signals a more ambitious ecological goal, aiming to restore biodiversity and landscape complexity. Accordingly, the SELINA Lower saxony test site (TS 27) applies the German threshold as X_{60} on a trial basis. This example also highlights that the concrete values used for the specific reference level roles should be aligned with the specific ecological and regulatory meaning of the potential candidate values (available policy thresholds, etc.).

Reference levels were mainly set by experts for MAES-HU agroecosystems condition sub-indicators (proportion of semi-natural areas / land subsidised under ‘High Nature Value Areas’ / maize / alfalfa / fallow land, number of cultivated crops, average parcel size, proportion of land included in an agri-environmental scheme (AES), proportion of green fallow, etc.), based on literature of agroecosystem characteristics and their relationship with biodiversity, pollinators and other taxa (Tanács et al., 2024). Expert judgement was often used in combination with other sources of data where the latter was not concrete enough to “stand on its own”. For instance, 20% of semi-natural habitats was estimated to have a very good impact on the condition, which refers to the RL role X_{80}. This estimation corresponds to available scientific thresholds, according to which at least 20% of native habitats is needed for good ecosystem functioning and ecosystem services provision (Garibaldi et al., 2020).

Even though it is difficult to define or identify sites in a "reference condition" in croplands, there are also sporadic examples for this approach in Europe. For example, the Dutch Soil Monitoring Network has a series of sites or farms chosen based on expert judgements, which are used as reference sites for soil biodiversity metrics (e.g., metrics related to microorganisms, nematodes, enchythraeids, earthworms, mites and springtails and soil processes). These variables can be applied in conventional or intensive farms, organic farms or even nature and parks (on a national scale). Because the sites in this network were selected using expert judgment, the resulting thresholds can also be seen as expert thresholds. (Rutger et al., (Eds.) 2008).

Historical data has also been used occasionally in agroecosystems for specific condition variables. For example, for soil moisture content, the German SEEA EA implementation compares historical records with long-term expected values (Destatis, 2023). The ENVASSO-project (Environmental Assessment of soil for Monitoring) suggested determining the upper reference level (X_{100}) for soil compaction using historical datasets from before the use of heavy machinery in agriculture began, or from contemporary reference sites where heavy farm machinery has never been used (Huber et al., 2008). For example, Håkansson et al., (1996) found that soil penetration resistances were 40% higher in croplands where heavy machinery was used to produce potatoes and sugar beets compared to fields where farm machinery has never been used.

Scientific thresholds for the “soil bulk density” variable are provided in the EEA (2022) classification, intended to be used for detecting harmful subsoil compression due to anthropogenic impacts. The value of 1.6g/cm³ (threshold between ‘normal’ and ‘dense’) is tested as X_{60} within the context of the SELINA Test Site (TS 27), however, the additional class breaks can potentially also be used for an even finer classification (e.g., as 1.2g/cm³ = X_{80}, 1.9g/cm³=X_{20}, etc). Instead of setting concrete reference values, the classes could as well be directly used for a discrete condition indicator scale, similar to the examples provided below.

Condition indicator scales

Table 7.1.2.2: Examples for discrete scales used for condition variables in European croplands. (N: Number of condition/status classes distinguished)

N	Class names	Condition variables	Context	Area
4	A (very good), B (good), C (poor), D (bad)	Scoring system of 7 aggregated variables	ELME (Estonian national MAES)	Estonia
3	Best case, medium, worst case	Land use intensity	Project ModBien	Germany
5	(5) most favourable, …, (1) least favourable	Composite agroecosystems condition indicator	TS 28 (MAES-HU)	Hungary
5	Very high, high, medium, low, very low	Field capacity	German DIN Standard	Germany

The condition categories used in MAES Estonia use pre-aggregated variables in a scoring system, which is an approach that presents considerable challenges to adapting the respective condition scale to other contexts. The discrete condition classes are derived from the sum of scores: A (10-13), B (7-9), C (4-6) and D (1-3). The scoring system includes organic farming (4 points), semi-natural grasslands in classes AB (3 points), semi-natural grasslands in classes CD (2 points), presence of forest edges or vegetation strips in the field (2 points), presence of small forested patches <0.1 ha (1 point), environmentally-friendly management measures (1 point), peat soils and thin calcareous soils (-1 point) (Helm et al., 2020).

Project ModBien aimed to model pollinator habitat suitability in Lower Saxony for developing sustainable regional landscape management. The condition of different habitat types, influencing the foraging and nesting suitability for pollinators, was estimated using land use intensity as a proxy on a discrete scale with three classes: best case, medium case and worst case (Hinsch et al., 2024).

There are also examples of discrete scales that are comparable to established EU policies, i.e., the WFD graded ordinal scale: the MAES-HU project used a five-step condition scale from one to five, where one was the least favourable and five the most favourable, to model cropland condition using variables like the number of cultivated crops, proportion of fallow land / alfalfa or green fallow / maize, average parcel size and proportion of protected or semi-natural areas: (Tanács et al., 2024). Similarly, the German Institute for Standardization measures field capacity using a five-step condition scale that ranges from very low to very high (GIS, 2020).

7.1.3 Heathlands

Reference condition

There are relatively few studies addressing the condition of heathland ecosystems, which is also apparent in the relatively low number of examples we could find for RCs or RLs used for rescaling condition variables in this ecosystem type. Much like grasslands, traditional management of European heathlands often requires maintenance like grazing or burning to prevent encroachment by higher vegetation. The heathland reference condition can therefore be considered as ‘seminatural’ with the key feature of “openness”, although some areas could also be kept open by natural wildfires. However, many areas also face problems from overuse and degradation.

In an interesting South African example, while the remaining sites of local indigenous shrubland (Renosterveld) are considered to represent the reference condition for the habitat type, paleo-environmental evidence suggests that their overall condition has changed over time due to increasing grazing pressure and more frequent wildfires. This example highlights the danger of a ‘shifting baselines syndrome’, if we define a habitat’s reference condition without access to sufficient long-term data (Forbes et al., 2018). Similarly, in a German example regarding “plaggen” soil heathlands, the reference condition was determined from patches of the habitat that had not been heavily degraded by the military (Heitkamp et al., 2008).

Setting reference levels using sample-based methods

Table 7.1.3.1: Examples of approaches using reference sites for setting reference levels for condition variables in European and African heathlands.

Approach	Condition variable	RC type	RL role	Context	Area
Reference sites representing indigenous shrubland (Renosterveld)	Coprophilous fungi, plant diversity	Seminatural	X_{100}	Forbes et al., 2018	Cape Floristic Region, South Africa
Undisturbed patch of plaggen soil heathland	Soil bulk density and pH, SOC, NP contents, roots density, microbial biomass & activity	Seminatural	X_{100}	Heitkamp et al., 2008	Heathlands, Germany

The South African study from the Cape Floristic Region used palaeo-environmental records (fossil pollen, coprophilous fungal spores and charcoal samples) taken from a local reference site, representing indigenous shrubland (Renosterveld), to track long-term changes in its vegetation, herbivory and fire regimes (Forbes et al., 2018).

The study from Germany examined the effects of different ecosystem restoration techniques to improve the condition of a heavily degraded plaggen soil heathland habitat. The upper reference levels (X_{100}) for soil bulk density and pH, soil organic carbon, nitrogen and phosphorus contents, roots density as well as microbial biomass and activity were derived from nearby patches of heathland that represented the same ecosystem but had avoided heavy use and degradation (Heitkamp et al., 2008).

Other methods to set reference levels

Table 7.1.3.2: Examples for other methods used for setting reference levels for condition variables European heathlands.

Approach	Condition variable	Method	RL role	Context	Area
Scientific threshold	Butterfly (*Phengaris alcon* & *Plebejus idas*)	Absence/presence of a key species is treated as a condition variable	X_0, X_{100}	N2000 evaluation	Wet heathlands, France

In the context of evaluating the conservation status of wet heathlands in France, the presence or absence of certain key butterfly species have been used as a binary indicator (Mistarz & Grivel, 2020), which by default are on an absolute (0-1) dimensionless scale with a clear normative meaning: presence denotes good and absence bad. This kind of variable is quite unique in the sense that it can directly be used as a dimensionless (0-1) condition indicator, without any additional rescaling steps.

Condition indicator scales

Table 7.1.3.3: Examples for discrete scales used for condition variables in European heathlands. (N: Number of condition/status classes distinguished)

N	Class names	Condition variables	Context	Area
3	Favourable (good), Unfavourable-Inadequate (poor), Unfavourable (bad)	Conservation status of wet heaths of EU interest	EU Habitat directive	France, temperate shrub heathlands / Atlantic wet heaths
2	Presence / absence (dimensionless)	Butterflies (*Phengaris alcon* & *Plebejus idas*)	EU Habitat directive	France, temperate shrub heathlands / Atlantic wet heaths
3	Favourable (good), Unfavourable-Inadequate (poor), Unfavourable (bad)	Vegetation cover, surface alteration	EU Habitat directive	Germany

The few heathland examples we could gather all use condition scales from the EU Habitat Directive, which include the three-part scale ranging from favourable to unfavourable and the dimensionless binary scale that relates to the presence or absence of some key species or process. The French evaluation of wet heaths of EU interest included the status of butterflies (*Phengaris alcon* & *Plebejus idas*), vegetation cover (woody, trees, structuring species, moor-grass and bracken) and surface alteration (Mistarz & Grivel, 2020). The German example regarding dry heaths involved the status of, i.a., coverage of the key species *Calluna* (<30%), stand structural diversity and a species index reflecting co-occurring vegetation (Schmidt et al., 2017).

7.1.4 Grasslands

Reference condition

The reference condition for grasslands ecosystems is typically semi-natural, as management through e.g., traditional grazing or burning is usually required to maintain open grasslands. As such, one key criterion of a grasslands reference condition is “openness”. For example, a Belgian study regarding the restoration of seminatural calcareous grasslands identified grassland occurrences that were deemed to be in the reference condition, defined as having existed for more than two centuries as open grasslands without woody encroachment, using historical maps, aerial pictures and field surveys (Piqueray et al., 2011). Another study from Spain examined the effects of traditional livestock grazing on seminatural grassland ecosystems using areas of (comparatively) high grazing pressure as reference sites to represent the reference condition (Serrano et al., 2024).

A study from Hungary investigated the correlation of bird communities and a vegetation-based naturalness metric for several different ecosystem types, including semi-natural grasslands (e.g., hay meadows). The "naturalness" of the landscapes was assessed by comparing the actual vegetation to the "ideal" reference state based on simple criteria regarding vegetation composition and structure. For semi-natural ecosystems, e.g., hay meadows, an equilibrium condition with traditional land use was considered as the ideal reference condition (Nagy et al., 2017). The national-scale MAES-HU condition mapping used a similar biodiversity-based approach to complement the other methods, where the indicator was the relative richness of characteristic farmland bird species. It measures the proportion of bird species actually present compared to the number of species experts would expect to be present in a high-value (semi-)natural grassland landscape (Tanács et al., 2024).

Setting reference levels using sample-based methods

Table 7.1.4.1: Examples of approaches using reference sites or paleoenvironmental records for setting reference levels for condition variables in European grasslands.

Approach	Relevant variable(s)	RC type	RL role	Context	Area
Calcareous grasslands reference sites known to have existed for more than two centuries	Soil NPK, plant species composition	Seminatural	X_{100}	Piqueray et al., 2011	Calestienne, Belgium
Mediterranean grasslands reference sites with stable grazing pressure	Soil NPK, SOM, microbial composition, enzymatic activity, forage quality	Seminatural	X_{100}	Serrano et al., 2024	Spain
Historical data from the Apennines (1966-1992)	Proxies of species composition, habitat structure, and landscape patterns	Seminatural	X_{100}	Carli et al., 2018	Apennines, Italy

The Belgian and Spanish studies both derived the upper RL (X_{100}) for different variables (i.a., soil nutrient contents, plant species composition, soil organic matter, microbial composition, enzymatic activity and/or forage quality) from direct measurements in reference sites representing open grasslands (Piqueray et al., 2011, Serrano et al., 2024).

In lieu of a contemporary reference site, a study from Italy used historical species data from the Apennines gathered between 1966-1992 to develop potential indicators to assess the conservation status of semi-natural grasslands habitats in the Habitats Directive context. For example, the change in species composition was quantified as the difference between historical habitat species pools (= X_{100}) and current communities (Carli et al., 2018).

Other methods to set reference levels

Table 7.1.4.2: Examples for other methods used for setting reference levels for condition variables in European grasslands.

Approach	Condition variable	Method	RL role	Context	Area
Scientific threshold	Coprophagous activity	Absence/presence of coprophagous activity is treated as a condition variable	X_0 / X_{100}	EU Habitat directive	Grasslands in France

Coprophagous activity was one of the variables considered during the evaluation of the conservation status of grasslands in French Natura 2000 sites. The simple presence or absence of this activity was used as a binary indicator (Maciejewski et al., 2015), which by default is on an absolute (0-1, dimensionless scale) with a clear normative meaning: presence denotes good and absence bad. This kind of example is quite unusual and does not require rescaling.

Condition scale(s)

Table 7.1.4.3: Examples for discrete scales used for condition variables in European grasslands. (N: Number of condition/status classes distinguished)

N	Class names	Condition variable	Context	Area
4	Favourable, Altered, Degraded, Bad	Conservation State of Pastures	EU Habitat directive	Grasslands in France
2	Presence/absence (dimensionless)	Coprophagous activity	EU Habitat directive	Grasslands in France

The grasslands-specific examples of condition scales are from the EU Habitat directive context of evaluating the condition of agropastoral grassland habitats (calcicolous grasslands, megaphorbiaes, mowed grasslands, pastures, juncion prairies, and molinion prairies) in French Natura 2000 sites. They used a four-step scale ranging from favourable to bad and a dimensionless binary scale that relates to the presence or absence of the key process of coprophagous activity (Maciejewski et al., 2015).

7.1.5 Urban ecosystems

Reference condition

Urban ecosystems are, by definition, highly anthropogenic and controlled, and typically their "green" and "blue" areas occur in patches of various sizes and can represent various ecosystem types (forests, grasslands, wetlands, freshwater and marine…). Trying to define a suitable "best attainable" reference condition for such environments may be difficult or even impossible (see section 6.1). It may be a better approach to begin with identifying suitable reference levels for each variable – often, there are plenty of "special values" available (management targets, policy, or regulation thresholds) (Browning et al., 2024). However, these thresholds are often connected to specific administrative areas (municipalities, counties, planning regions, etc.), which makes them problematic for large-scale (national, EU) ecosystem condition assessments.

There have been some indirect attempts to define urban reference conditions through, for example, the application of scoring systems such as the Blue Green Factor in Oslo, the Biotope Area Ratio in Berlin and the Polish Ratio of Biologically Vital Areas. These regulations prescribe a minimum score that construction projects should gain to be approved by the municipality. Points are normally assigned based on the presence of certain land covers (e.g., green areas, forest patches, lawns) and green elements (trees, permeable soil, etc.). The Norwegian Standard for the Blue Green Factor also differentiates the reference condition (i.e., minimum score) to be applied to different urban morphologies (Stange et al., 2022; Oslo Commune, 2023).

Setting reference levels

Table 7.1.5.1: Examples of approaches for setting reference levels for condition variables in EU urban areas.

Approach	Condition variable	Method	RL role	Context	Area
Policy thresholds	Share of urban green space / canopy cover	Share of green space in urban centres / clusters > 45 %, share of urban tree canopy cover > 10 %	X_{40}	NRR	EU
	Air pollutants concentration	Alignment with WHO Air Quality Guidelines	X_{60}	Air quality directive	EU
	Proportion of protected natural areas	Defined by taking into account Target 11 of the Aichi Biodiversity Targets (= 17%)	X_{100}	Singapore index on cities' biodiversity	Any city
Scientific thresholds	Minimum required share of tree canopy cover	Based on scientific evidence of human well-being in relation to tree cover	X_{60}	Astell-Burt & Feng, 2020, city planning rule of thumb 3-30-300	Any city
	Air temperature and humidity, floristic diversity, butterfly species richness, surface outflow	Correlation between Ratio of Biologically Vital Area (RBVA) and selected environmental features	X_{60}	Szulczewska et al., 2014	Warsaw, Poland
Expert thresholds	Proportion of natural areas in the city	The maximum score will be accorded to cities with natural areas occupying	X_{100}	Singapore index on cities' biodiversity	Any city

Approach	Condition variable	Method	RL role	Context	Area
		more than 20% of the total city area			
Expert thresholds	Percentage of native bird species in built up areas (relative to total native bird species)	Scoring is based on the relatively low number of native bird species in urban areas. A city gains the maximum score when in built-up areas it is possible to find more than 20% of the total native bird species that are present within the city	X_{100}	Singapore index on cities' biodiversity	Any city
	Proportion of invasive alien species	Scoring range is based on the negative impact of invasive alien species on the native species. No point is assigned to cities where the percentage of invasive alien species in a selected taxonomic group is higher than 30%	X_0	Singapore index on cities' biodiversity	Any city

"Special values" relating to policies are a logical and potentially useful approach to setting reference levels for urban habitats. Several EU policies (i.a., Nature Restoration Regulation, Air Quality Directive etc.) provide useful "benchmark" thresholds (X_{40}, X_{60}) for variables like the share of urban green space / canopy cover and air pollutants concentration (Vallecillo et al., 2022; EU, 2024a,b). Such policy thresholds are sometimes based on standards of human well-being, for example the air quality directive aims to align with the most up-to-date WHO Air Quality Guidelines (Vallecillo et al., 2022). As an internationally relevant example, Target 11 of the Aichi Biodiversity Targets (= 17%) is also used by the Singapore index on cities' biodiversity as the upper reference level (X_{100}) for the proportion of protected natural areas (Chan et al., 2021).

30% canopy cover in urban areas is a well-used "benchmark" threshold (X_{60}) based on scientific evidence. An Australian study suggests that this is the minimum canopy cover percentage that ensures city residents benefit in terms of their health and wellbeing (Astell-Burt & Feng, 2020). Many cities around the world have adopted a "3-30-300 rule of thumb" for city planning (3 mature trees visible from every home, 30 percent tree canopy cover in every neighbourhood, 300 metres from the nearest high-quality public park or other green space) (Konijnendijk, 2021).

A study from Warsaw evaluated the proper size of the Polish eco-spatial index (Ratio of Biologically Vital Areas; i.e., ratio of regions with ecological functions within the city) using empirical evidence. The relationship between the size of the RBVA and selected environmental features (e.g., air temperature and humidity, floristic diversity, butterfly species richness, surface outflow, etc.) was calculated in neighbourhoods characterised by different RBVA values (ca. 20% to ca.70%). They observed a nearly linear correlation between the RBVA and environmental performance and recommended an RBVA value of at least 45 % to guarantee ecological balance and environmental performance (Szulczewska et al., 2014). This value could be potentially useful as a threshold RL (X_{60}).

Urban ecosystems are also often “scored” against thresholds determined by experts. In the context of Singapore's biodiversity index alone, experts have set reference levels for several variables including the proportion of natural areas in the city, percentage of native bird species (relative to total native bird species) and proportion of invasive alien species (Chan et al., 2021).

Although it is difficult to define or identify sites in a "reference condition" for urban areas, a recent study from Madrid presents a framework (based on the SEEA EA) that focuses specifically on the condition of urban green areas and the ecosystem services they provide. Artificial green areas, urban trees and the SIOSE HR[62] forest categories with tree cover greater than 30 % were selected to identify reference sites for urban areas. Further analysis showed that Monte de El Pardo had the highest average values for the condition variables among the green areas and was chosen as the reference site to represent the “best attainable” reference condition, which was then used to derive the upper RL (X_{100}) for a great variety of variables, many of which directly relate to ecosystem services (Perviousness, NDMI, Cooling effect, Quietness index, Air pollution, Water vapour, Communitarian interest in bird species richness, NDVI, Proximity to riparian habitats, Tree canopy cover, Forest Area Density) (Álvarez-Ripado et al., 2024). However, the approach is not yet well-established and using a forest area as a reference site for urban ecosystems raises several conceptual and operational questions.

Condition scale(s)

Table 7.1.5.2: Examples for discrete scales used for condition variables in European urban areas. (N: Number of condition/status classes distinguished)

N	Class names	Condition variable	Context	Area
4	A (very important), B (important), C (less important), D (not assessed)	Recreation area valuation	Norwegian Env. Agency assessment standard	Norway
3	High, medium, low	Green areas (%), connectivity of green areas, % canopy cover of green areas	Study - analysis of green infrastructure in Burgas	Bulgaria

Many cities around the world use scoring systems to measure the “greenness” of urban areas. Typically, such systems are policy-driven and focus on the effects of urban greenery on human well-being. Often, they are also conceptually quite far from the SEEA EA methodology and from more “typical” ecosystem condition scales used in other ecosystem types. For example, a Bulgarian analysis of green infrastructure focuses entirely on its function in the regulation of air quality (pollution mitigation and dust reduction) using a simple discrete condition scale

[62] High Resolution SIOSE is a system built for the integration of high detail geospatial sources in Spain (https://www.siose.es/en/web/guest/siose-alta-resolucion)

with three status classes (high, medium and low) (Borisova et al., 2024). The Norwegian Guidance on Mapping and Valuation of Recreation Areas, on the other hand, features a four-part discrete scale from A to D, where A and B represent valuable areas, C means less important and D category areas have not been assessed. This valuation was performed entirely by local residents, using 13 qualitative criteria provided by a dedicated guidance (Cimburova & Barton, 2021).

7.1.6 Wetlands

Reference condition

The reference state for open wetlands in Norway has been defined as one where the ecosystems remain open with little to no woody encroachment or anthropogenic disturbance. For natural ecosystems this reflects a state where climate warming, wetland drainage, and alien species etc. have not led to any net loss of these nature types and their internal functioning (Kolstad et al., 2023; Venter, 2023 in Kolstad et al., 2023).

Setting reference levels using sample-based methods

Table 7.1.6.1: Examples of approaches using reference sites or historical / paleoenvironmental records for setting reference levels for condition variables in EU wetlands.

Approach	Condition variable	RC type	RL role	Context	Area
Reference sites with little to no woody encroachment	Encroachment	Natural	X_{100}	SEEA EA	Norway
Historical baseline	Wetland birds index	Natural	X_{100}	SEEA EA	Norway

For Norwegian open ecosystems, reference areas with good ecological condition (i.e., no woody encroachment) have been used to set the optima RL (X_{100}) for the variable of woody encroachment. This is also an example of a “natural scale endpoint”, as there can be no less than zero percent encroachment. (Kolstad et al., 2023).

Historical data has been used to set the optimal baseline (X_{100}) for the breeding birds index. The breeding bird indicators represent average trends relative to a reference year and are therefore scaled with respect to population size in the reference year. Therefore, the indicator value in the reference year is 100 %. The reference value 100 represents unchanged abundance relative to the reference year. Determining a threshold for what level of decrease in bird abundance is acceptable under “good ecological condition”, however, is a challenging task. Another approach to determining the threshold would be to assume that bird populations should not be declining at all if an ecosystem is in good condition, i.e., that 100 is the threshold for good ecological condition (X_{60}) instead of the upper reference value (Nater and Pavón-Jordán, 2023 in Kolstad et al., 2023).

Other methods to set reference levels

Table 7.1.6.2: Examples for other methods used for setting reference levels for condition variables in EU wetlands.

Approach	Condition variable	Method	RL role	Context	Area
Natural scale endpoints	Anthropogenic disturbance to soils and vegetation (ADSV)	No disturbance (Sites without wearing, vehicle tracks, erosion)	X_{100}	IBECA (Jakobsson et al., 2020, 2021)	Norway
	Encroachment	50^{th} percentile of LiDAR-derived vegetation heights (i.e. forested areas)	X_0	IBECA	Norway
Expert thresholds	ADSV	The threshold RL is set to no more than 10% damage compared to pristine condition	X_{60}	IBECA	Norway

The variable "anthropogenic disturbance to soils and vegetation" (ADSV) has negative directionality, so the upper RL (X_{100}) can be set using the scale end point where no disturbance (no wearing, vehicle tracks, erosion) has occurred. Following this, the threshold for good condition (X_{60}) for ADSV is set to <10 % damage compared to the optimal state. This is an expert judgement, and a rather weak one at that. In the future, empirical validation of this threshold might feature comparing different variable scores with erosion rates or the potential for plant growth. This would need to be done specifically for each nature type (Kolstad et al., 2023).

The approach to set the lower endpoint (X_0) for the "encroachment" variable was to take the 50th percentile of LiDAR-derived vegetation heights within forest polygons to define a climax vegetation successional stage where encroachment is at its most extreme ("worst possible") (Kolstad et al., 2023).

Condition scale(s)

Table 7.1.6.3: Examples for condition scales used for condition variables in European wetlands. (N: Number of condition/status classes distinguished)

N	Class names	Condition variable	Context	Area
Continuous		Anthropogenic disturbance to soils and vegetation (ADSV), encroachment, etc.	IBECA	Norway

The ecosystem condition indicators developed in Norway for three of their major ecosystems (wetlands, open areas below the forest line and semi-natural land) are all rescaled from 0 to 1 according to the IBECA methodology (Jakobsson et al., 2020, 2021), where 0 refers to the maximally degraded condition, 1 to the reference condition and the threshold between good and not good condition corresponds to 0.6 (Kolstad et al., 2023).

7.1.7 Rivers and lakes

Reference condition

For most bodies of water, the Water Framework Directive prescribes the use of a natural reference condition (representing “high ecological status”). The RC is usually determined from reference sites in (near) pristine condition or sites in “least disturbed condition”, where modelling “pressure removal” can be used to estimate the environmental state without the pressure. If no reference sites are available, historical / paleoenvironmental records are used. When all else fails, the RC can even be determined by experts (EC, 2003b).

For artificial or heavily modified water bodies, an anthropogenic reference condition can be used instead, defined as their “maximum ecological potential”. The biological status of such water bodies reflects that of the closest comparable water body as far as possible, taking the modified characteristics of the water body into account (EC, 2003c).

As an example outside the WFD, in the Ogosta Valley of Bulgaria, the reference condition for surface water and groundwater pollution was estimated based on reference sites in the Ogosta River floodplain (Kotsev et al., 2019). This approach remains very close to WFD conceptually.

Setting reference levels using sample-based methods

Table 7.1.6.1: Examples of approaches using reference sites or paleoenvironmental records to set reference levels for condition variables in EU freshwater ecosystems.

Approach	Condition variable	RC type	RL role	Context	Area
Reference sites in (near) natural or “least disturbed” condition, paleoenvir-onmental records	WFD quality elements (biological + supporting physico-chemical and hydromorphological elements)	Natural	X_{80} / X_{100}	WFD	EU lakes, rivers, transitional waters and coastal waters
Reference sites in the Ogosta River floodplain	Dissolved Oxygen (DO), pH, arsenic	Natural	X_{100}	INES project	Ogosta, Bulgaria

WFD indicators include the primary biological quality elements (phytoplankton, macrophytes, phytobenthos, benthic invertebrate fauna and fish) and supporting physico-chemical (nutrients, oxygen condition, temperature, transparency, salinity and river basin specific pollutants) and hydromorphological quality elements. The condition of these elements is estimated on a five-part condition scale with five status classes: bad, poor, moderate, good, high (EC, 2003b). Reference sites in (near) pristine condition are the recommended approach to quantify the upper RL (X_{100}) or the boundary between high/good status (X_{80}). The H/G boundary is often also derived as the 10 % percentile of the upper RL (EC, 2003b). If reference sites in natural condition are not available, reference levels can sometimes be extrapolated from reference sites in "least disturbed condition" based on modelling ('pressure removal') or quantifying an alternative benchmark (e.g., X_{60}) (see section 6.2). If available, WFD also recommends the use of available paleo-environmental or historical data (sediment cores, past monitoring records, etc.) (EC, 2003b).

The INES project (INtegrated assessment and mapping of water-related Ecosystem Services for nature-based solutions in river basin management) used a similar approach, defining the upper reference levels (X_{100}) for dissolved oxygen (DO), pH and arsenic from reference sites located in the floodplain of Ogosta River (Kotsev et al., 2019).

Other methods to set reference levels

Table 7.1.6.2: Examples for other methods used for setting reference levels for condition variables in EU lakes, rivers, transitional waters and coastal waters.

Approach	Condition variable	Method	RL role	Context	Area
Natural scale endpoints	Biological quality elements	For biological quality elements, zero population level is the worst possible state	X_0	WFD	EU
Policy thresholds	Priority substances and certain other pollutants	WFD, GWD and EQSD	X_{60}	WFD	EU
Interpolation & extrapolation	All quality elements	The remaining class boundaries are typically derived from the good-moderate boundary (X_{60}) and reference condition (X_{100})	X_{20}, X_{40}, X_{80}	WFD	EU
Expert thresholds	All quality elements	A "stopgap" approach used when neither reference sites nor historical / paleoenvironmental data is available	X_{60}, X_{100}	WFD	EU

In the WFD context, there are several different ways to set the threshold RL (X_{60}) and remaining class boundaries. For biological quality elements, the natural scale end point representing zero population level is often used for the lower endpoint (X_0), as this is the worst possible state (EC, 2011). For water quality metrics representing priority substances or pollutants, policy thresholds set under several different EU directives (WFD, GWD, EQSD) are the logical choice for the RL between moderate and good condition (X_{60}) (EC, 2011). The remaining class boundaries (X_{20}, X_{40}, X_{80}) for all variables are typically derived from the good-moderate boundary (X_{60}) and reference condition (X_{100}), e.g., X_{80} is usually defined as the 10% percentile of X_{100} (EC, 2011). Expert thresholds are also commonly used as the "last resort" for deriving the upper RLs (X_{80}, X_{100}) or the threshold of good condition (X_{60}) when other means are not available.

Condition scale(s)

Table 7.1.6.3: Examples for condition scales used for condition variables in EU freshwater ecosystems. (N: Number of condition/status classes distinguished)

N	Class names	Condition variable	Context	Area
5	High, good, moderate, poor, bad	WFD quality elements (total P / N, Dissolved Oxygen, pH, Chlorophyll, Water Quality Index, Taxonomic Diversity, *etc.*)	WFD	EU lakes, rivers, transitional waters and coastal waters
3-5	Nitrate thresholds	Nitrates	Nitrate directive	EU lakes, rivers, transitional waters and coastal waters
4	Excellent/Good Quality, Slightly Polluted,Moderatel y Polluted, Heavily Polluted	Arsenic	INES project, Global Water Partnership (GWP)	Ogosta river, Bulgaria

The overwhelming majority of freshwater ecosystem condition indicators use the standard five WFD status classes ranging from bad to high (EC, 2003b). The nitrate indicator scale is an exception, using 3-5 thresholds regarding the degree of pollution (EC, 2024). The arsenic pollution of the Ogosta river in Bulgaria has also been measured on a four-step scale from heavily polluted to excellent / good quality in the context of the INES project (Global Water Partnership) (Kotsev et al., 2019).

7.1.8 Marine ecosystems

Reference condition

The reference condition in marine contexts typically represents the "natural" or "ideal" state of the environment. This state has been defined using all approaches detailed in Chapter 6.1: reference sites, modelling approaches (pressure removal) and historical (or paleo-environmental) data. For certain pressures and synthetic pollutants, 'natural zero' values are considered to reflect the natural state. If the pristine condition cannot be determined, other potential approaches to establish a reference condition include the use of 'past state' (least impacted known environmental state) from available (sufficiently long) time series, 'current state' (set at the beginning of an environmental policy or first assessment of state) and ‘potential future state’ (Arcangeli et al., 2022). The use of the current state is sometimes considered an interim definition, until a more robust definition can be established.

In coastal waters, WFD typically also considers the reference condition as a pristine or natural state of the environment, and for natural waterbodies this is defined as “high ecological status”. However, for heavily modified or artificial water bodies the reference condition is instead derived from their “maximum ecological potential”. This is a state where the biological status reflects that of the closest comparable water body as far as possible, taking the modified characteristics of the water body into account (EC, 2003c). In the marine context, heavily modified or AWBs include port areas or dammed coastal areas (like reservoirs).

Setting reference levels using sample-based methods

Table 7.1.8.1: Examples of approaches using reference sites, historical data or paleoenvironmental records for setting reference levels for condition variables in EU regional seas.

Approach	Condition variable	RC type	RL role	Context	Area
RL was derived from Swedish bays in a natural state	Coastal fish key species	Natural	X_{100}	MSFD, HELCOM	Baltic Sea
RL was derived from a population in reference condition	Blubber thickness	Natural	X_{100}	MSFD, HELCOM	Baltic Sea
RL was derived from historical Secchi records	Water transparency	Natural	X_{100}	EU WFD (Finland)	EU coastal waters
Old non-scientific literature was used to determine old distribution (RL not quantitative)	Abundance of harbour porpoise	Natural	X_{100}	MSFD, HD	Baltic Sea

In the MSFD context, the reference levels that separate good from not good condition are typically set at a level of “safe” or “acceptable” deviation from the reference condition, or represent the maximum acceptable extent of loss or damage. However, it is debatable where such reference conditions can be found. For instance, there are not many applicable reference sites in the Baltic Sea because nearly all parts are under some level of pressure from human activity and / or eutrophication.

However, there are still two examples of the use of reference sites for determining upper reference levels for Baltic Sea indicators. The baseline (X_{100}) for seal blubber thickness was originally derived from as far as seal populations found in the Atlantic Ocean (HELCOM, 2023, Nutritional status of marine mammals). For the variable of coastal fish key species, the upper RL could be derived from certain bays in northern Sweden that were found to be in a more natural condition compared to the rest. Additionally, the period of observation (reference period) had to represent stable environmental conditions and cover at least ten years to extend over two generations of fish (HELCOM, 2023, Abundance of coastal fish key species).

There are also examples of using historical / paleo-environmental records. The upper reference level for water transparency was derived from historical records of Secchi depth over 120 years ago, when offshore waters were relatively unimpacted and coastal impacts were mainly found adjacent to cities and estuaries (Fleming-Lehtinen, 2016). Anecdotal evidence from old non-scientific literature, on the other hand, was used to determine the historical distribution of harbour porpoises in the Baltic Sea (notably, due to the anecdotal nature of the records, this is not a quantitative upper RL) (HELCOM, 2023, Harbour porpoise abundance).

Methods for setting all reference levels

Table 7.1.8.2: Examples for methods used for setting all reference levels for condition variables in EU regional seas.

Approach	Condition variable	Method	RL role	Context	Area
Natural scale endpoints	Gray seal population size	Zero population level is the worst possible state	X_0	MSFD, HD	Baltic Sea, NE Atlantic
Policy thresholds	Abundance of breeding waterbirds	Max 30% decrease of population size is allowed in relation to baseline year	X_{60}	MSFD, HELCOM, OSPAR	Baltic Sea, NE Atlantic
Scientific thresholds	Fishing mortality	Model-derived RL, producing optimal fishery catch without endangering the stock	X_{60}	MSFD, CFP, ICES	EU
Interpolation & extrapolation	Chlorophyll-a	The remaining class boundaries are derived from the good-moderate boundary (X_{60}) and reference condition (X_{100})	X_{20}, X_{40}, X_{80}	WFD	EU
RL transfer	Macrophytes depth distribution	Reference level for water transparency was transferred by linear correlation to this variable	X_{60}	EU WFD (Finland)	EU coastal waters
Expert thresholds	Cumulative impact from physical pressures on benthic biotopes	Experts define the impact level of each human activity	X_0-X_{80}	MSFD, HELCOM	Baltic Sea

In the absence of reference sites, the “special values” methods are well represented amongst MSFD indicators. For biodiversity indicators, such as grey seal population size, the natural scale end point of zero population is often used to represent the lower reference level i.e., “worst possible” state (X_0) (HELCOM, 2023, Gray seal population size). Similar examples can be found from all EU marine regions. Policy thresholds are also very commonly used and usually represent the threshold between good and not good condition (X_{60}). For example, in the case of breeding waterbirds abundance the reference level (of % change from a baseline) was agreed in HELCOM and OSPAR (e.g., HELCOM, 2023, Abundance of breeding waterbirds).

The threshold between good and not good (X_{60}) for fishing mortality / maximum sustainable yield combines policy (Common Fisheries Policy), scientific and “pressure removal” methods, using a modelling approach and scientific advice from ICES. The model-derived threshold represents optimal fishery catch without endangering the stock; when the spawning stock biomass drops below the maximum sustainable yield (MSY) biomass trigger reference points or the biomass threshold, fishing pressure should be reduced (usually below FMSY) to allow the stock to rebuild (ICES, 2025).

Several WFD indicators are also used for marine condition assessments under MSFD. Where WFD indicators use five ecological status classes ("bad", "poor", "moderate", "good", high"), in the MSFD context, only the upper RL (~reference condition) and the threshold value separating good and not good condition are required (Arcangeli et al., 2022). When WFD indicators are used in the marine context, the reference level separating moderate and good condition (X_{60}) represents the MSFD threshold between good / not good condition (X_{60}).

“Reference level transfer” is a very common approach to set reference levels for WFD indicators and therefore is well represented also amongst marine coastal indicators. For example, the threshold (X_{60}) for water transparency relates directly and could be transferred by linear correlation to represent the threshold (X_{60}) of macrophyte depth distribution (Fleming-Lehtinen & Laamanen, 2012).

Interpolation and extrapolation methods are typically used to derive the remaining WFD indicator class boundaries from the good-moderate boundary – this is usually not relevant to MSFD indicators, but the WFD/MSFD overlap includes the marine coastal ecosystems and relevant variables, such as Chlorophyll-a (HELCOM, 2023, Chlorophyll-a).

As there is yet little research on the effects of cumulative impacts on benthic biotopes, all reference levels for the indicator “Cumulative impact from physical pressures on benthic biotopes” were set by experts, where the threshold between low and moderate categories was defined as the level of cumulative pressure that would likely prevent achieving GES (X_{60}) (HELCOM, 2023, Cumulative impact […] on benthic biotopes).

Condition scale(s)

Table 7.1.8.3: Examples of the use of discrete scales for condition variables in the Baltic Sea. (N: Number of condition/status classes distinguished)

N	Class names	Condition variable	Context	Area
2	Good, not good	Grey seal population size + blubber thickness, abundance of wintering and breeding waterbirds, white-tailed eagle productivity, herring / cod / sprat biomass and fishing	MSFD, HELCOM	Baltic Sea

N	Class names	Condition variable	Context	Area
		mortality, abundance of coastal fish key groups and species, benthic quality index, *etc*.		
6	High, moderate 1-3, low, very low	Cumulative impact from physical pressures on benthic biotopes	MSFD, HELCOM	Baltic Sea
5	High, good, moderate, poor, bad	Macroalgae depth distribution, Chlorophyll-a, total phosphorus and nitrogen, water transparency, DIN, DIP, *etc*	WFD	Baltic Sea coastal waters

Nearly all MSFD indicators use a two-part condition scale (good / not good), where Good Environmental Status is defined as "the environmental status of marine waters where these provide ecologically diverse and dynamic oceans and seas which are clean, healthy and productive". The index for cumulative benthic impact is a notable exception, separating the spatial extent of disturbance into six different impact levels (from *very low* to *high*). In this case the threshold between "good" and "not good" is set between the low and moderate categories (EU, 2017 & 2022, HELCOM, 2025).

The WFD indicators used in marine coastal waters all feature the five standard WFD status classes from bad to high. To facilitate integration with MSFD-specific indicators, the threshold between good and not good condition is set between the moderate and good categories (= 0.6 on the continuous scale).

7.2 EU Water Framework Directive

Here we will focus on the main steps of rescaling (identification of RC, linking RLs to the RC, identification of further RLs, and the construction of the rescaling function, "translating" each step to the common "rescaling language" established in Chapters 5-6.[63] By making this link we will also assess the possible misalignments between the studied framework and our system, highlighting also potential issues with SEEA EA compatibility. Finally, we aim to illustrate each framework with 1-2 "typical" rescaling functions, which will be documented in detail. These examples will illustrate how the given framework handles RC and RLs at a very practical level, which can be highly useful / inspiring for SELINA case studies and beyond.

The main goal of the EU Water Framework Directive (2000/60/EC) is to establish a framework for the protection of surface waters (and groundwater, which is omitted here). The units of analysis are *surface water bodies* (*~ecosystem assets*), which are delineated separately as four *water categories* (*~ETs*: lakes, rivers, transitional waters, and coastal waters). The *ecological status* of water bodies is to be reported on a graded ordinal scale with five *status classes* ("bad", "poor", "moderate", "good", "high", see also Fig. 6.2.), which can be used to formulate statements/goals on the ecological status of the water bodies in a simple language. The main objectives of the WFD, to be achieved by year 2027 (originally 2015, later postponed

[63] To improve readability, such "translations" will be in *italics* and start with a tilde (~) sign.

several times), are (i) to prevent deterioration of the status and (ii) to reach at least *good ecological status* for all surface water bodies. Less ambitious objectives (termed "*good ecological potential*") are set for *artificial* and *heavily modified* water bodies (EC, 2023c).

The WFD distinguishes several groups of ecosystem condition characteristics ("quality elements"), with slightly different standardisation and aggregation procedures. Here we review the procedure followed for "biological quality elements" (BQE, largely covering ~*ET classes B1-B3*). While the WFD prescribes the list of BQEs (~*characteristics*) to be monitored (WFD: Annex V), the implementation is left to the national level, i.e., each country is expected to develop reliable metrics (~*variables*) that can make these BQEs measurable. For each meaningful combination of BQE and water category each country is expected to develop at least one metric. Metrics are also expected to be linked to a specific pressure, and if there are multiple pressures affecting a BQE in a water type, then these pressures are recommended to be addressed by separate metrics.[64] Finally, each country is also expected to identify "water types" following a harmonised approach (EC, 2003a) in distinguishing ~*subtypes* of lakes, streams, rivers, etc.[65] The most important purpose of defining water types is to minimise the within-type natural variability of the metric values for water bodies *in reference condition* (EUWI, 2020; in line with Section 6.2 above). Accordingly, all RLs are specific to (groups of) water types, i.e., most of the procedures described below have to be repeated for several water types.

For water bodies that are neither artificial nor heavily modified, WFD prescribes the use of a (close to) **natural reference condition** (~*natural RC*) representing a "high ecological status". WFD provides several different options to operationalise this RC and quantify the corresponding ~*RLs*. The first recommendation is to establish a set of *reference sites* (per *water type*) in near pristine condition (with "very minor disturbance" allowed), which allows RLs to be quantified as ~*simple statistical parameters*. The second choice is to select a set of reference sites in "least disturbed condition" and use modelling (~*pressure removal*), or quantify an "alternative benchmark" (see below) and ~extrapolate the RC with it. If available, WFD also recommends the use of ~*paleo-environmental* or historical data, and if everything else fails ~*expert thresholds* can also be used for setting the RLs (EC, 2003b).
If suitable near pristine *reference sites* could be established, then the data from these sites are used to derive two RLs:

- the "reference value" (~X_{100}) is typically quantified as the *median* or the *mean* of the sample values, and

[64] If there are several metrics proposed for a specific combination of *water category*, *BQE*, and *pressure*, and none of them is clearly superior than the others, then it is a common practice to combine them into a single *multimetric index* (Buffagni et al., 2007). The way how these multimetric indices are constructed (linear rescaling followed by aggregation with a (weighted) arithmetic means) makes them fully compatible with our general rescaling framework. Accordingly, (as the composition of two piecewise linear transformations is also a piecewise linear transformation) the contribution of the individual sub-indices will be identifiable (disaggregatable) in the final EQR / nEQR values, and it will also be possible to "project" the class boundaries of the multimetric index back to the original scales of the component metrics. Therefore the conclusions that we make here about WFD metrics will also be valid for the individual components of such multimetric indices. Nevertheless, for the sake of simplicity we do not discuss such multimetric indices here any further.

[65] This leads to ~100-150 different subtypes in a medium-sized EU country.

- the "boundary of high/good status" (often shortened as "H/G"; ~X_{80}) is typically derived as the *10% percentile* of them.

If the near pristine sites are missing, but data for the relevant pressure/land use gradients are available and "least disturbed sites" can be identified, then the following two methods are recommended (EC, 2015):

- "alternative benchmark": instead of trying to set the "high status" (~X_{100}) or the "H/G boundary" (~X_{80}), another solution is to choose a "lower" status class, and identify reference sites ("alternative benchmark sites") which can be associated with that class boundary (e.g., the G/M boundary, ~X_{60}). The quantification of this RL happens with ~*simple statistical parameters* (EUWI, 2020), and the "higher" RLs can then be determined using ~extrapolation.
- The other option is to use modelling (~*pressure removal*, hindcasting) to "predict" an estimated value (and a confidence interval) for the natural (non-disturbed) state.

After the "upper" RLs (or the "alternative benchmark") has been set, the remaining status class boundaries (G/M: ~X_{60}, M/P: ~X_{40}, and P/B: ~X_{20}) can also be addressed. There are also several possibilities here:

- The pessimal RL (X_0) is typically set to zero for BQE metrics.[66]
- If there are well-established ~*special values* available (policy/scientific thresholds) that can be clearly matched to a class boundary, then it should also be used.[67]
- Class boundary RLs can also be "borrowed" from other closely correlated pressure or condition variables using ~*RL transfer* which have a clearly recognised special value or a well-established RL.
- If some of the class boundaries cannot be set in a meaningful way. then they should be ~*interpolated* instead (described as "equal width classes" in the guidance documents).

In addition, it is required that the class boundaries defined in a specific country should be regularly "*intercalibrated*" with those of similar variables in similar water types in the neighbouring countries (so called *Geographical Intercalibration Groups*). Technically an intercalibration exercise consists of a large number of ~RL transfers following structured choreography on a big joint database. The outcomes of such an exercise are then used to modify (harmonise) the RLs in all of the participant countries (EC, 2015, Fig. 7.1).

Based on the RLs, WFD provides a piecewise linear function to normalise (~rescale) the values of the metrics (~variables) to a 0-1 scale. This rescaling is usually described to happen in two consecutive steps (EC, 2011):

- first each metric is transformed into an *ecological quality ratio* (EQR) value using a simple linear transformation with X_{100} and X_0 (similar to Eq. (1)),

---

[66] With some exceptions (including negative directionality variables).

[67] There is a series of complex instructions in EC (2011), which aim at identifying "discontinuities" (breakpoints) in the relationship between the metric and the associated pressure, which can also be matched to class boundaries as special ~*expert thresholds*.

- then the EQR is further transformed into a *normalized EQR* (nEQR) using a stepwise linear transformation using 6 RLs (X_0=0, X_{100}=1, and the EQR values corresponding to the four class boundaries)

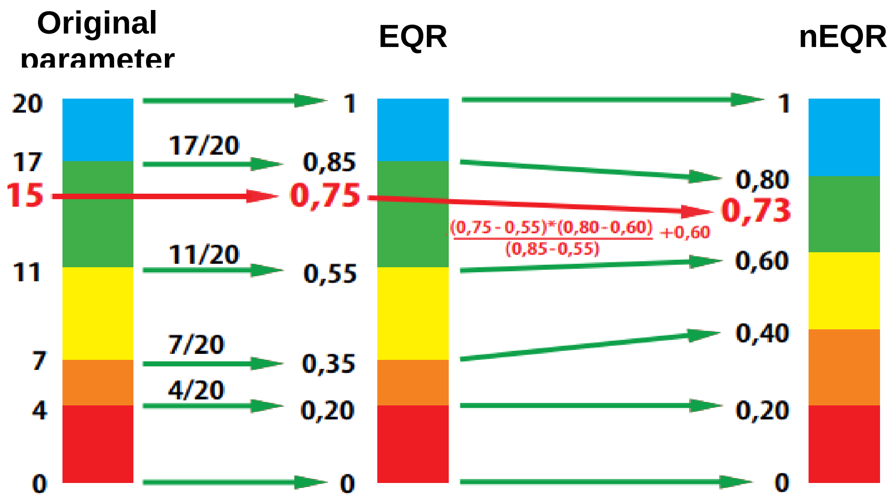


Figure 7.1. A visual example of the two steps of how the *normalized ecological quality ratio* (nEQR, *~indicator*) is calculated in two steps from an unscaled WFD *parameter/index* (*~variable*) (source: DV, 2018: p38).

While EQR is slightly simpler, nEQR offers more straightforward and intuitive policy interpretation (e.g., an nEQR value of 0.21 always indicates a "poor" condition, an nEQR of 0.75 is always "good", etc; while an EQR value of 0.75 can possibly be in any of the (higher) status classes). According to the approach in this report, we consider nEQR to be the condition indicator, and EQR is just an interim step for calculating this indicator. As the composition of the two consecutive piecewise linear transformations is also a piecewise linear transformation, an nEQR value can also be calculated "directly" from the original metric in a single step. This "direct" rescaling function can be described using the following equation:

$$Y = \begin{cases} \text{...if } X > X_{100} & 1 \\ \text{...if } X_{80} < X < X_{100} & (X-X_{80}) / (X_{100}-X_{80}) + 0.8 \\ \text{...if } X_{60} < X < X_{80} & (X-X_{60}) / (X_{80}-X_{60}) + 0.6 \\ \text{...if } X_{40} < X < X_{60} & (X-X_{40}) / (X_{60}-X_{40}) + 0.4 \\ \text{...if } X_{20} < X < X_{40} & (X-X_{20}) / (X_{40}-X_{20}) + 0.2 \\ \text{...if } X_{0} < X < X_{20} & (X-X_{0}) / (X_{20}-X_{0}) \\ \text{...if } X < X_{0} & 0 \end{cases} \qquad \text{(Eq. 2)}$$

Figure 7.2 shows two examples for such nEQR rescaling functions.

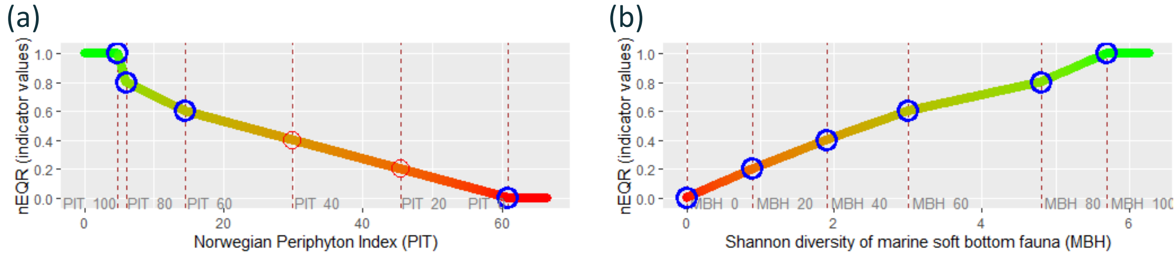


Figure 7.2. RLs and rescaling functions applied to calculate dimensionless indicator values (nEQR) for two WFD variables in Norway. (a) Norwegian periphyton index for trophic status of rivers (PIT; Schneider & Lindstrøm, 2011) in very lime-poor rivers. (b) Shannon diversity of marine soft bottom fauna (MBH; Rygg, 2006). (Source: Schartau et al., 2016)

7.3 EU Marine Strategy Framework Directive

7.3.1 Descriptors and criteria of assessment

The Marine Strategy Framework Directive (MSFD) was adopted by EU in 2008 with the main goal of achieving Good Environmental Status in EU marine waters by 2020, where GES is defined as "The environmental status of marine waters where these provide ecologically diverse and dynamic oceans and seas which are clean, healthy and productive". MSFD further identifies eleven qualitative condition descriptors (loosely corresponding to ~ecosystem characteristics), including both state elements (i.a., biodiversity, food webs and sea floor integrity) and anthropogenic pressures and impacts (i.a., eutrophication, contaminants and pollutants, litter and input of energy) (EU, 2008).

Table 7.3. Qualitative descriptors of good environmental status of marine environment sensu MSFD.

D1: Biodiversity is maintained
D2: Non-indigenous species do not adversely alter ecosystems
D3: Populations of commercial fish and shellfish species are healthy
D4: Food webs ensure long-term abundance and reproduction of species
D5: Eutrophication is reduced
D6: Sea floor integrity ensures the proper functioning of ecosystems
D7: Permanent alteration of hydrographical conditions does not adversely affect ecosystems
D8: Concentrations of contaminants give no pollution effects
D9: Contaminants in seafood are at safe levels
D10: Marine litter does not cause harm
D11: Introduction of energy (inc. underwater noise) does not adversely affect the ecosystem

Each descriptor assessment builds from element assessments which are biological elements (species, habitats, trophic guilds) or pressure elements (biological, physical, substances, litter, energy) using one or more primary criteria and optional secondary criteria and associated indicators. These criteria (for each element) are assessed by one or several parameters

(~variables). Primary criteria are defined at the EU-level, and they are mandatory for all Member States to ensure consistency across assessments, while secondary criteria are used by the Member States where they are relevant to complement the primary criteria. Secondary criteria may e.g., represent the major environmental problem of the region (such as eutrophication in the Baltic Sea), add precision to the assessment of a criterion failing to reach the threshold value, validate a primary criterion that achieves or almost achieves the threshold and fill gaps where the primary criterion is still under development (Arcangeli et al., 2022). The criteria are listed in Commission Decision (EU) 2017/848 (so-called GES Decision).

7.3.2 Regional Sea Conventions and assessment areas

MSFD acknowledges four regional seas within the Union and mandates that Member States collaborate and coordinate to develop and implement their marine strategies. However, the Directive does not provide specific legal frameworks or governing structures to facilitate this cooperation. This role is filled by four European Regional Sea Conventions: the Oslo-Paris Convention (OSPAR) for the North-East Atlantic, the Helsinki Convention (HELCOM) for the Baltic Sea, the Barcelona Convention (UNEP-MAP) for the Mediterranean Sea and the Bucharest Convention (BSC) for the Black Sea. Within these four regions, GES is determined at various spatial scales so as to properly reflect the ecology of species and habitats, and the requirements of management and administration in relation to pressure elements. For example, whales are a wide-ranging species and are addressed more appropriately at a regional scale, while nutrient enrichment is linked to land-based sources and local management needs, and is therefore better addressed at a smaller localised scale. The GES Decision and the regional sea conventions have set up a nested system of assessment areas of varied spatial resolutions, applicable for each criteria, element, and feature. These scales include regions and their subregions and subdivisions, as well as the 'national parts of a subdivision' (set within national marine areas) and coastal waters as defined under WFD. (SWD, 2020)

7.3.3 Reference condition and threshold values

The Commission Decision (EU) 2017/848 acknowledges only a single reference level, i.e., the 'threshold value' between good and not good condition. This is set to the levels of pressure or impact the marine environment can withstand without suffering significant or irreversible damage to its life or its habitats. This interpretation emphasises sustainability of use over a pristine ecological condition. Threshold values in MSFD context can thus be interpreted as environmental objectives in relation to a desirable condition (EU, 2017). They are typically set at a level of "safe" or "acceptable" deviation from a reference condition that represents the natural or ideal state of the environment, or represent the maximum acceptable extent of loss or damage (Magliozzi et al., 2021).

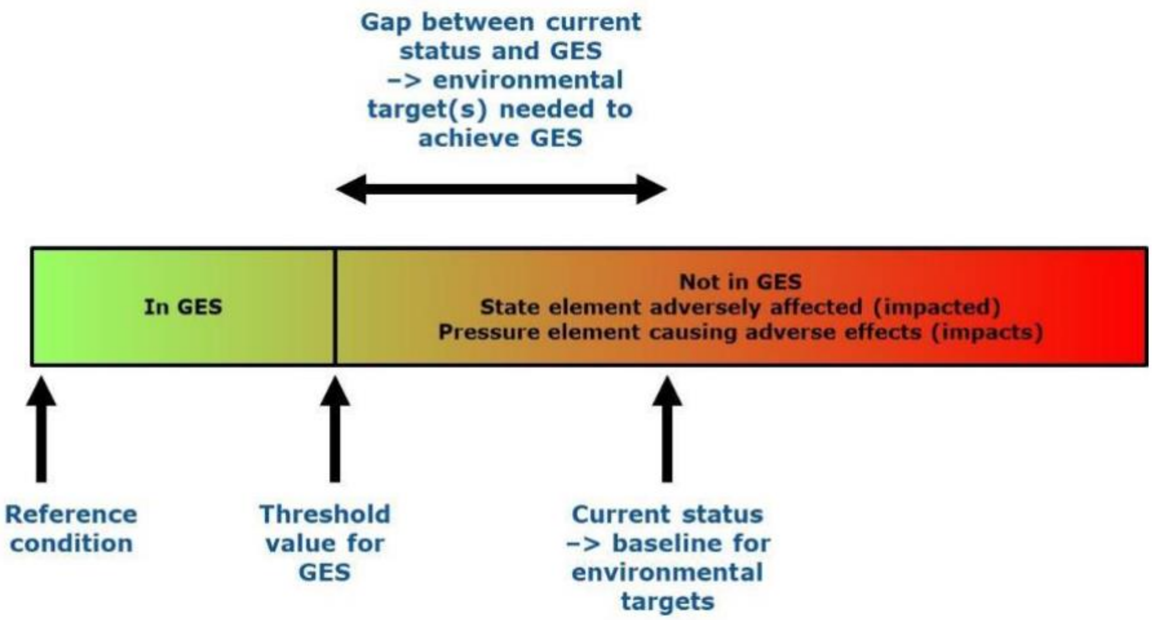


Figure 7.3. An example for the relationship between the RC (~X_{100}), the threshold value (~X_{60}), and the current value in an MSFD variable that is currently not in good environmental status (GES) (source: https://commission.europa.eu/system/files/2020-06/swd202062final.pdf: p52)

MSFD does not specify how a threshold is set, but these are set for the parameters under the element criteria. The GES Decision lists certain thresholds which are to be defined on EU level and some on regional level. Therefore, the regional sea conventions and technical groups of member state experts under the MSFD common implementation strategy have taken the role of defining some of the thresholds. The rest are legally the Member States responsibility, but in practice the regional sea conventions have supported the work and hence ensured better coordination. The MSFD does not legally require rescaling of the parameters (~variables) but this is commonly done in EEA and Regional Sea Convention integrated assessments for the MSFD variables (see EEA, 2029; HELCOM, 2023).

The common approaches to derive the baseline in the MSFD context are, roughly in order of scientific robustness, 'reference condition', 'past state', 'current state', 'potential future state' and 'policy threshold'. 'Reference condition' represents a state with little to no impact to the environment and the upper RL (X_{100}) can be derived from existing 'pristine' areas (~reference sites), historical data, modelling (or a combination thereof) or even from a 'natural zero' value when it comes to pollutants not typically found in nature.[68] 'Past state' refers to the least impacted known environmental state from an available data time series. To account for interannual variability, the upper RL is often calculated as the mean of several years. This approach is considered relatively robust but there is a risk that the RL is based on an already significantly impacted condition.[69] 'Current state' can be set as the upper RL at the beginning of a particular environmental policy or first assessment of state.[70] 'Potential future

[68] The reference condition approach is used for e.g., Descriptor 5 (Eutrophication) criteria C1 (nutrient concentrations) and C2 (chlorophyll *a* concentrations).

[69] The past state approach has been used for e.g., the HELCOM waterbirds wintering abundance indicator.

[70] The current state approach has been used in e.g., OSPAR assessments of Atlantic grey seal abundance.

state' requires modelling of a future condition or a trend-based approach of a continuous improvement in the environmental state.[71] The policy threshold approach is used for e.g., the sea-floor integrity indicator, where specific percentage values of unimpacted habitat area were set on the EU level. Whichever approach is taken, baselines are usually reviewed periodically to consider the effects of e.g., changes in human activities in the area or climate change (Arcangeli et al., 2022).

Each spatial assessment unit typically requires its own threshold values for each indicator to account for differences in the biotic and abiotic characteristics between areas, or the availability of data (EU, 2017, 2022).

The thresholds between good and not good condition (X_{60}) are set at different stages of the EU hierarchy. The key principles for thresholds are described at Union-level (Article 9 of MSFD). This also includes thresholds "borrowed from" other policies, e.g., threshold values (called "reference points") taken from the Common Fisheries Policy, thresholds of the Habitats Directive or status class boundaries for water quality elements as described in WFD.

Although the threshold values for the scientific indicators are often considered on a case-by-case basis, MSFD guidelines (Arcangeli et al., 2022) identify certain guiding principles and narratives that are taken into account, including the precautionary principle, risk considerations and the legal requirement of non-deterioration (Article 1 MSFD).

Common narratives and approaches when setting threshold values are as following:

- 'Acceptable deviation from reference condition/baseline': an approach that accounts for natural variability in certain parameters such as zooplankton populations and nutrient loads.
- 'Cut-off values' are implemented in cases where a certain concentration of a substance is known to cause adverse effects (harmful substances, low oxygen levels).
- 'Removal and conservation targets' are used to measure population viability for e.g., HELCOM indicators concerning drowned mammals and waterbirds in fishing gear.
- Maximum sustainable yield indicators in relation to commercially exploited fish populations follow the principle of 'lowest endpoint', indicating a low risk for over-exploitation.
- The 'limit reference level', used in e.g., the HELCOM seal abundance indicators, is derived from population viability analyses where a low risk of extinction is accepted.
- The 'vulnerability approach' involves the use of specific pressures (e.g., noise, bycatch, entanglement by litter) as a proxy ('reference level transfer') to measure the state of species or habitats that are especially vulnerable to that pressure.

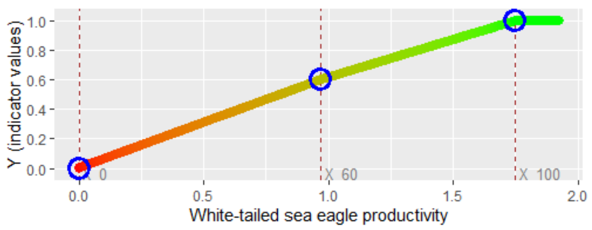
[71] The potential future state approach has been used when modelling future marine bird population size.

7.3.4 Expression of GES

Unlike the five-step condition scale of the Water Framework Directive, in MSFD assessments GES status is expressed as good status achieved, not achieved, not assessed or unknown. Regional Sea Conventions may report the status of marine condition with more nuance in their own reporting, e.g., in HELCOM holistic reports where the integrated assessment results can be expressed in five categories with three degrees of poor status and two degrees of good status (HELCOM, 2023).

The final GES expression for the entire environment ‘builds up’ from threshold values set for single parameters (~variables), which are then combined to increasingly higher levels of integration. During each step of the assessment hierarchy, the parameter, criterion, element or feature either meets or fails the threshold value. The final expression of GES builds up on these ‘quality standards’ through different integration rules (Arcangeli et al., 2022). Integration as high as the descriptor level is not, however, required. Such high-level integration is still used by HELCOM and EEA in their assessment reports with the aim to communicate the state of marine environment in an understandable way (EEA, 2019, HELCOM, 2023).

Variable name:
White-tailed sea eagle productivity
([https://indicators.helcom.fi/indicator/ white-tailed-sea-eagle/](https://indicators.helcom.fi/indicator/white-tailed-sea-eagle/))
Variable unit:
Number of nestlings (dimensionless)
ET code/name: Marine
ET subtype / ecoregion (if relevant): --
RC type: Natural RC



	X value	RL method*	Source (and description/justification)
X_0	0	Natural zero	A zero value that represents local extinction of the population is meaningful as a “maximally degraded value”
X_{60}	0.97	Policy threshold based on historical value	The reference level for productivity 0.97 has been derived by combining the reference levels for brood size and breeding success. Breeding success data cover the period 1915-1953 and nestling brood size includes data from the period 1858-1950. The threshold value of breeding success (successful nests) is 59% (the lower end of the 95% CI 59-86%). The threshold value of brood size is 1.64 (the lower end of the 95% CI 1.64-2.04).
X_{100}	1.75	Historical value	The upper end of the 95% CI.

Figure 7.4. An example for RLs used for a HELCOM variable “While-tailed eagle productivity”, a variable used for the evaluation of contaminant impacts for a top predator under the MSFD descriptor 8. * See Section 6.3 for the description of RL methods.

8 Recommendations

This chapter concludes the report with recommendations about the application of the SELINA reference framework (Chapters 5–6). We grouped our recommendations according to the four steps presented in Chapter 5.1, preceded by an additional preliminary step:

- **Step 0**: Select a specific "condition model", consisting of
    - a set of "status classes" (e.g., "good" vs "not good" (2 classes); or "good" vs. "moderate" vs. "poor" (3 classes)
    - a set of status class boundaries (one fewer than status classes) and specify the indicator values to which they are linked (e.g., the "good/not-good boundary" is at 0.6).
    - normative definitions for the status class boundaries (What kind of criterion should separate/ distinguish the "good" cases from the "not good" cases?)

  One can, of course, freely choose their condition model, but within a single assessment (or account) one should always use the same condition model (for all ETs). It is often possible (and advisable) to "reuse" a condition model from the major EU policies (e.g., the MSFD / IBECA model with two classes and a single boundary at 0.6). The condition model (i.e., the status classes) should be intuitive for the main end users of the assessment / accounts. The number of RLs (per EC variable) will then be determined by the condition model selected (one will always need one more RLs than the number of status classes).

- **Step 1** (for each ET): Select / define the RC for the ET
    - for anthropogenic ETs like urban or croplands proceed to Step3
    - for natural ETs use their stable natural state as RC, for seminatural ETs use a seminatural RC (see Section 6.1), and draft a "normative definition" describing the RC (i.e.: what kind of changes / variability should be "accepted" as part of the RC, and what kind of changes should already be seen as degradation).

- **Step 2** (for each variable): set a RL that corresponds to the RC.

  There are many ways to do this. We recommend the following procedure:
    - identify the relevant ET subtypes (for which different RLs should be used)
        - check if (a representative sample of) paleo-environmental data are available for some of the subtypes. (If yes, use them to estimate the RL. → procedure A)
    - try to identify suitable reference sites (following the normative definition from Step 1)
        - check if you have enough sites with enough data that are still (nearly) in the RC. (If yes, sample these sites. → procedure A)
        - select a sample that includes the best available "least disturbed condition" and all other more degraded ones along a degradation

gradient. Also collect data about the main pressures/drivers for these sites. → procedure B)

    - If there seem to be substantial environmental gradients in the samples, set up appropriate ecoregions. Ideally the ecoregions should be harmonised across all EC variables (for the same ET) – this can improve the consistency and reduce the complexity of the exercise. Alternatively, try to adjust the variables in a way that reduces the environmental gradients in the reference sites.

  - estimate the upper RLs based on the reference sample for each variable/ subtype / ecoregion:
    - procedure A: if the reference sample includes the RC, then use a simple statistical parameter (e.g., a high percentile) to estimate X_{100}. Alternatively, it might be more robust to use a suitably selected percentile to estimate X_{95} or X_{90}, and then estimate X_{100} by extrapolation. A good theoretical thumb rule for X_{100} is that, if an additional increase above (or decrease below) this value can still be seen as a meaningful improvement in condition, then the value is probably set too low (or high).
    - procedure B: If the sample does not include the RC (just a least disturbed condition) then use pressure removal / hindcasting to make a predictive estimation for the value of the EC variable for the RC.
  - In all cases the "validity domain" of the RL is given by the subtype & ecoregion for which it was determined.

- **Step 3** (for each variable): set all the "remaining" RLs (that are prescribed by your "condition model" – in anthropogenic ETs and for negative variables this means *all* of the RLs) using one of the following approaches (try the approaches in this order):
  - check if endpoint RLs (X_0, X_{100}) can be meaningfully set to the natural scale endpoints (natural zero values) of the original variable:
    - for positive variables this will be X_0 (and in (semi)natural ETs this is actually the only remaining endpoint RL, because X_{100} has already been set in Step2)
    - for negative variables this will typically be X_{100} (the optimal condition is where the negative variable is zero)
    - in some cases (variables measured as percentages of probabilities) the variable might have a second scale endpoint which might then be used as the other endpoint RL.
  - to set mid-range RLs (~status class boundaries), check if any of the following options are available:
    - policy thresholds and/or clear scientific thresholds (carefully examine the validity (subtype, ecoregion) and the indicator value (class boundary) to which this value could be assigned)

        - "other transferable thresholds": thresholds that have been identified as meaningful for other variables (EC, main pressure) for the given ET (subtype, ecoregion). It may be possible to "transfer" RLs across variables using the same representative sample that was used for "pressure removal" in Step 2.
    - if there are at least 2 RLs for a variable: use inter- or extrapolation to fill in the remaining RL values (as prescribed by the "condition model").

- **Step 4** (for each variable):
    - do the rescaling (using the right RLs – i.e., the ones that are valid for the given subtype and ecoregion)
    - Ensure transparency and clarity in reporting reference levels and rescaling functions:
        - "RL tables": one table per variable, listing RLs for each RL role (e.g., X_0, X_{60}, X_{100}, as in the condition model) and each ET subtype / ecoregion. In the case of mid-range RLs, the corresponding indicator thresholds also need to be specified. Table 8.1 provides a template that can be used to efficiently specify and document a rescaling function, together with the origins and justifications of the RL values identified. The development of reporting standards for the rescaling operation is an important (yet typically overlooked) step towards making the accounts fully transparent and scientifically reproducible.
        - "RL maps" for visualising the spatial changes in the RLs (if ecoregions were also applied)
        - metadata: for each individual RL: document how it was estimated.

Table 8.1. Recommended reporting template for documenting the reference levels (RL) applied in the context of a piecewise linear rescaling. (See an example for a filled in template in section 7.3.4.

Variable name: .... *Variable unit: ....* *ET code/name: ....* *ET subtype / ecoregion (if relevant): ....* *RC type: ....*			
	X value	**RL method***	**Source (and description/justification)**
X_0	0	e.g., Natural zero	A zero value that is meaningful as a "maximally degraded value"
…	…	…	
X_{60}	…	Policy threshold	Critical level of XXX in regulation YYY
X_{100}	…	…	

** See Chapter 6 for the full list of methods*

9 Conclusions

The SELINA reference framework is a key piece in the broader effort to develop systematic ecosystem condition assessment in the context of SEEA EA. It provides significant support towards the implementation of EU policies related to ecosystem condition by providing a structured and practical methodology to identify reference conditions, set reference levels and rescale EC variables in a way that is scientifically robust and internationally comparable.

Building upon the work started by SEEA EA and the "EU-wide methodology to map and assess ecosystem condition" on standardising and operationalising the concept of references (UN 2021, Vallecillo et al. 2022), the SELINA reference framework especially addresses the previously identified need for international comparability and the lack of coherence between different existing frameworks by harmonising the variable terminology and establishing a reference setting and rescaling workflow, compatible with SEEA EA principles, that can be used for any type of condition variable. This report discusses all the available methods that can be used for identifying meaningful reference conditions and reference levels in diverse ecological and policy contexts, explains the SEEA EA approach for rescaling condition variables into condition indicators and presents practical mathematical solutions for implementing the rescaling with the help of reference levels and added context from the WFD and MSFD frameworks. The difficulties in identifying meaningful references for highly anthropogenic ecosystems (e.g. urban systems and croplands), are addressed by recommending to "skip" defining a (fictive and potentially erroneous) reference condition and focusing on identifying RLs for each variable.

Tables 9.1 and 9.2 summarise the recommended reference conditions for each ET and methods available for quantification of optimal RLs.

Table 9.1. Recommended reference condition (RC) types for the main broad ETs

	Natural RC	Seminatural RC	Anthropogenic RC*	No RC (data-driven methods)*
Urban			x	
Cropland			x	
Grassland		x	(x)**	
Woodland and forest	x	x	(x)**	
Heathland and shrub	x	x		
Sparsely vegetated land	x	x		
Wetlands	x	x		
Rivers and lakes	x	x	(x)**	
Marine inlets and transitional waters	x	x		
Coastal	x	x		
Shelf	x			
Open ocean	x			

* While data-driven approaches avoid the concept of reference condition (RC), they are still presented in this table for the sake of completeness.
** Depending on the ET typology applied, these ETs may have completely anthropogenic subtypes for which an anthropogenic RC can be considered (c.f. also the handling of anthropogenic water bodies in WFD).

Table 9.2. Methods that are available in the case of the different RC types for the quantification of the optimal RLs of the EC variables

	Natural RC	Seminatural RC	Anthropogenic RC*	No RC (data-driven approach)*
RC methods				
Reference sites	x	x		
Pressure removal	x	x		
Paleo-environmental data	x	x		
RL methods				
Natural scale endpoints	x	x	x	
Policy thresholds	x	x	x	
Scientific thresholds	x	x	x	
Expert thresholds	x	x	x	
RL transfer	x	x	x	
Interpolations, extrapolations	x	x	x	
Data-driven methods				
Range standardisation (min-max method)				x

* While data-driven approaches avoid the concept of reference condition (RC), they are still presented in this table for the sake of completeness.

To cross the bridge between theory and practice, this report also provides relevant real-life examples from different ETs regarding different aspects of the framework and a step-by-step guide that can be easily followed by policy makers, researchers and experts regardless of their circumstances. It should be noted that while there are considerable differences in data availability and state of methodology between different ETs (freshwater and marine ETs standing out as the most advanced thanks to the WFD and the MSFD), elements of the framework – the reference condition, reference levels and different rescaling frameworks – are already well used in many different contexts in Europe and the wider world. The SELINA framework, while by no means the "ultimate answer" to all questions related to references in (and beyond), should prove to be a useful guideline for everyone working on the topic of ecosystem condition assessment, no matter where their "starting line" is.

Acknowledgements

We would like to thank several helpful reviewers from Norwegian Environment Agency and JRC who contributed valuable comments to earlier versions of this document, and everyone who participated in the survey to validate and contextualise the SELINA reference framework. Your active and constructive cooperation has significantly improved the quality and comprehensibility of this document.

References

Allain, S., Plumecocq, G., & Leenhardt, D. (2018). Spatial aggregation of indicators in sustainability assessments: Descriptive and normative claims. *Land Use Policy*, 76, 577–588. https://doi.org/10/gdzc6c

Álvarez-Ripado, A., Bruzón, A.G., Álvarez-García, D. and Arrogante-Funes, P. 2024. Assessing urban ecosystem condition account with object-based methods. *Ecological Indicators*, 168, 112727, ISSN 1470-160X. https://doi.org/10.1016/j.ecolind.2024.112727

Andreasen, J. K., O'Neill, R. V., Noss, R., & Slosser, N. C. (2001). Considerations for the development of a terrestrial index of ecological integrity. Ecological Indicators, 1(1), 21–35. https://doi.org/10.1016/S1470-160X(01)00007-3

Arcangeli, A., Authier, M., Baccetti, N., Borsani, J., Cachera, M., Canals, V., Casier, M., Cavenne, E., Claro, F., Valentina, C., Conner, D., Consalvo, I., Cardoso, A., Dalu, M., Dierschke, V., Ducommun, L., Fleming, V., Fortibuoni, T., Galgani, F. & Zenatello, M. (2022). European Commission, 2022. MSFD CIS Guidance Document No. 19, Article 8 MSFD, May 2022.

Astell-Burt, T. & Feng X. (2020) Urban green space, tree canopy and prevention of cardiometabolic diseases: a multilevel longitudinal study of 46 786 Australians. *International Journal of Epidemiology*, 49(3), pp. 926-933. doi: 10.1093/ije/dyz239.

Babí Almenar, J., Marando, F., Vallecillo, S., Zulian, G., Cortinovis, C., Zurbaran-Nucci, M., Chrysoulakis, N., Parastatidis, D., Heris, M. & Grammatikopoulou, I. (2023). Urban Ecosystem accounts following the SEEA EA standard: a pilot application in Europe. *Luxembourg: Publications Office of the European Union*. https://doi.org/10.2760/741116

Bana e Costa, C.A., Ensslin, L., Cornêa, E.C. & Vansnick, J.-C. (1999). Decision Support Systems in action: Integrated application in a multicriteria decision aid process. *European Journal of Operational Research,* 113, pp. 315-335. https://doi:10.1016/S0377-2217(98)00219-7

Bennion, H., Battarbee, R. W., Sayer, C. D., Simpson, G. L., & Davidson, T. A. (2011). Defining reference conditions and restoration targets for lake ecosystems using palaeolimnology: A synthesis. *Journal of Paleolimnology*, 45(4), 533–544. https://doi.org/10.1007/s10933-010-9419-3

Bigaret, S., Hodgett, R. E., Meyer, P., Mironova, T., & Olteanu, A.-L. (2017). Supporting the multi-criteria decision aiding process: R and the MCDA package. *EURO Journal on Decision Processes*, 5(1), 169–194. https://doi.org/10.1007/s40070-017-0064-1

Birk, S., Willby, N. J., Kelly, M. G., Bonne, W., Borja, A., Poikane, S., & Van De Bund, W. (2013). Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. *Science of The Total Environment*, 454–455, 490–499. https://doi.org/10.1016/j.scitotenv.2013.03.037

BMEL (German Federal Ministry of Food and Agriculture) (2024). CAP-Strategic Plan for the Federal Republic of Germany. V5.1, C(2024)7250. 2216pp. https://www.bmel.de/SharedDocs/Downloads/DE/_Landwirtschaft/EU-Agrarpolitik-Foerderung/gap-strategieplan-version-5-1.pdf?__blob=publicationFile&v=3

Borisova, B., Semerdzhieva, L., Dimitrov, S., Valchev, S., Iliev, M., & Georgiev, K. (2024). Geospatial Prioritization of Terrains for "Greening" Urban Infrastructure. *Land*, 13(9), 1487. https://doi.org/10.3390/land13091487

Browning, M. H. E. M., Locke, D. H., Konijnendijk, C., Labib, S. M., Rigolon, A., Yeager, R., ... & Nieuwenhuijsen, M. (2024). Measuring the 3-30-300 rule to help cities meet nature access thresholds. *Science of the Total Environment*, 907, 167739.

Bruelheide, H., Jandt, U., Marmol, A., Fernandez, N., Smets, B., Buchhorn, M., Giagnacovo, L., Milli, G., Jiménez-Alfaro, B. & Alvarez-Martinez, J.M. (2024). D5.2 Past-to-present EBV modelled datasets and status indicator for selected terrestrial habitats in the Habitats Directive. *Martin Luther University Halle-Wittenberg (MLU); Flemish Institute for Technological Research (VITO); University of Oviedo (UniOvi); Europa Biodiversity Observation Network: integrating data streams to support policy (EUROPABON)* https://doi.org/10.3897/arphapreprints.e128158

Brunsdon, C. (2018). Quantitative methods III: Scales of measurement in quantitative human geography. *Progress in Human Geography*, 42(4), 610–621. https://doi.org/10.1177/0309132517717008

Bruzón, A. G., Arrogante-Funes, P., & Santos-Martín, F. (2023). Modelling and testing forest ecosystems condition account. *Journal of Environmental Management*, 345, 118676.

Buffagni, A., Erba, S., and Furse, M. T. (2007). A simple procedure to harmonize class boundaries of assessment systems at the pan-European scale. *Environmental Science & Policy*, 10(7), 709–724. https://doi.org/10.1016/j.envsci.2007.03.005

Burge, O. R., Richardson, S. J., Wood, J. R., & Wilmshurst, J. M. (2023). A guide to assess distance from ecological baselines and change over time in palaeoecological records. *The Holocene*, 09596836231169986. https://doi.org/10.1177/09596836231169986

Carli, E., Giarrizzo, E., Burrascano, S., Alós, M., Del Vico, E., Di Marzio, P., Facioni, L., Giancola, C., Mollo, B., Paura, B., Salerno, G., Zavattero, L. & Blasi, C. (2018) Using vegetation dynamics to face the challenge of the conservation status assessment in semi-natural habitats. Rendiconti Lincei. *Scienze Fisiche e Naturali,* 29, 363–374. https://doi.org/10.1007/s12210-018-0707-6

Chan, L., Hillel, O., Werner, P., Holman, N., Coetzee, I., Galt, R., & Elmqvist, T. (2021). Handbook on the Singapore Index on Cities' Biodiversity (also known as the City Biodiversity Index). *Montreal: Secretariat of the Convention on Biological Diversity and Singapore: National Parks Board, Singapore*. 70 Pages

Chen, K., Olson, J. R., Vander Laan, J. J., Hill, R. A., Wang, B., & Hawkins, C. P. (2019a). Improving the performance of ecological indices by balancing reference site quality and representativeness. *Hydrobiologia*, 837(1), 177–194. https://doi.org/10.1007/s10750-019-3970-3

Chen, K., Rajper, A. R., Hughes, R. M., Olson, J. R., Wei, H., & Wang, B. (2019b). Incorporating functional traits to enhance multimetric index performance and assess land use gradients. *Science of the Total Environment*, 691, 1005–1015. https://doi.org/10.1016/j.scitotenv.2019.07.047

Cimburova, Z. & Barton, D.N. (2021). Testing GIS data-driven mapping and valuation of recreation areas in Oslo. *NINA Report* 1931. Norwegian Institute for Nature Research.

Courtney Jones, S.K., Geange, S.R., Hanea, A., Camac, J., Hemming, V., Doobov, B., Leigh, A. & Nicotra, A.B. (2023). IDEAcology: An interface to streamline and facilitate efficient, rigorous expert elicitation in ecology. *Methods in Ecology and Evolution*, 14, 2019-2028.

Czúcz, B., Keith, H., Driver, A., Jackson, B., Nicholson, E., & Maes, J. (2021a). A common typology for ecosystem characteristics and ecosystem condition variables. *One Ecosystem*, 6, e58218. https://doi.org/10/gjsn2c

Czúcz, B., Keith, H., Maes, J., Driver, A., Jackson, B., Nicholson, E., Kiss, M., & Obst, C. (2021b). Selection criteria for ecosystem condition indicators. *Ecological Indicators*, 133, 108376. https://doi.org/10.1016/j.ecolind.2021.108376

Czúcz, B., Simensen, T., Skrindo, A. B., Rusch, G. M., Nybø, S., Grainger, M., & Töpper, J. P. (2025). No net loss accounting: Aligning biodiversity offsets with ecosystem accounts. *Ecological Solutions and Evidence*, 6(1), e70006. https://doi.org/10.1002/2688-8319.70006

Destatis (2023). *Methode der Zustandsbilanzierung der Ökosysteme*. 64pp. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/UGR/oekosystemgesamtrechnungen/Publikationen/Downloads/methode-zustandsbilanzierung-5853202239004.pdf?blob=publicationFile&v=9

Dias, E. & Pereira, D. (2022). *Cartography of the terrestrial habitats of the Natura 2000 Network of the Azores by the LIFE IP AZORES NATURA Project* (LIFE17 IPE/PT/000010) – 2022. Available from: https://portal.azores.gov.pt/web/drpm/life-ip-azores-natura [upon request] (May 29, 2025).

Dias, E., Mendes, C., Branco, S. (2022). Índice de Naturalidade aplicado à vegetação dos Açores e Habitats de Turfeiras. *Gabinete de Ecologia Vegetal e Aplicada (GEVA)*. CBA – Universidade dos Açores. Produced for the LIFE IP AZORES NATURA Project (LIFE17 IPE/PT/000010). Available from: https://portal.azores.gov.pt/web/drpm/life-ip-azores-natura [upon request] (May 29, 2025).

Dudley, N. (Ed.). (2008). Guidelines for applying protected area management categories. IUCN.

DV (Direktoratsgruppa Vanndirektivet) (2018). Veileder 02: 2018. Klassifisering av miljøtilstand i vann. Økologisk og kjemisk klassifiseringssystem for kystvann, grunnvann, innsjøer og elver (p. 220). Miljødirektoratet.

Ellis, B. (1966). Basic Concepts of Measurement. *Cambridge University Press*, Cambridge, UK.

ETC/BD (2006). The indicative map of Euro-pean biogeographical regions: Methodology and development. ETC/BD.

ETC/BD (2006). The indicative map of Euro-pean biogeographical regions: Methodology and development. ETC/BD. https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-2005/methodology-description-pdf-format/methodology-description-pdf-format/download

EEA (2019). Marine messages II, Navigating the course towards clean, healthy and productive seas through implementation of an ecosystem based approach. EEA Report 17/2019. ISBN: 978-92-9480-197-5

EEA (2022). Soil monitoring in Europe – Indicators and thresholds for soil health assessments. EEA Report 08/2022, 181 pp.

EC (2003a). Identification of Water Bodies. WFD CIS Guidance Document No. 2. Published by the Directorate General Environment of the European Commission, Brussels, ISBN No. 92-894-5123-8, ISSN No. 1725-1087.

EC (2003b). River and lakes – Typology, reference conditions and classification systems. WFD CIS Guidance document No 10 produced by WFD CIS Working Group 2.3 (REFCOND). Office for Official Publications of the European Communities, Luxembourg.

EC (2003c). Identification and designation of heavily modified and artificial water bodies. WFD CIS Guidance Document No. 4. produced by EU Working Group 2.2. Luxemburg: Office for Official Publications of the European Communities

EC (2011). Guidance on the intercalibration process 2008-2011. WFD CIS Guidance document No 14, WFD-CIS. Office for Official Publications of the European Communities, Luxembourg.

EC (2015). Procedure to fit new or updated classification methods to the results of a completed intercalibration exercise. WFD CIS Guidance document No 30; Technical Report 2015-085. Office for Official Publications of the European Communities, Luxembourg.

EC (2020). EU COMMISSION STAFF WORKING DOCUMENT Review of the status of the marine environment in the European Union Towards clean, healthy and productive oceans and seas Accompanying the Report from the Commission to the European Parliament and the Council on the implementation of the Marine Strategy Framework Directive (Directive 2008/56/EC). SWD/2020/61 final

EC, JRC, Nikolaidis, N., Poikane, S., Bouraoui, F., Salas Herrero, M.F., Free, G., Varkitzi, I. and Van De Bund, W. (2024). Comparison and harmonization of eutrophication assessment criteria under the WFD and the Nitrate Directives. Publications Office of the European Union, Luxembourg. https://data.europa.eu/doi/10.2760/23475, JRC137238

EU (2005). Overall approach to the classification of ecological status and ecological potential. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance document n.o 13. Office for Official Publications of the European Communities, Luxembourg.

EU (2017). COMMISSION DECISION (EU) 2017/848 of 17 May 2017 laying down criteria and methodological standards on good environmental status of marine waters and specifications and standardised methods for monitoring and assessment, and repealing Decision 2010/477/EU. Official Journal of the European Union L125/43

EU (2024a). Regulation (EU) 2024/1991 of the European Parliament and of the Council of 24 June 2024 on nature restoration and amending Regulation (EU) 2022/869. Official Journal of the European Union, 1991, 1.

EU (2024b). Directive 2024/2881 of the European Parliament and of the Council of 23 October 2024 on ambient air quality and cleaner air for Europe (recast) 2024/2881. *Official Journal of the European Union*, 2024, 11. http://data.europa.eu/eli/dir/2024/2881/oj

EUWI (2020). Guidance document on the methodology for the development of a WFD-compliant ecological status classification system (No. ENI/2016/372-403; p. 22). *Office International de l'Eau* (IOW). https://www.euwipluseast.eu/images/2021/09/20-11-05_EUWI_Guidance_on_ESCS_final.pdf

Fisher, M., & Keil, F. C. (2018). The Binary Bias: A Systematic Distortion in the Integration of Information. *Psychological Science*, 29(11), 1846–1858. https://doi.org/10.1177/0956797618792256

Fleming-Lehtinen, V. & Laamanen, M., (2012). Long-term changes in Secchi depth and the role of phytoplankton in explaining light attenuation in the Baltic Sea. *Estuarine, Coastal and Shelf Science* 102-103:1-10.

Fleming-Lehtinen, V. (2016). Secchi depth in the Baltic Sea – an indicator of eutrophication. University of Helsinki, Faculty of Biological and Environmental Sciences, Helsinki. 42 pages.

Forbes et al., 2018. Shifting baselines in a changing world: Identifying management targets in endangered heathlands of the Cape Floristic Region, South Africa. *Anthropocene*, 22:8193. https://doi.org/10.1016/j.ancene.2018.05.001

Framstad, E., Kolstad, A. L., Nybø, S., Töpper, J. & Vandvik, V. (2022). The condition of forest and mountain ecosystems in Norway. Assessment by the IBECA method. *NINA Report* 2100. Norwegian Institute for Nature Research.

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., & Granger, C. B. (2015). *Fundamentals of Clinical Trials (5th edition)*. Springer.

Garibaldi, L. A., Oddi, F. J., Miguez, F. E. et al., (2020). Working landscapes need at least 20% native habitat. *Conservation Letters*, 14 (2), e12773, https://doi.org/10.1111/conl.12773

Gauer J. & Kroiher F. (2012). Waldökologische Naturräume Deutschlands– Forstliche Wuchsgebiete und Wuchsbezirke Digitale Topographische Grundlagen – Neubearbeitung Stand 2011. Sonderheft 359.

Johann Heinrich von Thünen-Institut Bundesforschungsinstitut für Ländliche Räume, Wald und Fischerei (vTI), https://literatur.thuenen.de/digbib_extern/dn050265.pdf
German Institute for Standardization (2020). Soil description – Terms, definitions, applications. *Beuth Verlag*. DIN 4220:2020-11. https://dx.doi.org/10.31030/3181018
Gorrod, E. J., Bedward, M., Keith, D. A., & Ellis, M. V. (2013). Systematic underestimation resulting from measurement error in score-based ecological indices. *Biological Conservation*, 157, 266–276. https://doi.org/10.1016/j.biocon.2012.09.002
Heitkamp, F., Glatzel, S., Michalzik, B., Fischer, E. & Gerold, G. (2008). Soil microbiochemical properties as indicators for success of heathland restoration after military disturbance. *Land Degrad*, Dev., 19: 408-420. https://doi.org/10.1002/ldr.848
HELCOM (2012). Development of a set of core indicators: Interim report of the HELCOM CORESET project. PART A. Description of the selection process. Balt. Sea Environ. Proc. No. 129A. Available at: https://helcom.fi/wp-content/uploads/2019/08/BSEP129A.pdf
HELCOM (2013). HELCOM core indicators: Final report of the HELCOM CORESET project. *Balt. Sea Environ. Proc.* No. 136. Available at: https://helcom.fi/wp-content/uploads/2019/08/CORESET_Final_BSEP136.pdf
HELCOM (2023). Abundance of coastal fish key functional groups. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/coastal-fish-key-groups/]. ISSN 2343-2543
HELCOM (2023). Abundance of coastal fish key species. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/coastal-fish-key-species/]. ISSN 2343-2543
HELCOM (2023) Abundance and population trends of harbour porpoises. *HELCOM pre-core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/harbour-porpoises-abundance/].
HELCOM (2023). Abundance of waterbirds in the breeding season. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/waterbirds-breeding-season/]. ISSN 2343-2543.
HELCOM (2023). Abundance of waterbirds in the wintering season. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/waterbirds-wintering-season/]. ISSN 2343-2543.
HELCOM (2023) Cumulative impact from physical pressures on benthic biotopes. *HELCOM indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/cumulative-impact-on-benthic-habitats/]. ISSN 2343-2543
HELCOM (2023). Diatom/Dinoflagellate index. *HELCOM pre-core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/diatom-dinoflagellate-index/]. ISSN 2343-2543
HELCOM (2018) Nutritional status of marine mammals. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/nutritional-status-of-seals/]. ISSN 2343-2543
HELCOM (2023) Population trends and abundance of seals. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/grey-seal-abundance/]. ISSN 2343-2543
HELCOM, (2023). State of the Baltic Sea 2023 — The third HELCOM holistic assessment (HOLAS 3) report, 2023. HELCOM (Helsinki: European Commission).
HELCOM (2023). Seasonal succession of functional phytoplankton groups. *HELCOM pre-core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/phytoplankton/]. ISSN: 2343-2543
HELCOM (2023). Size structure of coastal fish (Coastal fish size). *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/coastal-fish-size/]. ISSN 2343-2543.
HELCOM (2023). State of the soft-bottom macrofauna community. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/soft-bottom-macrofauna/]. ISSN 2343-2543
HELCOM (2023) White-tailed sea eagle productivity. *HELCOM core indicator report.* Online. [11.4.2025], [https://indicators.helcom.fi/indicator/white-tailed-sea-eagle/]. ISSN 2343-2543
HELCOM (2023) Zooplankton mean size and total stock. *HELCOM core indicator report*. Online. [11.4.2025], [https://indicators.helcom.fi/indicator/zooplankton/]. ISSN 2343-2543
HELCOM Indicator Manual - 2025 update. (2025) Baltic Sea Environment Proceedings n° 175. HELCOM.
Helm, A., Kull, A., Veromann, E., Remm, L., Villoslada, M., Kikas, T., Aosaar, J., Tullus, T., Prangel, E., Linder, M., Otsus, M., Külm, S., Sepp, K., 2020 (täiend. 2021). Metsa-, soo-, niidu- ja põllumajanduslike ökosüsteemide seisundi ning ökosüsteemiteenuste baastasemete üleriigilise hindamise ja kaardistamise lõpparuanne. ELME projekt. Tellija: Keskkonnaagentuur (riigihange nr 198846). https://loodusveeb.ee/sites/default/files/inline-files/elme-ost-baastasemed_l6pparuanne_14-06-21.pdf
Hinsch, M., Zulian, G., Stekker, S. et al., 2024. Assessing pollinator habitat suitability considering ecosystem condition in the Hannover Region, Germany. *Landscape Ecology*, 39, 47. https://doi.org/10.1007/s10980-024-01851-x

Huber, S., Prokop, G., Arrouays, D., Banko, G., Bispo, A., Jones, R. J. A., Kibblewhite, M. G., Lexer, W., Möller, A., Rickson, R. J., Shishkov, T., Stephens, M., Toth, G., Van den Akker, J. J. H., Varallyay, G., Verheijen, F. G. A. & Jones, A. R. (eds.), 2008. *Environmental Assessment of Soil for Monitoring: Volume I Indicators & Criteria*. EUR 23490 EN/1, Office for the Official Publications of the European Communities, Luxembourg, 339pp. https://esdac.jrc.ec.europa.eu/projects/Envasso/documents/ENV_Vol-I_Final2_web.pdf

Håkansson, I., Grath, I. and Olsen, H.J. (1996). Influence of machinery traffic in Swedish farm fields on penetration resistance in the subsoil. Swedish Journal of Agricultural Research 26, 181-187. International Council for the Exploration of the Sea (ICES). (2025). Latest advice, ICES website. https://www.ices.dk/advice/Pages/Latest-Advice.asp

Jakobsson, S., Töpper, J.P., Evju, M., Framstad, E., Lyngstad, A., Pedersen, B., Sickel, H., Sverdrup-Thygeson, A., Vandvik, V., Velle, L.G., Aarrestad, P.A. & Nybø, S. (2020). Setting reference levels and limits for good ecological condition in terrestrial ecosystems – Insights from a case study based on the IBECA approach. Ecological Indicators, 116, 106492.

Jakobsson, S., Evju, M., Framstad, E., Imbert, A., Lyngstad, A., Sickel, H., Sverdrup-Thygeson, A., Töpper, J. P., Vandvik, V., Velle, L. G., Aarrestad, P. A., & Nybø, S. (2021). Introducing the index-based ecological condition assessment framework (IBECA). Ecological Indicators, 124, 107252. https://doi.org/10/gh2qhj

JRC: Institute for Environment and Sustainability, Huber, S., Prokop, G. and Arrouays, D., (2008). Environmental assessment of soil for monitoring. Volume I, Indicators & criteria, Publications Office of the European Union, https://data.europa.eu/doi/10.2788/93515

Keith, H., Czúcz, B., Jackson, B., Driver, A., Nicholson, E., & Maes, J. (2020). A conceptual framework and practical structure for implementing ecosystem condition accounts. One Ecosystem, 5, e58216. https://doi.org/10/ghjjbx

Kervinio, Y., Surun, C., Comte, A., & Levrel, H. (2023). Defining ecological liabilities and structuring ecosystem accounts to support the transition to sustainable societies. One Ecosystem, 8, e98100. https://doi.org/10.3897/oneeco.8.e98100

Kibblewhite, M.G., Jones, R.J.A., Montanarella, L., Baritz, R., Huber, S., Arrouays, D., Micheli, E. and Stephens, M. (eds) (2008). Environmental Assessment of Soil for Monitoring Volume VI: Soil Monitoring System for Europe. EUR 23490 EN/6, Office for the Official Publications of the European Communities Luxembourg, 72pp. https://esdac.jrc.ec.europa.eu/Projects/Envasso/documents/ENV_Vol-VI_Final2_web.pdf

Kilgour, B. W., & Stanfield, L. W. (2006). Hindcasting Reference Conditions in Streams. American Fisheries Society Symposium, 48, 1–18.

King, G., & Powell, E. N. (2008). How Not to Lie Without Statistics. https://gking.harvard.edu/files/abs/nolie-abs.shtml

Kolstad, A.L., Grainger, M.J., Sandvik, H., Venter, Z., Töpper, J., Speed, J.D.M, Nater, C.R., Evju, M., Åström, J., Davey, M. (2023) Indicators for Ecosystem Condition in Norway. https://doi.org/10.5281/zenodo.7854835

Konijnendijk, Cecil. (2021). The 3-30-300 Rule for Urban Forestry and Greener Cities. 4.

Kourdounouli, C., & Jönsson, A. M. (2019). Urban ecosystem conditions and ecosystem services – a comparison between large urban zones and city cores in the EU. Journal of Environmental Planning and Management, 63(5), 798–817. https://doi.org/10.1080/09640568.2019.1613966

Korpinen, S., Laamanen, L., Bergström, L., Nurmi, M., Andersen, J.H., Haapaniemi, J., Harvey, E.T., Murray, C.J., Peterlin, M., Kallenbach, E., Klančnic, K., Stein, U., Tunesi, L., Vaughan, D. & Reker, J. (2021) Combined effects of human pressures on Europe's marine ecosystems. Ambio 50:1325–1336, https://doi.org/10.1007/s13280-020-01482-x

Kotsev, T., Tcherkezova, E. and Stoyanova, V. (2019). A concept of an integrated geodatabase for surface water, soil and groundwater pollution with arsenic in the upper part of Ogosta valley, northwestern Bulgaria. European Journal of Geography 10:6-23 https://eurogeojournal.eu/articles/1_Tcherkezova_et_al.pdf

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., ... & Peng, S. (2018). Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015). Earth System Science Data, 10(1), 219-234.

Lopez, D.R., Cavallero, L., Brizuela, M.A. & Aguiar, M.R. (2011). Ecosystemic structural-functional approach of the state and transition model. Applied Vegetation Science, 14, 6-16.

Maes, J., Bruzón, A. G., Barredo, J. I., Vallecillo, S., Vogt, P., Rivero, I. M., & Santos-Martín, F. (2023). Accounting for forest condition in Europe based on an international statistical standard. Nature Communications, 14(1), Article 1. https://doi.org/10.1038/s41467-023-39434-0
Maciejewski L., Seytre L., Van Es J., Dupont P., (2015). État de conservation des habitats agropastoraux d'intérêt communautaire, Méthode d'évaluation à l'échelle du site. Guide d'application. Version 3. Avril 2015. Rapport SPN 2015 - 43, Service du patrimoine naturel, Muséum national d'Histoire naturelle, Paris. 194 pp.
Magliozzi, C., Druon, J. N., Palialexis, A., Artigas, L., Boicenco, L., González-Quirós, R., ... & Varkitzi, I. (2021). Pelagic habitats under MSFD D1: current approaches and priorities: An overview of approaches towards D1C6 assessment.
Marcos, B., Gonçalves, J., Alcaraz-Segura, D., Cunha, M., & Honrado, J. P. (2023). Assessing the resilience of ecosystem functioning to wildfires using satellite-derived metrics of post-fire trajectories. Remote Sensing of Environment, 286, 113441.
Mazziotta, M., & Pareto, A. (2022). Normalization methods for spatio-temporal analysis of environmental performance: Revisiting the Min–Max method. *Environmetrics*, 33(5), e2730. https://doi.org/10.1002/env.2730
Metzger, M. J., Bunce, R. G. H., Jongman, R. H., Mücher, C. A., & Watkins, J. W. (2005). A climatic stratification of the environment of Europe. Global ecology and biogeography, 14(6), 549-563.
Metzger, M. J. (2018). The Environmental Stratification of Europe, [dataset]. University of Edinburgh. https://doi.org/10.7488/ds/2356
Miller, S., Eldred, P., Muldoon, A., Anlauf-Dunn, K., Stein, C., Hubler, S., Merrick, L., Haxton, N., Larson, C., Rehn, A., Ode, P., & Vander Laan, J. (2016). A Large-Scale, Multiagency Approach to Defining a Reference Network for Pacific Northwest Streams. Environmental Management, 58(6), 1091–1104. https://doi.org/10.1007/s00267-016-0739-6
Mistarz, M. & Grivel, L. (2020). - Évaluation de l'état de conservation des landes humides d'intérêt communautaire. Cahiers d'évaluation à l'échelle des sites Natura 2000. Version 1. UMS Patrinat – OFB/CNRS/MNHN. 88p.
Morrone, J. J. (2018). The spectre of biogeographical regionalization. Journal of Biogeography, 45(2), 282–288. https://doi.org/10.1111/jbi.13135
Nater, C. and Pavón-Jordán D. In: Kolstadt et al., (2023). Indicators for Ecosystem Condition in Norway, chapter 11 (Breeding bird indicators). https://ninanor.github.io/ecosystemCondition/breeding-birds.html
Nagy, G.G., Ladányi, M., Arany, I., Aszalós, R., Czúcz, B., (2017). Birds and plants: Comparing biodiversity indicators in eight lowland agricultural mosaic landscapes in Hungary. Ecological Indicators 73, 566–573. https://doi.org/10.1016/j.ecolind.2016.09.053
Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD. https://doi.org/10.1787/533411815016
Nicholson Thomas, I. Lecomte, X., Grêt-Regamey, A., Roche P. (2025). SELINA D3.2. Derive a minimum set of key ecosystem condition indicators per ecosystem type. SELINA project, https://project-selina.eu
Nygård, H., Murray, C., Andersen, J.H., Martin, G., Torn, K. & Korpinen, S. (2018). BEAT 3.0 – a Tool for Integrated Biodiversity Assessments. *Journal of Open Research Software*, 6(1), p.19. DOI: http://doi.org/10.5334/jors.226
O'Hagan, A. (2019). Expert knowledge elicitation: subjective but scientific. *The American Statistician*, 73, 69-81.
Oslo Kommune (2023). Norm for vegetasjon og vannhåndtering (blågrønn faktor). Available at: https://www.oslo.kommune.no/getfile.php/13494305-1698840957/Tjenester%20og%20tilbud/Plan%2C%20bygg%20og%20eiendom/Byggesaksveiledere%2C%20normer%20og%20skjemaer/Norm%20vegetasjon%20vannh%C3%A5ndtering%20-%20bl%C3%A5gr%C3%B8nn%20faktor.pdf
Parrochia D. (2019). Classification. In: The Internet Encyclopedia of Philosophy (ISSN 2161-0002). https://www.iep.utm.edu/classifi. Accessed on: 2019-5-30.
Pedersen, B., & Skarpaas, O. (2012). Statistiske egenskaper til Naturindeks for Norge. Usikkerhet i datagrunnlaget og sensitivitet. In 55 s. Norsk institutt for naturforskning. https://brage.nina.no/nina-xmlui/handle/11250/2385833
Pedersen, B., Nybø, S., & Sæther, S. A. (2016). Nature Index for Norway 2015. Ecological framework, computational methods, database and information systems. In 84 pp. https://brage.nina.no/nina-xmlui/handle/11250/2374610

Pereira, P. S., Souza, N. F., Baptista, D. F., Oliveira, J. L. M., & Buss, D. F. (2016). Incorporating natural variability in the bioassessment of stream condition in the Atlantic Forest biome, Brazil. Ecological Indicators, 69, 606–616. https://doi.org/10.1016/j.ecolind.2016.05.031

Piqueray, J., Bottin, G., Delescaille, L.-M., Bisteau, E., Colinet, G. and Mahy, G. (2011). Rapid restoration of a species-rich ecosystem assessed from soil and vegetation indicators: The case of calcareous grasslands restored from forest stands. Ecological Indicators, 11 (2), pp. 724-733. ISSN 1470-160X. https://doi.org/10.1016/j.ecolind.2010.06.007

Ramon, P., Andersen, J.H., Murray, C.H. & Frigstad, H. (2025). Temporal and spatial trends in eutrophication status in the Oslofjord, Norway. Science of The Total Environment 978: 179462. DOI: https://doi.org/10.1016/j.scitotenv.2025.179462

Rastetter, E. B., King, A. W., Cosby, B. J., Hornberger, G. M., O'Neill, R. V., & Hobbie, J. E. (1992). Aggregating Fine-Scale Ecological Knowledge to Model Coarser-Scale Attributes of Ecosystems. Ecological Applications, 2(1), 55–70. https://doi.org/10.2307/1941889

Rendon, P., Watson, M., Ruf, K., Peters, A., Kleeschulte, S., Santos, F. (2023). SELINA D3.1. Integrating data streams to define and map ecosystem types. SELINA project, https://project-selina.eu

Rowland, J. A., Bland, L. M., James, S., & Nicholson, E. (2021). A guide to representing variability and uncertainty in biodiversity indicators. *Conservation Biology*, cobi.13699. https://doi.org/10/gh2bqm

Rutgers, M., Mulder, C., & Schouten, A. J. (Eds.) (2008). Soil ecosystem profiling in the Netherlands with ten references for biological soil quality. RIVM Report 607604009/2008.

Rygg B. (2006). Developing indices for quality-status classification of marine soft-bottom fauna in Norway. NIVA-rapport 5208/2006

Schartau, A.K., Gundersen, H. & Pedersen, B. (2016). Likheter og forskjeller i vannforskriftens klassifisering og naturindeksen for ferskvann/kystvann. Evaluering av metodikk og datagrunnlag. NINA Kortrapport 5. Norsk institutt for naturforskning. http://hdl.handle.net/11250/2384734

Schmidt, J., Fassnacht, F.E., Förster, M. and Schmidtlein, S. (2018). Synergetic use of Sentinel-1 and Sentinel-2 for assessments of heathland conservation status. Remote Sens Ecol Conserv, 4: 225-239. https://doi.org/10.1002/rse2.68

Schneider, S. & Lindstrøm, E.A. (2011). The periphyton index of trophic status PIT: A new eutrophication metric based on nondiatomaceous benthic algae in Nordic rivers. Hydrobiologia 665:143–155.

Serrano, A. R., Peco, B., Morillo, J. A. and Ochoa-Hueso, R. (2024). Abandonment of traditional livestock grazing reduces soil fertility and enzyme activity, alters soil microbial communities, and decouples microbial networks, with consequences for forage quality in Mediterranean grasslands. Agriculture, Ecosystems & Environment, 366. 108932, ISSN 0167-8809. https://doi.org/10.1016/j.agee.2024.108932

Skärby, L., Ro-Poulsen, H., Wellburn, F.A., Sheppard, L.J., (1998). Impacts of ozone on forests: a European perspective. New Phytol. 139 (1), 109–122.

Song, X. P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. Nature, 560(7720), 639-643.

Soranno, P. A., Cheruvelil, K. S., Stevenson, R. J., Rollins, S. L., Holden, S. W., Heaton, S., & Torng, E. (2008). A framework for developing ecosystem-specific nutrient criteria: Integrating biological thresholds with predictive modeling. Limnology and Oceanography, 53(2), 773–787. https://doi.org/10.4319/lo.2008.53.2.0773

Stange, E.E., Barton, D.N., Andersson, E. and Haase, D. (2022). Comparing the implicit valuation of ecosystem services from nature-based solutions in performance-based green area indicators across three European cities. Landscape and Urban Planning, 219. 104310, ISSN 0169-2046. https://doi.org/10.1016/j.landurbplan.2021.104310

Statistisches Bundesamt (Destatis). (2023). Methode der Zustandsbilanzierung der Ökosysteme. Statistisches Bundesamt. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/UGR/oekosystemgesamtrechnungen/Publikationen/Downloads/methode-zustandsbilanzierung-5853202239004.pdf. Available in German

Stoddard, J. L., Larsen, D. P., Hawkins, C. P., Johnson, R. K., & Norris, R. H. (2006). Setting Expectations for the Ecological Condition of Streams: The Concept of Reference Condition. Ecological Applications, 16(4), 1267–1276. https://doi.org/10.1890/1051-0761(2006)016[1267:SEFTEC]2.0.CO;2

Stoddard, J. L., Herlihy, A. T., Peck, D. V., Hughes, R. M., Whittier, T. R., & Tarquinio, E. (2008). A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society, 27(4), 878–891. https://doi.org/10.1899/08-053.1

Szulczewska, B., Giedych, R., Borowski, J., Kuchcik, M., Sikorski, P., Mazurkiewicz, A. and Stańczyk, T. (2014).

How much green is needed for a vital neighbourhood? In search for empirical evidence. Land Use Policy, 38, pp. 330-345. ISSN 0264-8377. https://doi.org/10.1016/j.landusepol.2013.11.006
Tanács, E., Bede-Fazekas, Á., Báldi, A., (2024). Surprisingly good fit of pressure-based cropland condition map and bird census data at the national scale. Ecological Indicators 167, 112665. https://doi.org/10.1016/j.ecolind.2024.112665
UN (United Nations), European Union, Food and Agriculture Organization of the United Nations, International Monetary Fund, Organisation for Economic Co-operation and Development, & The International Bank for Reconstruction and Development/The World Bank (2024). System of Environmental Economic Accounting - Ecosystem Accounting (Statistical Papers No. Series F No. 124). New York.
Vallecillo, R. S., Maes, J., Teller, A., Babi, A. J., Barredo, C. J. I., Trombetti, M., Abdul, M. D., Paracchini, M.-L., Carré, A., Addamo, A., Czucz, B., Zulian, G., Marando, F., Erhard, M., Liquete, G. M. D. C., Romao, C., Polce, C., Pardo, V. A., Jones, A., … Gumbert, A. (2022). EU-wide methodology to map and assess ecosystem condition. https://doi.org/10.2760/13048
VanderMeulen, M. A., Hudson, A. J., & Scheiner*, S. M. (2001). Three evolutionary hypotheses for the hump-shaped productivity–diversity curve. Evolutionary Ecology Research, 3(4), 379–392.
Vrasdonk, E., Palme, U., & Lennartsson, T. (2019). Reference situations for biodiversity in life cycle assessments: Conceptual bridging between LCA and conservation biology. The International Journal of Life Cycle Assessment, 24(9), 1631–1642. https://doi.org/10.1007/s11367-019-01594-x
Wallace, K. J., & Jago, M. (2017). Category mistakes: A barrier to effective environmental management. Journal of Environmental Management, 199, 13–20. https://doi.org/10.1016/j.jenvman.2017.05.029
Wurtzebach, Z., & Schultz, C. (2016). Measuring Ecological Integrity: History, Practical Applications, and Research Opportunities. BioScience, 66(6), 446–457. https://doi.org/10.1093/biosci/biw037
Zoltán, L., Tanács, E. and Standovár, T. (2023). Validation and limitations of large-scale forest condition indicators – An example from Hungary. Ecological Indicators 154, 110539. https://doi.org/10.1016/j.ecolind.2023.110539

SELINA

SCIENCE FOR EVIDENCE-BASED AND SUSTAINABLE DECISIONS ABOUT NATURAL CAPITAL

https://project-selina.eu/

Annex 1: Glossary of key terms

Baseline: A value of a time series variable to which other (subsequent) values are compared. Baselines are typically associated with the initial point in time of the time series. (new)

Benchmark: An extreme value of a variable measured across a range of similar 'objects', to which other objects are compared. Benchmarks are typically associated with 'best in class' performances. (new)

Continuous indicator scale: A measurement scale used for dimensionless ecosystem condition indicators, that covers all real numbers in the [0,1] interval, with the endpoints anchored to normative concepts (0: bad, 1: good). In the context of a rescaling operation, this scale corresponds to the 'transformed scale'. (new)

Directionality: The capacity of a variable for conveying a clear 'normative' meaning. A variable has a clear directionality if for any change (increase, decrease) in its value it can clearly be decided if the change is an improvement or a decline. (based on Czúcz et al., 2021b)

Discrete indicator scale: An ordinal scale used for characterising ecosystem condition. Many major EU biodiversity policies (e.g., WFD, MSFD, HD) use discrete indicator scales. The categories of a discrete indicator scale are often called 'status classes'. (new)

Ecosystem characteristics: The system properties of the ecosystem and its major abiotic and biotic components (water, soil, topography, vegetation, biomass, habitat and species) with examples of characteristics including vegetation type, water quality and soil type. (SEEA EA Glossary)

Ecosystem condition: The quality of an ecosystem unit measured in terms of its abiotic and biotic characteristics. The term 'ecosystem state' is considered to be a synonym of 'ecosystem condition' in the context of this study. (SEEA EA Glossary)

Ecosystem integrity: The ecosystem's capacity to maintain composition, structure, functioning and self-organization over time using processes and elements characteristic for its ecoregion and within a natural range of variability. (SEEA EA Glossary)

Ecosystem type (ET): Categories of a classification system (typology) for ecosystem units, reflecting a distinct set of abiotic and biotic components and their interactions. (based on SEEA EA Glossary)

Ecosystem unit: A small contiguous part of the environment that can be considered homogeneous in all of its relevant characteristics. (Simplified from the SEEA EA definitions of 'ecosystem asset' & 'basic spatial unit')

Index, sub-index: Ecosystem condition indices (and sub-indices) are composite indicators that are aggregated from the combination of individual ecosystem condition indicators recorded in the ecosystem condition indicator account. (SEEA EA Glossary)

Indicator: Ecosystem condition indicators are selected and rescaled versions of ecosystem condition variables with the aim to represent and simplify condition of an ecosystem type. Indicators are dimensionless, and they can be measured either on a continuous or a discrete scale. (SEEA EA Glossary, extended)

Pressure: A process or activity exerted by society that negatively affects the condition of ecosystems. (new)

Reference: A point of comparison for a variable that goes beyond the current data context, e.g., a theoretical threshold or a societal norm. References can provide a context-independent interpretation to the data. (new)

Reference condition (RC): The state (condition) of the ecosystem against which past, present and future ecosystem condition is compared. (based on SEEA EA Glossary)

Reference level (RL): A value of a variable against which it is meaningful to compare other measured values of the variable. Reference levels are used to rescale a condition variable to derive a condition indicator. Mathematically, a RL is a value measured on the original (untransformed) scale of the variable, which is linked ('anchored') to another concrete value, measured on the 'transformed scale' (i.e., the 0-1 continuous indicator scale). (Keith et al., 2020, extended)

Reference level role: Based on their position on the continuous indicator scale (transformed scale) several types of RLs can be distinguished. *Upper RL* (X_{100}) are linked to the 'good endpoint' (=1) of the continuous indicator scale, *lower RLs* (X_0) locate the 'bad' scale endpoint (=0), whereas *mid-range RLs* can be used to separate different status classes (e.g., 'good condition' and 'bad condition') of a discrete indicator scale, linked to a specific value of the continuous indicator scale (e.g., 0.6, X_{60}). The SELINA references framework recommends a simple notation consisting of the symbol of the variable (e.g., X) and a subscript identifying the 0-1 value to which the RL is anchored. To avoid decimal points in the subscript, this "anchoring value" is presented in % (e.g., 0.6 = 60%). Accordingly, a statement like X_{60} = 12 t/ha can be translated as "*for variable X, an original (untransformed) value of 12 t/ha will correspond to a value of 0.6 on the transformed indicator scale*". (The terminology 'RL type' is sometimes also used as a synonym for 'RL role'.) (new)

Reference site: A site representing the studied ET, that is considered to be "*in the reference condition*", and where the (upper) RL values of the different condition variables can be observed or measured. (new)

Rescaling: The act of transforming a condition variable (measured an arbitrary biophysical scale, i.e., the 'original scale') to a dimensionless 0-1 condition indicator (i.e., the 'transformed scale'), with the help of reference levels (RLs) and a rescaling function. (new)

Rescaling function: A mathematical function which takes an arbitrary real number (the value of an EC variable) and assigns it to a value on the [0,1] interval (the scale of the continuous indicator). There are several classes (functional forms) of rescaling functions, which can be adjusted with the help of parameters (RLs). (new)

Status class: A category of a discrete (ordinal, graded) scale used for assessing condition (or the ecological/environmental/conservation status) of the ecosystems, often applied by EU biodiversity policies (e.g., WFD, MSFD, HD). (new)

Typology: The operation of distributing objects into classes or groups that are less numerous than the original objects. Typologies need to be exhaustive and mutually exclusive: classes should not overlap and their union should cover all objects. Typologies can also be hierarchical if the individual divisions (classes) are further subdivided. The terms 'classification' and 'typology' are considered to be synonyms in the context of this study. (based on Parrochia, 2019)

Variable: Ecosystem condition variables are quantitative metrics describing individual characteristics of an ecosystem asset. Variables are measured on their original scales using their natural biophysical units. (SEEA EA Glossary, extended)