

Calculating the nature index

Definition, spatial units, elicitation of indicator observations, scaling functions and weights

Bård Pedersen, Stein Are Sæther, Jens Åström

Norwegian Institute for Nature Research, February ??, 2017

This document presents and explains the mathematical framework for calculating the nature index. It explains the structure of spatial units involved, format of indicator observations, scaling functions, reference values, calculation of weights, the elicitation process and parametric bootstrapping.

Definition

The nature index is a weighted average of scaled indicator observations

$$NI_{jkt} = \sum_{i=1}^n w_{ijk} S_{ikt} \quad (1)$$

where S_{ikt} , $i = 1, \dots, n$ is the n scaled indicator states included in the calculation. The weights (w_{ijk}) must conform to $\sum_{i=1}^n w_{ijk} = 1$.

Spatial units

The Nature Index is calculated for a specified major habitat (j) in a defined area (k), here referred to as an NI-area) and for a particular year (t) from a set of indicator-observations.

Indicator observations are however collected from indicator-specific spatial units (here referred to as indicator-areas) that may have a different spatial extent than the NI-area. The spatial delineation of indicator areas may also vary among indicators.

That is, for each indicator there is a set of non-overlapping spatial units from which indicator observations are collected. However, each indicator-area and NI-area must consist of one or more basic spatial units. In the Norwegian implementation of the Nature index this basic spatial unit is the municipality.

Thus, in order to calculate the index, the delineation of each indicator-area and each NI-area must be provided. That is, each indicator- and NI-area must be specified in terms of the basic spatial units they consist of. Further, all indicator observations must be linked to the correct indicator area. See section Input data below.

Indicator observations

Indicator observations may be associated with errors and are therefore uncertain. Thus, in the context of the Nature Index, the indicator observations are specified as probability distributions where the dispersion of the distribution represents this uncertainty, and the central tendency of the distribution represents the size of the indicator value. Thus, each observation is regarded as a stochastic variable. This presupposes that all input values used to calculate the index are provided with an estimate of how uncertain the values are. This estimate is provided by each expert in terms of the interquartile distance in the distribution, while the central tendency is given as an expected value.

Elicitation

Elicitation is the term for the process that generates probability distributions for uncertain quantities based on experts' knowledge and beliefs about them. In the context of the Nature Index, this is a two-step process. For all indicator values, experts estimate the distribution's expected value and its lower and upper quartiles. The R-scripts for calculating NI presupposes that these three parameters are provided for all indicator observations included in the calculation of the nature index.

During the calculation of the index, a probability distribution is fitted from a number of model distributions for each indicator value. The fitting is based on a least squares criterion. All the indicators are supposed to be nonnegative variables with zero as the minimum value. Hence, the model distributions are non-negative, univariate distributions. These are truncated normal-, lognormal-, Weibull-, "zeroinflated" exponential -, and gamma distributions. The truncated normal distribution is left-truncated at zero. All model distributions thus have two parameters.

Scaling

Scaling the indicator values to a common scale is done using nonlinear scaling functions (Figure 5.1). The scaling functions contain only one parameter, the so-called reference value (*Ref*). Reference values are specified for each indicator area in which an indicator is observed. The scaling functions' values are over the interval $[0,1]$.

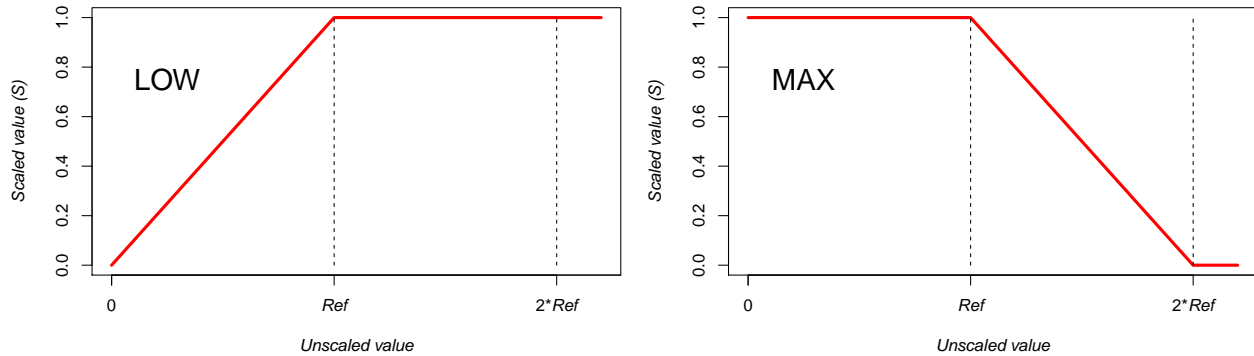


Figure 1: The scaling models *LOW* and *MAX*

There are two types of scaling functions, *LOW* and *MAX* (Figure 1). The choice of scaling function is determined by whether the indicator correlates positively or negatively with the aspect of biodiversity that the indicator represents. The *LOW* model is used when there is a positive correlation between the indicator and biodiversity. This holds true for most indicators. The *MAX* model is used when there is a negative correlation between the indicator and biodiversity. This applies only to certain indirect indicators that represent a negative effect that the measured indicator has on other components of biodiversity.

Weights

The nature index is a weighted average. In a weighted average, the state values used to calculate the average are weighted in relation to each other. Some indicator observations are given more weight than others. There are two motives for doing this in the context of the Nature Index. One motive is to correct for biases in the data from which the index is calculated. The second motive relates to the interpretation / definition of the index.

The current R-scripts calculate weights according to the Norwegian implementation of the Nature Index. However, the weighting system might be modified to better accord with the Bulgarian dataset and the

purposes of the Bulgarian implementation. The Norwegian weighting system is described below.

The weights assigned to individual indicator values for the calculation of the Norwegian Nature Index for a basic spatial unit (i.e. a municipality) take into account the indicators' specificity / fidelity to the respective major ecosystem and the indicators' ecological function. The weights can be written as a product of two factors, a trophic weight ($w^{trophic}$) and a fidelity weight ($w^{fidelity}$).

The indicators' **fidelities** (φ_{ij}) to the various major habitats are determined by the individual expert based on the degree to which the underlying data for the indicator reflect the state of one or more major ecosystems. An indicator's total fidelity is 100%, i.e. $\sum_{j=1}^p \varphi_{ij} = 100\%$. Many indicators belong to only one of the major ecosystems; for example, mussels are only an indicator for the coast bottom major ecosystem and have a fidelity of 100% to the coast bottom, with a fidelity of 0% to the other major ecosystems. Conversely, marine fish species will often belong to two or more of the marine major ecosystems, ocean pelagic and ocean bottom, coast pelagic and coast bottom. In these situations (unless ecosystem-specific data are available), the fidelity is divided over several major ecosystems based on the extent to which the data for the indicators are affected by conditions in the different systems. The effect of this is that indicators that belong to several major ecosystems will be included in the calculation of the Nature Index for all of these systems. But their weight is then reduced in each of these calculations compared with indicators that have a 100% fidelity to one system. In some cases, ecosystem-specific data is available for a species, and then separate indicators with 100% fidelity can be used. The **fidelity weight** of an indicator belonging to trophic group g is given by

$$w_{ijk}^{fidelity} = \frac{\varphi_{ij}}{\sum_{m \in g} \varphi_{mj}} \quad (2)$$

where the sum in the denominator is over all indicators within the trophic group observed in the geographical unit. The same holds for key indicators that are treated as a separate group in this context.

Trophic weight, ($w^{trophic}$), is equal to $\frac{1}{2}$ for the key indicators and equal to $\frac{1}{2r_{jk}}$ for other indicators, when observations of key indicators from municipality k is included in the calculation. When key indicators are lacking in k , the trophic weight for other indicators are $\frac{1}{r_{jk}}$. r_{jk} is the number of functional groups represented by indicator observations in the municipality k .

The product of an indicator observation's fidelity weight and trophic weight is the **total weight** given to the observation in a calculation of the nature index for a municipality k ,

$$w_{ijk}^p = w_{ijk}^{fidelity} * w_{ijk}^{trophic} \quad (3)$$

and the nature index for major habitat j in the municipality at time t is

$$NI_{jkt} = \sum_{i=1}^n w_{ijk}^p S_{ikt} \quad (4)$$

where the sum is over all indicators recorded in the municipality with $\varphi_{ij} > 0$ and $\sum_{i=1}^n w_{ijk}^p = 1$.

The consequences of these allocations are first, that the sum of the weights w_{ijk}^p for key indicators is $\frac{1}{2}$. So key indicators weigh 50% in the Nature Index for a municipality where observations of key indicators have been included in the calculation. Second, the different trophic groups represented in the municipality are weighted equally, regardless of the number of indicators represented within each group.

Municipalities are the basic spatial units for terrestrial major habitats and freshwater in the Norwegian implementation the Nature Index. However, in practice, aggregated indices for larger NI-areas (K) are calculated as weighted averages of municipal index values, e.g. for country parts or the whole country:

$$NI_{jKt} = \sum_{k \in K} w_{jk}^{area} NI_{jkt} = \sum_{k \in K} \sum_{i=1}^n w_{jk}^{area} w_{ijk}^p S_{ikt} \quad (5)$$

where w_{jk}^{area} are the **area weights**, $w_{jk}^{area} = \frac{a_{jk}}{\sum_{q \in K} a_{jq}}$, a_{jk} is the area that major ecosystem j covers in municipality k , and $\sum_{k \in K} w_{jk}^{area} = 1$.

Calculating the Nature Index

Since the Nature Index is calculated from indicator observations that are considered stochastic variables, the index itself is also a stochastic variable with an associated probability distribution. Parametric bootstrapping is used to simulate this distribution from the distributions of indicator values. An observation from the index's distribution is generated by randomly drawing one observation from each of the distributions of indicator observations used in the calculation. The Nature Index value is then calculated based on this. The index's distribution is simulated by repeating this procedure 1000 times, for example. Normally, the median of the simulated distribution is given as a point estimate for the Nature Index, while the 95% confidence interval given by the 2.5% and 97.5% quantiles of the distribution is used as a measure of uncertainty for the index value.

There are different opinions on whether uncertainties in reference values should be accounted for in the calculation of the nature index. On one hand reference values may be considered as fixed parameters that have been decided on the basis of the ecological framework described in Pedersen *et al.* (2016)¹ rather than estimated from the framework. On the other hand, reference values could be viewed as uncertain estimates derived from or modelled on the basis of the framework.

Further reading

The framework for calculating the nature index is further described and discussed in Certain *et al.* (2011)², Pedersen and Skarpaas (2012)³, Skarpaas *et al.* (2013)⁴, and Pedersen *et al.* (2016)

¹Pedersen, B., Nybø, S., Sæther, S. A. (eds.) 2016. Nature Index for Norway 2015. Ecological framework, computational methods, database and information systems – NINA Report 1226. 84 pp.

²Certain, G., Skarpaas, O., *et al.* 2011. The Nature Index: A General Framework for Synthesizing Knowledge on the State of Biodiversity. - PLoS ONE 6: e18930.

³Skarpaas, O., Certain, G. & Nybø, S. 2012. The Norwegian Nature Index – conceptual framework and methodology. - Norsk Geografisk Tidsskrift - Norwegian Journal of Geography 66: 250-256.

⁴Pedersen, B., Skarpaas, O. 2012. Statistical properties of the Nature index for Norway. Measurement error and sensitivity. – NINA Report 797. 55 pp. (In Norwegian)