Glossary

Abiotic

Non-living components of ecosystem – the chemical and physical environment.

Biotic

Living components of ecological community organisms, species.

Covariate

A variable that might predict the response of interest; also termed predictor or explanatory variable.

Extent

A component of spatial scale: the size of the study domain.

Extrapolation

Estimating the value of a response variable at unmeasured locations by extending a model to new places or times.

Geospatial data

Digital representations of data about Earth, including geographic locations, stored in a geographic information system (GIS).

Grain

A component of spatial scale: sample resolution, the size of a single observation.

Proximal

Direct or causal – for example, a factor directly related to species distribution and abundance (in contrast with distal factors that may be surrogates or proxies for causal factors).

Abundance data:

these data may be in the form of direct counts (i.e., how many individuals were observed) or some index of abundance derived from the raw counts.

Detection/non detection data:

term sometimes used to acknowledge that species presence/absence is usually imperfectly observed. More specifically, often used for data that are collected in a way that are informative about the detection process (e.g., via repeat surveys to sites, multiple independent observers, times to detection, etc.).

Integrated distribution modeling:

the practice of fitting species distribution models with more than one observation model.

Integrated population modeling (IPM):

the practice of simultaneously modeling population abundance and the demographic processes driving its variation, combining multiple sources of data into a single model (e.g., count or census-type data alongside mark–recapture and ring recoveries).

Latent state:

an unobserved, and often practically unobservable, property of the modeled ecological system (e.g., the actual distribution or abundance of a species) that we are trying to estimate.

Link function:

a function describing the relationship between the observations and the predicted mean of the latent state, to ensure that the predicted mean meets distributional criteria; for example, point counts are usually assumed to follow a Poisson or negative binomial distribution via a log link.

Multispecies model:

A statistical model in which some parameters are shared among species, often by treating species-specific parameters as random effects.

Observation model:

A statistical description of the data collection process. In a standard occupancy detection model, the observation (sub)model characterizes the likelihood of detecting the species at a site where it is present.

Occupancy-detection model:

a class of SDM where the data are collected so that they are informative about the detection process (e.g., by several repeat surveys to at least some of the sampling sites). With such data, the model can separately estimate the probability of species’ occurrence at a site, and the parameters driving the observation process, for example, the probability of detecting the species where present, or the probability that the species was present at a site where it was not observed.

Point process model:

A statistical model that describes how points (e.g., individuals) are distributed in space. The stochastic process used for this description has a so-called intensity, in which points are more likely to be present at locations where the intensity is high. The most common implementation is a Poisson point process model, which assumes independence in the location of individuals, after accounting for the intensity.

Presence–absence data:

records of whether a species is present or not at each of a number of sampling locations (e.g., quadrats or study sites). Detection does not have to be certain and its probability can be estimated with occupancy-detection models if there are multiple visits to a location. This term is often used for situations where detection/non detection data would be more appropriate.

Presence-only data: records of the locations where a species was observed (e.g., from museum samples). These data lack information about where individuals were not observed, in contrast to presence–absence data.

Species distribution model (SDM):

generally, refers to a statistical (correlative) model that relates environmental covariates to species’ records over a geographic region. In practice, SDMs are often fitted to presence only data, although are more robust when fitted to presence absence or detection/non detection data. SDMs include occupancy-detection models and abundance models.

State-space model:

A model that combines a latent state with one or more observation models that describe how the data were generated from this latent state. Structured data: data derived from a well-defined sampling protocol, such that observations are comparable in time and/or space (i.e., they can be described by a common observation model). Note that a structured survey protocol does not guarantee that the data are free from spatial bias; whilst some schemes select survey sites following a statistical sampling protocol (e.g., stratified random sampling; the UK Breeding Bird Survey), others do not (e.g., most butterfly monitoring schemes).

Submodels:

components of a hierarchical state-space model. In the integrated models described here, there are separate observation submodels for each dataset and one state submodel for the latent state.

Unstructured data:

Data collected without formal protocol or sampling design, or where the protocols are unknown. Most unstructured data are in the form of presence-only data, for example, those arising when members of the public submit records of wildlife observations.

AIC

Akaike’s information criterion, a within-sample non-Bayesian score for prediction Bayes factor the ratio of marginal data distributions pertaining to two models

BIC

Bayesian (Schwartz) information criterion, a within-sample non-Bayesian score for model averaging

CPO

conditional predictive ordinate, a within-sample score for leverage

Cross-validation

the iterative use of within-sample data to validate models in terms of out-of-sample predictive ability

DIC

deviance information criterion, a within-sample quasi-Bayesian score for prediction

Effective number of parameters

pD, a measure of model complexity as a penalty in Bayesian information criteria

Empirical Bayesian

the use of within-sample data to inform Bayesian model components such as priors

Out-of-sample data

an auxiliary set of data that are used for model comparison

Posterior predictive loss

an approach for scoring models based on decision theory

Regularization

constraining a statistical optimization problem (i.e., penalization or shrinkage)

Regulator

constraint, optimism, penalty, or prior

Score

a function used to evaluate models numerically, usually in terms of predictive ability

WAIC

Watanabe-Akaike information criterion, a within-sample fully-Bayesian score for prediction

Within-sample data

response data typically used to fit a model, but also to calculate information criteria