Model-based data integration: a primer and practical guide

To make the most accurate prediction of a species’ distribution it is important to make use of all relevant data. Although opportunistically collected species records are more readily abundant and contain lots of data, they are also more likely to be biased. Standardised surveys may suffer less observation bias, but they are expensive to conduct and are likely to be more limited in spatial extent. Combining the two data types not only gives more data to a model, but also provides a chance for structured survey data to help correct biases in opportunistic data. Model-based data integration explicitly accounts for imperfect observation processes in both datasets, and propagates information contained in each while accounting for appropriate biases. In this skills showcase we will explain the concepts behind model-based data integration, and demonstrate its implementation in a Bayesian framework for broad spatial-scale presence-only data and spatial abundance data from structured surveys. We will illustrate how Point process models can be used to facilitate this integration. We expect that participants will leave more confident in the theory and execution of integrated models.