

# Bayesian Methods for Ecological and Environmental Modelling

Combining diverse data sets

Peter Levy

2019-09-11



# This session

- ▶ No new theory
- ▶ Example applications of Bayesian approach to real problems
- ▶ Focus on
  - ▶ propagating uncertainty
  - ▶ combining different data sources

Bayesian approach particularly suited to these

## Combining diverse data sets

For any given ecological problem, we often have more than one source of information:

- ▶ old data and new data e.g. last year's experiment and this year's

## Combining diverse data sets

For any given ecological problem, we often have more than one source of information:

- ▶ old data and new data e.g. last year's experiment and this year's
- ▶ data from different methods e.g. regular surveys and citizen science

# Combining diverse data sets

For any given ecological problem, we often have more than one source of information:

- ▶ old data and new data e.g. last year's experiment and this year's
- ▶ data from different methods e.g. regular surveys and citizen science
- ▶ data from different space/time scales e.g. ground-based versus satellite

# Combining diverse data sets

For any given ecological problem, we often have more than one source of information:

- ▶ old data and new data e.g. last year's experiment and this year's
- ▶ data from different methods e.g. regular surveys and citizen science
- ▶ data from different space/time scales e.g. ground-based versus satellite
- ▶ data on different (but related) ecosystem properties e.g. evaporation and run-off, or population of species A and species B.

# Bayes Theorem

Bayes Theorem relates the parameters of a model to observed data:

$$P(\theta|D) = P(\theta) \frac{P(D|\theta)}{P(D)}$$

- ▶ posterior probability of the parameter set  $\theta$  given the observed data  $D$
- ▶ prior probability of the parameters  $P(\theta)$
- ▶ likelihood of the data given the parameters  $P(D|\theta)$
- ▶ marginal probability of observing the data  $P(D)$ .



# Bayes Theorem

Bayes Theorem relates the parameters of a model to observed data:

$$P(\theta|D) = P(\theta) \frac{P(D|\theta)}{P(D)}$$

We combine different data sources by specifying them in either the prior distribution or the likelihood function.

# How to combine data sources?

## Incorporating in the prior

- ▶ old data and new data e.g. last year's experiment and this year's

## Incorporating in the likelihood function

- ▶ data from different methods e.g. regular surveys and citizen science
- ▶ data from different space/time scales e.g. ground-based versus satellite
- ▶ data on different (but related) ecosystem properties

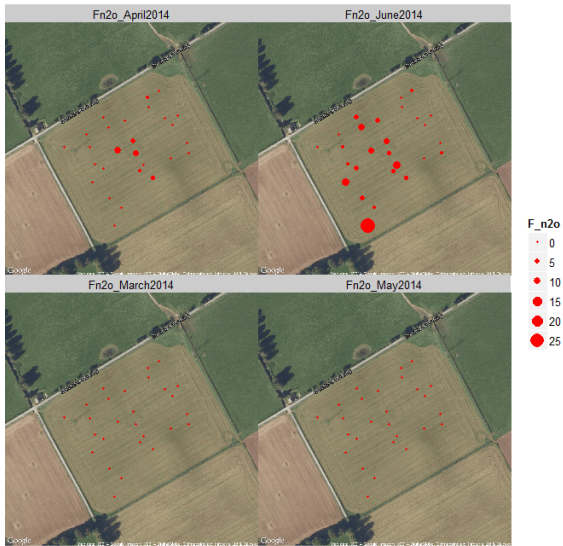
## Incorporating in the prior

$$P(\theta|D) = P(\theta) \frac{P(D|\theta)}{P(D)} \quad (1)$$

$D_{old}$  informs  $P(\theta)$

- ▶ Estimate model parameters with old data
- ▶ Obtain posterior distribution
- ▶ Now use this as subsequent prior distribution
- ▶ Re-estimate model parameters with new data
- ▶ examples
  - ▶ time series of river flow
  - ▶ N<sub>2</sub>O gas emissions from a pasture field

# Incorporating in the prior



# Data from different space/time scales

- ▶ Incorporating in the likelihood function
- ▶ Represent the effect of space/time scale explicitly
- ▶ We can consider small-scale observations as samples from a distribution in space/time.
- ▶ Large-scale observations measure the mean of this distribution.
- ▶ examples
  - ▶ satellite observations with coarse resolution and individual forest measurements
  - ▶ catchments naturally integrate over large areas, comprising a population of small sub-catchments

# Data from different methods

- ▶ Incorporating in the likelihood function
- ▶ May be very straightforward
  - ▶ just need appropriate  $\sigma$  for each observation
- ▶ May need a discrepancy term to account for systematic error in different methods
- ▶ data from different methods
  - ▶ soil moisture in the UK

## Data from different (but related) ecosystem properties

So long as the model makes a prediction of an observed quantity, or is at least correlated with it, we can include it in the likelihood function.

- ▶ example
  - ▶ land-use change in the UK - the next practical

# Practical

Brief introduction to BayesianTools R package - more this afternoon

- ▶ multiple terms for multiple data sets in the likelihood function
  - ▶ land-use change in the UK