Bayesian Methods for Ecological and Environmental Modelling

Data modelling and uncertainty propagation

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 can handle both

Data modelling and uncertainty propagation

Statistical data modelling is a mathematically-formalised representation of the process that generates your observations [1]

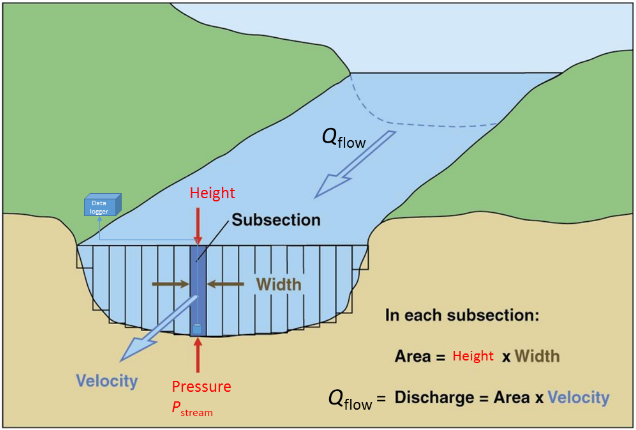
It is important in uncertainty propagation, but usually ignored.

Bayesian approach particularly suited to this

[1] "Data modelling" has other meanings in IT and database domain.

Streamflow example

We want to estimate streamflow with a sensor system shown below.



Streamflow example

- We would typically say we have “observations of streamflow”

Streamflow example

- We would typically say we have “observations of streamflow”
- But we actually have observations of stream height

Streamflow example

- We would typically say we have "observations of streamflow"
- But we actually have observations of stream height
- or rather water pressure ...

Streamflow example

- We would typically say we have “observations of streamflow”
- But we actually have observations of stream height
- or rather water pressure . . .
- or rather pressure transducer output voltage . . .

Streamflow example

- We would typically say we have "observations of streamflow"
- But we actually have observations of stream height
- or rather water pressure ...
- or rather pressure transducer output voltage ...
- or rather data logger measurements of voltage.

Streamflow example

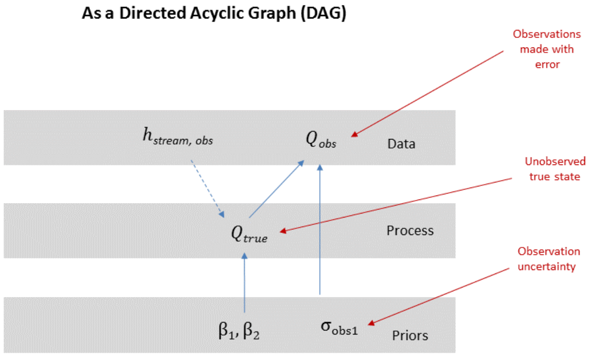
We have a series of four linear models:

$$
\begin{aligned}
Q_{flow} =& \beta_1 + \beta_2 h_{stream} + \epsilon_1 \\
h_{stream} =& \beta_3 + \beta_4 P_{sensor} + \epsilon_2 \\
P_{sensor} =& \beta_5 + \beta_6 V_{sensor} + \epsilon_3 \\
V_{sensor} =& \beta_7 + \beta_8 V_{logger} + \epsilon_4
\end{aligned}
$$

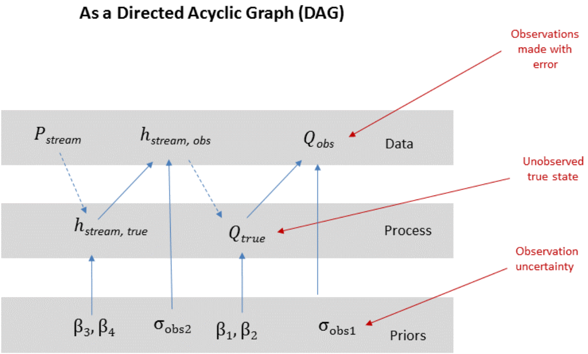
We effectively assume these models are perfect and the error terms ϵ 1-4 are zero.

This is a relatively simple case, and some of these errors may well be negligible. Many cases are not so simple.

Streamflow example as a DAG



Streamflow example as a DAG



Streamflow example

We can substitute one model in another models:

$$
\begin{aligned}
Q_{flow} =& \beta_1 + \beta_2(\beta_3 + \beta_4 P_{sensor} + \epsilon_2) + \epsilon_1 \\
P_{sensor} =& \beta_5 + \beta_6 V_{sensor} + \epsilon_3 \\
V_{sensor} =& \beta_7 + \beta_8 V_{logger} + \epsilon_4
\end{aligned}
$$

to the extreme case where we include the uncertainty in all four stages

$$
Q_{flow} = \beta_1 + \beta_2(\beta_3 + \beta_4(\beta_5 + \beta_6(\beta_7 + \beta_8 V_{logger} + \epsilon_4) + \epsilon_3) + \epsilon_2) + \epsilon_1
$$

Spot the observational uncertainties

Spot the observational uncertainties

Spot the observational uncertainties

Spot the observational uncertainties

Missing Values and Gap-filling

We often have missing values in time series.

We may need a cumulative sum, which requires us to fill the gaps.

These are commonly filled in by regression with another variable - just like calibration

Again, not a known, fixed relationship - we should include the uncertainty

When does it really matter?

When it introduces a bias that would be mis-interpreted as a real effect

When we miss a real effect because of bias in the measurements

- long-term monitoring
  - e.g. climate change detection
- long-term experiments
  - e.g. plant species composition change as a result of pollution
- when high-precision measurements are needed for cross-site comparisons
  - e.g. inverse modelling of trace gas concentrationa

Summary

We commonly assume the observations are perfectly known and error-free.

"Observations" often involve explicit or implicit modelling steps, but the model is assumed to have no uncertainty.

The Bayesian approach provides a way to make these steps explicit, and to propagate the uncertainties

- The practical exercise - https://github.com/NERC-CEH/BayesCourse/tree/master/11 Sep AM - CombiningData - PL/uncertaintyProp_Practical1.html