Grantham Institute
Climate Change and the Environment
An institute of Imperial College London

Two applications of Bayesian statistics in green energy

Jens Bendel

Mathematics of Planet Earth
Imperial College London
University of Reading

The Bayesian idea

The goal of a Bayesian statistical analysis is the *posterior distribution*. Instead of a point estimate or a confidence interval the goal is to find the full probability *distribution*. *Posterior* means "after taking experimental results into account".

This poster presents the Bayesian idea and two applications.

Parametric Bayesian Statistics

- Bayes' formula:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{1}$$

or in terms of probability density functions (PDFs):

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)} = \frac{f(x|\theta)p(\theta)}{\int_\Theta f(x|\tau)p(\tau)\,\mathrm{d}\tau} \tag{2}$$

with data / observations $x = (x_1, x_2, \ldots, x_n)$ and parameter θ.

- **A Bayesian statistical model:**

The likelihood $f(x|\theta)$ models the distribution of the data given the parameter.

The prior $p(\theta)$ contains a priori knowledge about the parameters and is usually specified by an expert.

- **The posterior** $p(\theta|x)$ is the distribution of the parameter given the data. It can be considered as an update of the prior distribution taking the available data into account.

Bayes in practice

- **The problem:**

Explicitly calculating $f(x) = \int_\Theta f(x|\tau)p(\tau)\,\mathrm{d}\tau$ is in practice often impossible.

- **The solution:**

It is "easy" to evaluate only $f(x|\theta)p(\theta)$. And it is possible to generate samples from the posterior $p(\theta|x)$ using algorithms that only require $f(x|\theta)p(\theta)$ but not $f(x)$.

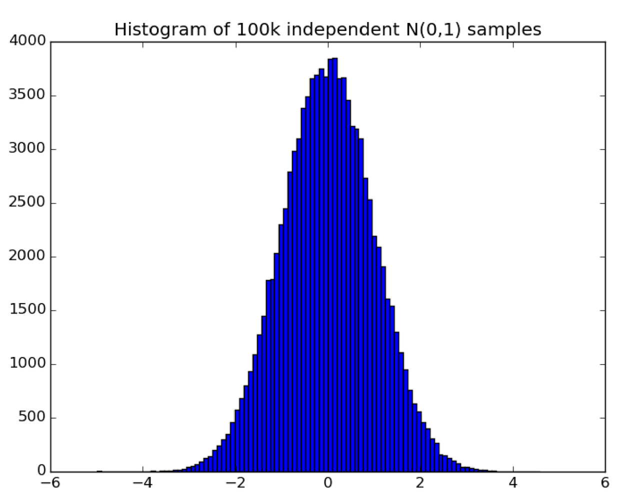

Figure : A histogram of samples from a standard Normal distribution. Having many samples is almost as good as the exact formula for the probability density function.

The Metropolis(-Hastings) algorithm is the most famous Markov chain Monte Carlo algorithm.

Algorithm 1 standard Metropolis algorithm

1. Choose a starting point θ^0 with $p\left(\theta^0 \mid x\right) > 0$.
2. For $k \in \{0, 1, 2, \ldots\}$:
   - Draw a sample θ^* from a symmetric PDF $J\left(\theta^* \mid \theta^k\right)$.
   - Calculate the acceptance ratio

$$r := \frac{p\left(\theta^* \mid x\right)}{p\left(\theta^k \mid x\right)}. \tag{3}$$

   - Accept the proposal θ^* with probability $\min(r, 1)$ i.e. draw a sample u from a standard uniform distribution and set

$$\theta^{k+1} = \begin{cases} \theta^* & \text{if } u < r; \\ \theta^k & \text{otherwise.} \end{cases}$$

To check that $p\left(\theta^0 \mid x\right) > 0$ and to calculate $r := \frac{p(\theta^* \mid x)}{p(\theta^k \mid x)} = \frac{f(x|\theta^*)p(\theta^*)}{f(x|\theta^k)p(\theta^k)}$ it is not necessary to compute $f(x)$ cf. equation (2).

1. Carbon dioxide storage

- *Storage of carbon dioxide:*

To reduce the amount of greenhouse gases in the Earth's atmosphere it is possible to **capture carbon dioxide (CO2) and store it**. Motivated by the idea of long term storage of CO2 in deep aquifers McBride-Wright et al. (2015) **investigate properties of solutions of CO2 in water**.

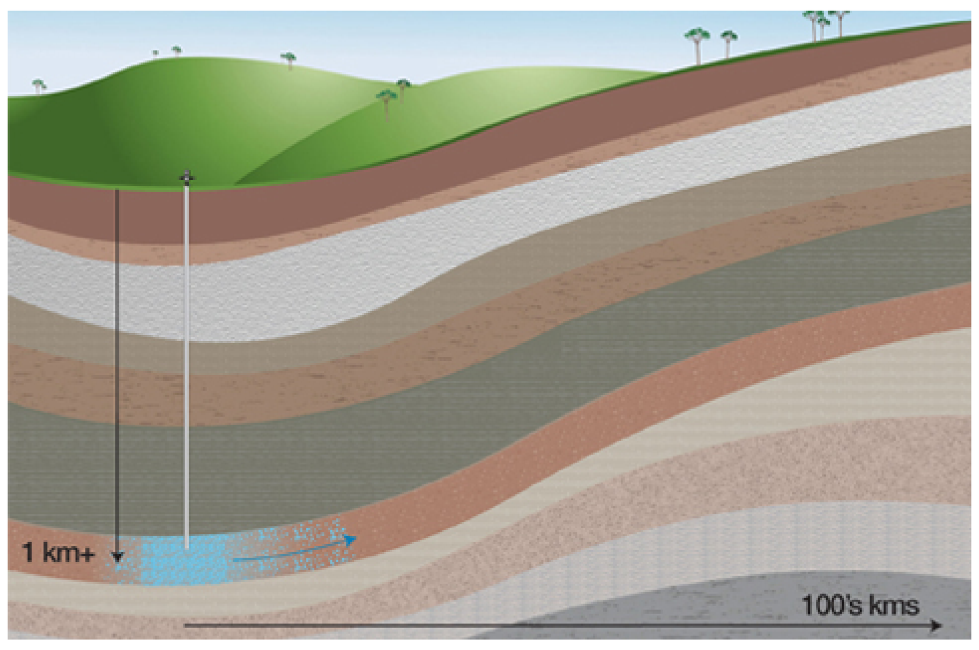

Figure : Deep aquifers can store carbon dioxide over geological time frames. (Image: CO2CRC.)

- **Experiment (and problem):**

The *quantity of interest* is the viscosity η. McBride-Wright et al. (2015) use a vibrating wire viscometer to determine η. The higher the viscosity, the slower the wire vibrates. However, the quantity that actually *can be directly measured* is a voltage V.

How do we use the measurements of V to draw inference about the viscosity η?

1.1 The mathematical model

- **Relationship between data and parameter:**

The following equations describe how the available data of V relates to the quantity of interest η:

$$V = \frac{\Lambda if}{f_0^2 - f^2(1+\beta) + if^2(\beta' + 2\Delta_0)} + a + bi + cif, \tag{4}$$

where β' and β denote the real and imaginary part of

$$\beta' + i\beta = \frac{\rho}{\rho_s}\left[i + \frac{4iK_1(\sqrt{i\Omega})}{\sqrt{i\Omega}K_0(\sqrt{i\Omega})}\right] \tag{5}$$

with $K_m(\cdot), m \in \{0, 1\}$, denoting the modified Bessel function of second kind with order m and

$$\Omega = \frac{2pf\rho R^2}{\eta}. \tag{6}$$

- **The likelihood:**

Our data x consists of experimental measurements $V_1, \ldots, V_n$ of the voltage V. The viscosity η is the parameter θ. We can use equations (4–6) to construct our statistical model in terms of the likelihood $f(x|\theta)$, i.e. in this case $f(V_1, \ldots, V_n|\eta)$.

- **The prior:**

A priori knowledge about the viscosity η is captured in the prior $p(\theta)$, i.e. $p(\eta)$.

1.2 First results

A Markov chain Monte Carlo method was custom built to obtain samples from the posterior distribution $p(\eta|V_1, \ldots, V_n)$.

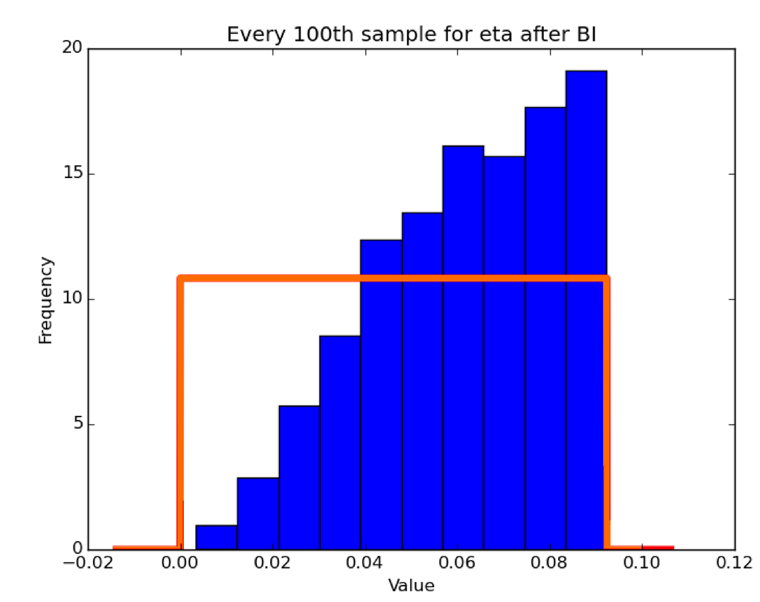

Figure : A plot of the prior distribution and a histogram of 2000 samples from the posterior distribution (blue). The posterior distribution is an updated version of the prior that takes the measured data into account.

2. Construction of offshore wind turbines

The planning and construction of offshore wind farms must account for the condition of the soil on which the wind turbines will be built.

Figure : An offshore wind farm west of Denmark. (Image: NOAA.)

Of particular interest are the number as well as the positions of boundary layers between different types of soil.

2.1 Identifying boundaries between soil layers

The usual procedure to find such boundary layers is the examination of core samples.

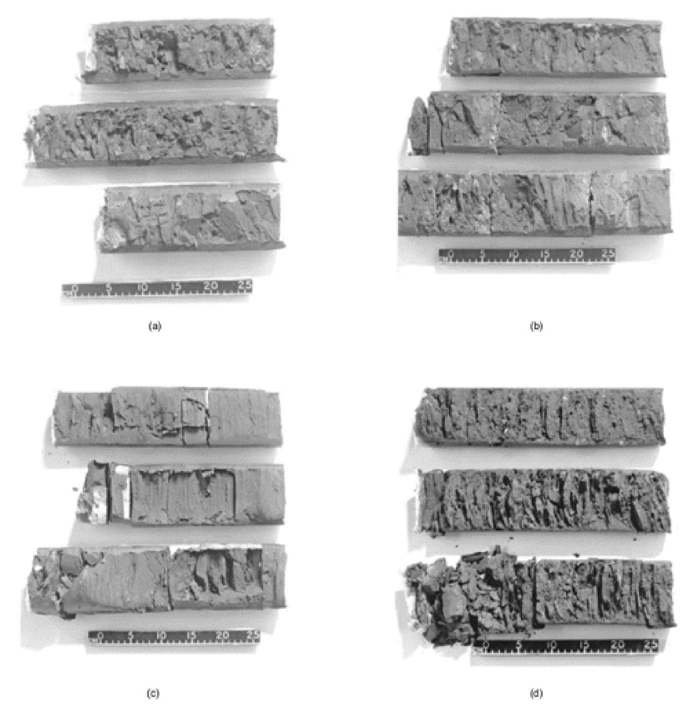

Figure : Core samples. (Image: Standing & Burland (2006).)

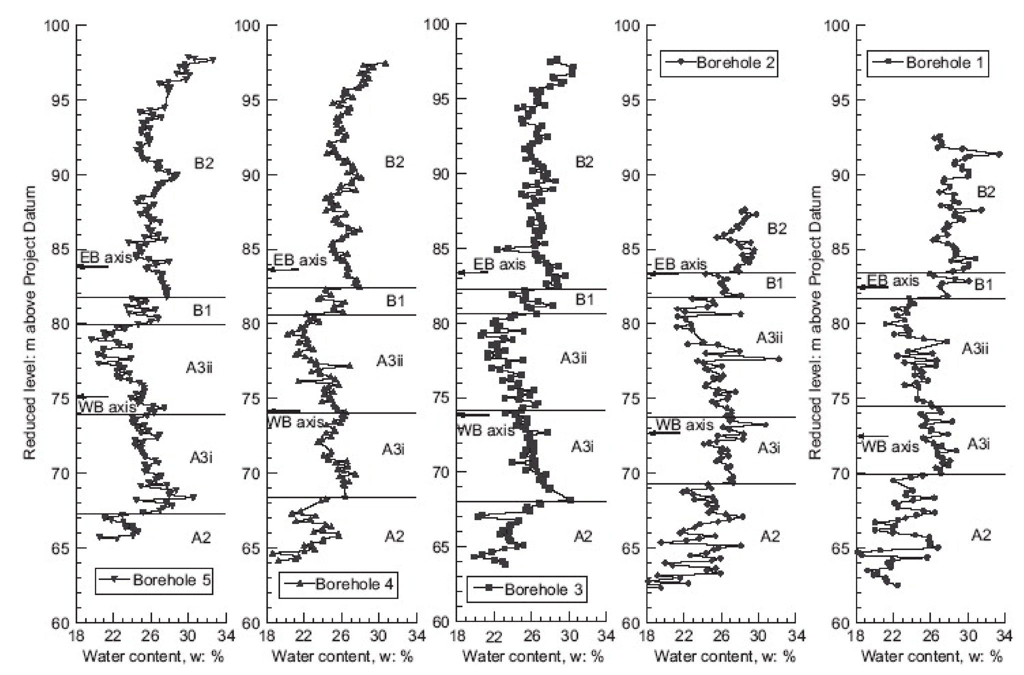

Figure : Water content of soil samples at different depths. It is not obvious where there is a change indicating a boundary layer or not. (Image: Standing and Burland (2006).)

Using sophisticated Bayesian methodology we can estimate the number of boundary layers and their location.

Acknowledgements

This poster presents joint work with Dr. Ben Calderhead (Dept. of Mathematics) and Prof. Martin Trusler (Dept. of Chemical Engineering).

EPSRC
Engineering and Physical Sciences Research Council

References

M. McBride-Wright, G. C. Maitland, and J. P. M. Trusler. Viscosity and density of aqueous solutions of carbon dioxide at temperatures from (274 to 449) K and at pressures up to 100 MPa. *Journal of Chemical & Engineering Data*, 60:171-180, 2015.

J. R. Standing and J. B. Burland. Unexpected tunnelling volume losses in the Westminster area, London. *Géotechnique*, 56(1):11-26, 2006.

j.bendel14@imperial.ac.uk