University of Edinburgh
School of Mathematics

Bayesian Data Analysis

Assessed Problem 1 Due 15 March, 11pm **Semester 2, 2016–2017**

Soil water characteristic curves (SWCC).

Description

Soil water retention data are useful for constructing soil water characteristic curves, a graphical representation which describes how soil stores and releases water. Usually, for each collected sample, the data is obtained experimentally after applying different tensions levels, x, and observing y, the proportion of water (soil water content) in the soil sample. Tension levels are measured in units of atmosphere (atm) pressure.

The observed proportion of water is limited from below by the residual soil water content and limited from above by the saturated proportion of water. The residual soil water content, $y_{\min}$, is the amount of water that cannot be drained from the soil even at high tension values and it is defined as the soil water content at $x = 15atm$. The saturated soil water content, $y_{\max}$, is the maximum water content able to be held in the soil before drainage takes place and it is defined as the water content at $0atm$. These two characteristics of the soil vary among soil samples and are different across soil depths. Therefore, for each sample, the observed data measured over several tension levels is rescaled using the residual water content value $y_{\min}$ and the saturated water content value $y_{\max}$ measured for this sample, so that the observed proportions of water are between 0 and 1:

$$\tilde{y} = (y - y_{\min})/(y_{\max} - y_{\min}).$$

Nonlinear models for conditional mean

Several analytical nonlinear expressions have been proposed in the literature for representing the soil water content $E(\tilde{y} \mid x) = \eta\,(x, \beta)$. Among the most widely used SWCCs expressions are Gardner's and van Genuchten's.

Gardner's expression is given by

$$\eta\,(x, \beta) = \frac{1}{1 + (\beta_1 x)^{\beta_2}}, \tag{1}$$

where $\beta_1 > 0$, $\beta_2 > 0$, and $\beta = (\beta_1, \beta_2)$.

The van Genuchten expression is given by

$$\eta\,(x, \beta) = \frac{1}{\left[1 + (\beta_1 x)^{\beta_2}\right]^{\beta_3}}, \tag{2}$$

where $\beta_1 > 0$, $\beta_2 > 0$, $\beta_3 > 0$, and $\beta = (\beta_1, \beta_2, \beta_3)$.

Data collection

Soil water retention data is usually measured by collecting soil samples along a given region and, at each collection site, soil samples are obtained for different soil depths and the measurements on the response variable are made in replicates. The reason why soil samples are collected from different depths at the same collection site is because soil water retention at different layers of the same soil is subjected to different hydraulic properties and dynamics. Thus, there is evidence of a possible variability in the data due to an unobserved effect of soil depth.

The data set consists of soil samples collected at depths $0-5cm$, $15-20cm$ and $60-65cm$, with 3 replicate samples at each depth. For each sample, the water content was measured at eight tension levels ranging from 0.01 to 10 atm, $x = (0.01, 0.03, 0.06, 0.10, 0.33, 0.80, 4, 10)$. For each sample, the data is rescaled using the residual soil water content and limited from above by the saturated proportion of water for that sample (i.e. the observations at $x = 0$ and at $x = 15$) as described in Section "Description". In particular, this implies that the rescaled observations $\tilde{y}_{kji}$ are between 0 and 1, $k = 1, \ldots, 8$, $j = 1, 2, 3$, $i = 1, 2, 3$.

The data is in the file DataSWCC.odc. The depths are coded as 1 for $0-5cm$, as 2 for $15-20cm$, and as 3 for $60-65cm$.

Aims

The main question of interest is the correct choice of the non-linear model for the conditional mean.

Also, accurate estimation of model parameters and their uncertainty is a major concern in SWCC studies. In both models, parameter β_1 is related to the inverse air entry value of the soil, and parameter β_2 characterises how fast the water content in the soil changes with the change in tension level. In the second model, β_3 is additional parameter adding flexibility to the model.

1. Propose a likelihood for the given rescaled water content data (conditional distribution of the rescaled water content $\tilde{y}$ given its mean), and find expressions for the mean and the variance of observations under the proposed likelihood. State the assumptions you have made.

2. Propose non-informative or weakly informative prior distributions for the unknown parameters for each of the two models (Gardner's and van Genuchten's) for the conditional mean of the rescaled water content $\tilde{y}$ (justify your choice).

3. Discuss how possible variability due to an unobserved effect of soil depth can be taken into the account for the given data, and include it in the model.

4. Fit the corresponding Bayesian model for each of the two models for the conditional mean (demonstrating convergence). Give the posterior summaries for the unknown parameters. Check model fit.

5. For each of the two models, obtain the predictive distribution of the rescaled water content of soil at depth $0-5cm$ at tension level 1atm (give mean, median, standard deviation, 95% confidence interval, and plot its density).

6. Check sensitivity of the two models to the priors (include plots to justify your conclusions).

7. Compare the two models for the conditional mean, and discuss which model fits best.

8. Discuss how the considered model can be improved using only given information and given data.