

Gibbs Sampler for the cBN Model

The network model can be expressed in probability distribution:

$$\begin{aligned}
 \log F_\mu &\sim N(\mu_F, \sigma_F^2) \\
 \mu_F &= \beta_{0F} + \beta_{1F}DA + \beta_{2F}Precip \\
 ml_{20} &\sim N(\mu_{ml20}, \sigma_{ml20}^2) \\
 \mu_{ml20} &= \beta_{0ml} + \beta_{1ml}P_{Natural} + \beta_{2ml}sgc + \beta_{3ml}Precip \\
 RichTOL &\sim N(\mu_R, \sigma_R^2) \\
 \mu_R &= \beta_{0R} + \beta_{1R}Temp + \beta_{2R}\mu_F + \beta_{3R}\mu_{ml20}
 \end{aligned}$$

These regression models are simplified representations of the model used. They do not include the interactions between some predictors and eco-region. The full model has 23 parameters to be estimated (Table 1 in the paper). Instead of using a Markov chain Monte Carlo simulation approach as in Qian and Miltner (2015) that is computationally demanding, we developed a Gibbs sampler algorithm based on linear model theories. It greatly reduces computational intensity and facilitates the predictive Monte Carlo simulation.

The network that connects these models forms a full set of conditional probability distributions: the joint posterior distribution of coefficients $\beta_F = \{\beta_{0F}, \beta_{1F}, \beta_{2F}\}$ is a multivariate normal distribution conditional on the data ($D_1 = \{\log F_\mu, DA, Precip, P_{Natural}\}$). So is the joint posterior distribution of $\beta_{ml20} = \{\beta_{0ml}, \beta_{1ml}, \beta_{2ml}, \beta_{3ml}\}$. These models are not affected by other variables in the network and we can directly draw posterior random samples based on standard linear model results (Qian 2016). For each set of these model coefficients, we obtain posterior samples of μ_F, μ_{ml20} for each observation in the data. Combining with temperature data, we have the conditional posterior distribution of $\beta_R = \{\beta_{0R}, \beta_{1R}, \beta_{2R}, \beta_{3R}, \beta_{4R}\}$, which is also a multivariate normal distribution. As a result, parameters of these models can be estimated using the Gibbs sampler. Specifically, the two models of flow metrics are independent from the model of *RichTOL*. The conditional distribution of coefficients of these two models are directly defined by the linear regression results. Conditional on $\mathbf{B} = \{\beta_F, \beta_{ml20}\}$ (i.e., μ_F, μ_{ml20} are known), the posterior conditional distribution of β_R (given data $D_2 = \{Temp, P_{Natural}, RichTOL\}$) is a multivariate normal with mean $\hat{\mu}_R = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T RichTOL$ and $\mathbf{X} = (1, Temp, P_{Natural}, \mu_F, \mu_{ml20})$ is the design matrix for the regression model. The posterior distribution of $\pi(\beta_R | D_1, D_2) = \int_{\mathbf{B}} \pi(\beta_R | \mathbf{B}, D_2) \pi(\mathbf{B} | D_1) d\mathbf{B}$.

As the posterior distributions of $\mathbf{B} | D_1$ is independent of β_R , the Gibbs sampler for this example model is simplified to a Monte Carlo simulation. Specifically, the Monte Carlo simulation starts with drawing random samples of $\sigma^2 = \{\sigma_F^2, \sigma_{ml20}^2, \sigma_{dh20}^2\}$ and \mathbf{B} as outlined in Qian (2016, Chapter 9), followed by random samples of σ_R^2 and β_R from the conditional distribution $\pi(\beta_R | \mathbf{B}, D_1, D_2)$ using the same algorithm. Repeating the process many times, we have samples of \mathbf{B} and β_R from their joint posterior distribution. These random samples represent the posterior distribution of all model coefficients.