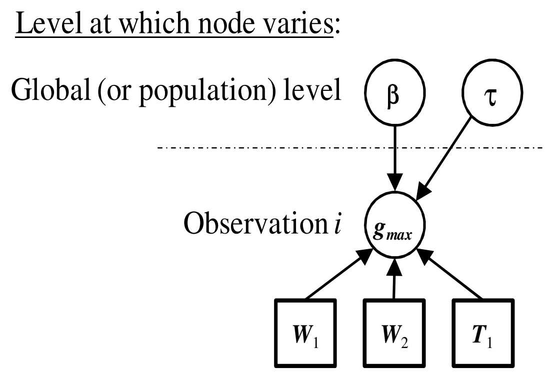
STOMATA EXAMPLE

In this project, you will gain experience implementing a Bayesian regression model and evaluating model fit. You are provided with data on maximum daily stomatal conductance (g_{max}) of a desert shrub. Plants are expected to regulate g_{max} such that they tend to close their stomata (reduce g_{max}) under water stress conditions (e.g., under low soil water content), and g_{max} may also be affected by air or leaf temperature. The data you are given include measurements of g_{max}, and associated covariates including “shallow” (0-30 cm soil depths) soil water (W_1), “deep” (30-60 cm) soil water (W_2), and nighttime temperature (T) preceding the daytime g_{max} measurement. Thus, for observation i (i = 1, 2, …106), you have $g_{max,i}$, $W_{1,i}$, $W_{2,i}$, and T_i. Based on plotting the data (which you might want to do yourself), you should log-transform g_{max}, which you will assume is normally distributed about a mean that is expressed as a regression model. You will conduct a linear regression of log(g_{max}) on *covariate-centered* values of W_1, W_2, and T (call these Wc_1, Wc_2, and Tc), and you should include a quadratic effect for temperature (i.e., Tc^2) and all 2-way interactions involving Wc_1, Wc_2, and Tc (i.e., three 2-way interactions) and Wc_1, Wc_2, and Tc^2 (an additional two 2-way interactions since Wc_1 x Wc_2 will already have been included). Thus, the vector of regression coefficients (β) should contain 10 elements (intercept, 3 main effects, 1 quadratic effect for Tc, and 5 2-way interactions). The DAG for this model is given below (note: τ, the precision is a scalar). The g_{max} data are in units of moles m^{-2} sec^{-1}, W_1 and W_2 have units of cm^3 water / cm^3 soil, and T has unit of Celsius (^{o}C).



The complete statistical model is as follows; for observation i and y = log(g_{max}):

Likelihood: $y_i \sim Normal(\mu_i, \sigma^2)$

Mean model: $\mu_i = \beta_1 + \beta_2 Wc_{1i} + \beta_3 Wc_{2i} + \beta_4 Tc_i + \beta_5 Tc_i^2 + \beta_6 Wc_{1i} Wc_{2i} +$
$\beta_7 Wc_{1i} Tc_i + \beta_8 Wc_{2i} Tc_i + \beta_9 Wc_{1i} Tc_i^2 + \beta_{10} Wc_{2i} Tc_i^2$

Where the centered covariates are defined as:

$Wc_{1i} = W_{1i} - \overline{W}_1,\ Wc_{2i} = W_{2i} - \overline{W}_2,\ Tc_i = T_i - \overline{T}$

where $\overline{W}_1, \overline{W}_2$, and $\overline{T}$ are the corresponding sample means

Priors: $\beta_k \sim Normal(0, 100000)$ conjugate, relatively non-informative
k = 1, 2, …, 10 (large variance [small precision])

$\tau \sim Gamma(0.1, 0.1)$ conjugate, relatively non-informative for precision ($\tau = 1/\sigma^2$)

Activities:

1. **1.** Program the model described above in JAGS by completing the "starter code" (Stomata_model_starter.R); if you get stuck, you can refer to the "solution code" (Stomata_model_SOLN.R). As part of this, include code for computing the standard deviation (σ) from the precision (τ), and for computing the covariate-centered covariates, Wc_1, Wc_2, and Tc (e.g., in R prior to providing the data to JAGS).
2. **2.** Since we are including a quadratic temperature effect (i.e., Tc^2), it is of interest to estimate the optimal temperature (T_{opt}, ºC); i.e., T_{opt} is defined as the temperature that "yields" the maximum value of g_{max} that would be predicted at average values of W_1 and W_2 (i.e., at $Wc_1 = 0$ and $Wc_2 = 0$). Solve for T_{opt} (this should be a function of your regression coefficients, β), and include code in your JAGS model to compute T_{opt} so you can obtain its posterior statistics. NOTE: You will implement the regression model with *centered* temperature data, and you will need to compute T_{opt} on the "raw" / "non-centered" scale.
3. **3.** Specify starting values (initials) for 3 chains, for all root nodes. Initialize the JAGS model via jags.model, and use coda.samples to update the jags model and to monitor all parameters of interest (β, τ, σ, T_{opt}, deviance). Evaluate burn-in, mixing, convergence, and autocorrelation.
4. **4.** How many total iterations do you need to run to computer posterior statistics?
5. **5.** Produce posterior results from your above model, including posterior means and 95% CIs for the parameters of interest (i.e., β, τ, σ, and T_{opt}).
6. **6.** How would you evaluate model fit?
7. **7.** Interpret your posterior results with respect to the following.
   - **a.** How well did the model fit your data? Is the model biased in any way?
   - **b.** Which covariates seem to be important predictors of g_{max}? For these "important" covariates, how do they affect g_{max}? E.g., do your results support the hypothesis that water stress (lower soil water content) stimulates stomatal closure (reduces g_{max})? Is the response to soil water affected by nighttime temperature conditions?
   - **c.** How does nighttime temperature affect g_{max}, and what is the optimal temperature for g_{max}?