The problem: Specific leaf area (SLA) data were mined from the primary literature for species of pines that are found in the United States. For a particular study, observations of SLA are obtained by measuring the surface area of pine needles and dividing this area by the dry weight of the leaves, thus SLA has units of cm^2/g. From each paper (or study), we extracted the SLA value reported in the article, which is usually a sample mean, and these values are the data that you will work with in this analysis. One potentially complicating factor when analyzing SLA data from different studies is that some studies report SLA as the "1-sided leaf area" (i.e., projected area of the needles) whereas others use the "2-sided leaf area" (i.e., total surface area of the needles), and some studies do not report "sidedness." Thus, for studies that used "2-sided" leaf area, their SLA values should be about 2 times higher than studies that used "1-sided" leaf area. The goal of this analysis is to obtain estimates of the mean (or "true") 1-sided (or projected) SLA value for each species. We would like to compare these SLA estimates across species to determine which species appear to have "unusually" low or high SLA values. We are interested in SLA because it is an important plant trait that affects a tree's ability to capture light, acquire carbon, and grow, and it is often a key parameter in vegetation or ecosystem models.

The data: We have a total of 396 SLA values (**sla**) for 25 species of pine (Table 1). Again, the SLA data are reported sample means. For some studies, we also have (1) the sample size (n) associated with the SLA sample mean (the data file contains n-2 = **dfMinus1**, i.e., "degrees of freedom minus 1"), (2) the standard error of the mean SLA reported (the data file contains **se2** = the squared standard error), and (3) the sidedness of the leaf area measurement (**Side**, coded as 0 = 1-sided, 1 = 2-sided), which varies at the level of the study. We have the species identity (**spp**) and source or study index (**src**) associated with each reported SLA value. It is possible that latent or species-specific SLA's are more similar for more closely related species, thus, for each species, we are also given the subsection that it occurs in (**subsec**), which we will use in a hierarchical "taxonomy" model for the species-specific SLA's.

The model: We will begin with a simple model that "ignores" the sidedness, sample size, standard error, and subsection data. E.g., assume that the reported SLA value for record i comes from a normal distribution with a species-specific mean and a common variance (or precision) such that:

$$sla_i \sim Normal\left(\mu_{sp(i)}, \sigma^2\right)$$

Where the $sp(i)$ indicates species (sp) associated with record i. We will define a hierarchical parameter model (or prior) for the species-specific latent SLA values (μ) that assumes that these species are a "random sample" from a population of pine species such that:

$$\mu_{sp} \sim N\left(\hat{\mu}, \hat{\sigma}^2\right)$$

We will assign standard, relatively non-informative priors to the remaining global parameters (or "root nodes"; i.e., $\hat{\mu}$, $\hat{\tau}$, and τ, where the τ's are precisions such that $\tau = 1/\sigma^2$). Then we will move on to implementing more complicated models that incorporate (1) the sidedness data and estimate the missing sidedness values, (2) the subsection data, which will be used to revise the hierarchical parameter model for μ, and (3) the sample size and standard error data for more accurate estimation of the species-specific SLA values, while simultaneously estimating the missing sample size and standard error data.

Table 1. List of species and the number of SLA records and the subsection index for each species.

Spp ID	Species	# records	Subsection ID	Spp ID	Species	# records	Subsection ID
1	*Pinus albicaulis*	1	7	14	*Pinus nigra*	2	5
2	*Pinus aristata*	1	4	15	*Pinus palustris*	9	1
3	*Pinus banksiana*	40	3	16	*Pinus ponderosa*	65	6
4	*Pinus contorta*	20	3	17	*Pinus radiata*	14	1
5	*Pinus coulteri*	1	6	18	*Pinus resinosa*	56	5
6	*Pinus echinata*	2	1	19	*Pinus rigida*	3	1
7	*Pinus edulis*	6	2	20	*Pinus serotina*	2	1
8	*Pinus elliottii*	23	1	21	*Pinus strobus*	19	7
9	*Pinus flexilis*	2	7	22	*Pinus sylvestris*	69	5
10	*Pinus jeffreyi*	1	6	23	*Pinus taeda*	48	1
11	*Pinus lambertiana*	1	7	24	*Pinus torreyana*	1	6
12	*Pinus monticola*	6	7	25	*Pinus virginiana*	3	3
13	*Pinus muricata*	1	1				

Activities:

1. Implement the above model in JAGS and R. For coda.samples, monitor all parameters of interest. Evaluate behavior of the MCMC chains (e.g., mixing, convergence, autocorrelation, and number of samples required).

2. Based on the Raftery diagnostic, ensure that you've run coda.samples for a sufficient number of iterations. Compute posterior statistics.

3. Compare species-specific SLA estimates (e.g., a caterpillar plot may be useful). Which species have the lowest and which have the highest SLA? Which species are associated with the lowest and highest uncertainty?

4. Modification to try: Model the species-level SLA values (μ_{sp}) hierarchically around their corresponding subsection SLA value, and model subsection-level latent SLA hierarchically around an overall, population-level SLA. This will introduce (unknown) variance terms that describe variability among species with each subsection, and among subsections. Code, implement, run this model, as done for the first model. Compare posterior results among the two different versions.

5. Modification to try: Account for sidedness with the following modifications:

$$sla_i \sim Normal\left(\eta_i, \sigma^2\right)$$

$$\eta_i = \mu_{sp(i)}\left(Side_{src(i)} + 1\right)$$

Where μ_{sp} is modeled that same as previously, but it is now interpreted as the latent ("true") one-side, species-specific SLA. Multiplication by $Side + 1$ converts the predicted SLA to 2-sides if Side = 1; $src(i)$ indicates the source associated with record i. Note that there are many missing values for *Side*, so we need a stochastic model (likelihood) for the

Side data, which provides a mechanism to predict/estimate the missing values. As a binary (0, 1) variable, we will assume:

$$Side_{src} \sim bern(p)$$

Where p is the probability that a study (source) reported or measured 2-sided SLA. Assign a relatively non-informative, *Uniform*(0,1) prior to p. Code, implement, run this model, as done for the first model. Compare posterior results among the different model versions.

6. <u>Modification to try</u>: Following Ogle et al. (2013), try to incorporate the sample size and standard error information in an attempt to account for the uncertainty associated with reported SLA value. In particular, Ogle et al suggest:

$$sla_i \sim Normal\left(\eta_i, \frac{\sigma^2}{N_i}\right)$$

$$N_i - 2 = \mathrm{dfMinus1}_i \sim Poisson(\lambda)$$

$$se2_i \sim Gamma\left(\frac{N_i - 1}{2}, \frac{N_i - 1}{2} \frac{N_i}{\sigma^2}\right)$$

Assign "appropriate" and relatively non-informative priors to the root node parameters, τ ($1/\sigma^2$) and λ; η_i is the same as define in part 5, or it can be replaces with $\mu_{sp(i)}$ if one wishes to "ignore" the sidedness data at this point. Code, implement, run this model, as done for the first model. Compare posterior results obtained from this model compared to other model versions.

References:

Ogle, K., J.J. Barber, K. Sartor (2013). Feedback and modularization in a Bayesian meta-analysis of tree traits affecting forest dynamics. *Bayesian Analysis* 8:133-168.

Ogle, K., Pathikonda, S., K. Sartor, J.W. Lichstein, J. Osnas, S.W. Pacala (2014). A model-based meta-analysis for estimating species-specific wood density and identifying potential sources of variation. *Journal of Ecology* DOI: 10.1111/1365-2745.12178