**Spatiotemporal Models for Ecologists**

**Homework #2 – Generalized linear mixed models in Template Model Builder**

Goal: Practice and demonstrate ability to (1) estimate parameters for generalized linear mixed models in Template Model Builder, (2) use a simulation experiment to demonstrate that correctly-specified models are statistically consistent, and (3) explore impact of mis-specifying a model on variance estimation.

Files to turn in:

1. Please submit a written description of your results. The whole thing should be (I imagine) less than 3 pages.
2. Please also submit a single R script, and a single TMB Template file provided code that can replicate the analysis.

**Simulation experiment**

Imagine you’ve collected data involving 10 counts of a species (e.g., clams) at each of 10 sites, where sites have some biological difference in habitat suitability. Please simulate data using the following design:

where is the log-mean for expected counts, is the log-mean for site , is the expected count for observation , is the variance of “among-site variability”, and is the variance of overdispersion. Note that we could also re-parameterize the model…

where is the site-level deviations away from (the global log-mean), and is deviations for each observation away from the prediction for that site (which we call “overdispersion”). If you are not used to these two parameterizations, we encourage you to confirm that they are identical (using algebra or a numerical example).

Then, try fitting this model using four alternative estimation models:

1. a generalized linear model, without any among-site variability or overdispersion.
2. a generalized linear mixed model (GLMM) with only among-site variability
3. a GLMM with only overdispersion
4. a GLMM with both among-site variability and overdispersion

Now, generate 100 simulated data sets from this model, and for each data set fit the model using each of the four estimation models. For each estimation model, record the estimate of and its standard error. Then, for the 100 simulated data sets, record the proportion of replicates where the true value ( is within the estimated confidence interval. Please provide a table summarizing the average estimate of and the proportion of replicates where the true value is within the interval for each estimation model.

Questions:

1. What do you notice about confidence interval coverage when the model is mis-specified?
2. Why might this pattern arise?