

Spatiotemporal Models for Ecologists

Homework #1 – Generalized linear models in Template Model Builder

Goal: Practice and demonstrate ability to (1) estimate parameters for generalized linear models in Template Model Builder, (2) use a simulation experiment to demonstrate that correctly-specified models are statistically consistent, and (3) use cross-validation to evaluate model performance in using a real-world data set.

Files to turn in:

1. Please submit a written description of your results, involving at least 2 tables (described below) and some explanatory text for each. 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.

Part 1 – Case study demonstration

First, please load survey catch rate data for Alaska Pollock in the eastern Bering Sea:

```
devtools::install_github("nwfs-assess/geostatistical_delta-GLMM")
data(EBS_pollock_data, package="SpatialDeltaGLMM")
```

This data set contains the catch rate (labelled “catch”) as well as some potential covariates.

Then, write a template file in TMB that can be used to estimate parameters for three alternative generalized linear models. Please use a single TMB template (i.e., a single “.cpp” file), where the file can be used to run all three models depending upon a user-defined input. Please use the following three models:

1. A generalized linear model (GLM) with a delta-lognormal distribution (see Week 1 lab) using design matrix $X=1$ and a log-link (i.e., only a single intercept)
2. A GLM with a delta-gamma distribution only using a single intercept and a log-link
3. A GLM with a distribution of your choice, potentially involving covariates, but which in some way differs from #1 and #2

Please provide a results table listing the following three results for each model:

1. The log-likelihood of the data;
2. The number of parameters
3. The log-predictive score per datum from a 10-fold crossvalidation experiment (see Week 1 lab powerpoint for definition, and Week 1 lab R script for example)

Part 2 – Simulation experiment

Please conduct a simulation experiment that involves a 3 x 3 factorial design involving all nine combinations of three simulation models and three estimation models. Use the same three estimation models as in Part 1, and define the three simulation models to perfectly match the estimation models (i.e, having the same linear predictors, probability distribution for the response variable, etc.). Parameters for the simulation models should be fixed at the values estimated when

using the EBS Pollock data set. For each combination of simulation and estimation model, please conduct 100 simulation replicates. Each replicate should involve generating 12,210 observations (i.e, the same number as in the pollock data set).

Please use this simulation experiment to show that the three cases where the simulation and estimation model are matching provide unbiased estimates of the intercept parameter. Please also interpret which estimation model has relatively high or low error for each simulation model.