**Bayesian Reanalysis of The Effects of Wetland Plant Functional Group Diversity on Ecosystem Services**

Lindsey Pett

PBIO 294

**Introduction**

Wetlands provide a multitude of ecosystem services, including but not limited to nutrient storage and production of primary productivity. The research being discussed in this paper is a Bayesian reanalysis of a past master’s thesis in which a stratified mesocosm experiment was used to determine the effects of wetland plant functional group diversity on ecosystem services of storage of carbon and nitrogen and production of aboveground biomass. Data was collected for two consecutive years, 2015 and 2016.

In 2015 mesocosms with the most plant functional groups stored the most carbon, with an average storage capacity of 433 g/m², although not statistically significant. Nitrogen storage for 2015 was also found to be the highest in mesocosms with the highest level of functional group diversity present, with an average storage capacity of 8.70 g/m², also not significant findings. Stored carbon and nitrogen positively increased as functional group diversity increased. Aboveground biomass for 2015 was also highest when functional group diversity was highest (three functional groups present), with an average biomass of 339 g/m².

2016 data results followed a similar trend to 2015 data results. 2016 had marginally the greatest amount of stored carbon when functional group diversity was highest, with an average storage of 310.61 g/m². This however did not follow a linear pattern, mesocosms with the lowest amount of functional groups present (one functional group) had the second highest amount of average biomass for 2016. 2016 also had the greatest amount of stored nitrogen when diversity was highest, with an average storage of 6.53 g/m². Finally, aboveground biomass was the greatest when high diversity of functional groups were present with an average biomass of 349 g/m², following a linear trend.

From the above-mentioned data, a Bayesian approach of inferring probability based on prior data was used to reanalyze the effects of wetland plant functional group diversity on ecosystem services. Bayesian analysis will allocate credibility based on the prior probabilities of 2015 data on the 2016 data. The prior probability based on 2015 observations allow us to see how allocation of probability differs from what the actual 2016 findings were. Comparing the Bayesian analysis of 2016 data against the MLEs of 2015 and 2016 findings can be beneficial in seeing how well your data fit the predicted, and for future predictions in modeling data.

Bayesian analysis of the above-mentioned data occurred using Markov Chain Monte Carlo (MCMC) samplers. A Metropolis and Gibbs sampler showed the differences between actual maximum likelihood estimates and the maximum likelihood estimate with a prior from the first year of sampling. Random effects from both the year in which the data was collected as well as the specific mesocosm will also be plotted to determine their effects on the data.

Sampling an exhaustive amount of representative points can be used to get a precise depiction of the posterior distribution (Kruschke, 2015). To sample these representative points as mentioned above a Markov Chain Monte Carlo (MCMC) was used. One type of the MCMC algorithm that was used is the Metropolis sampler. The Metropolis sampler was chosen because the algorithm applies to the data set from this project. For the algorithm to be applicable the data must be continuous. The algorithm itself is quite general and can be applied to any number of dimensions and overall apply to general proposal distributions (Kruschke, 2015). The Metropolis sampler will result in a highly-resolved model of the posterior distribution of the 2016 data with a 2015 prior (Kruschke, 2015). This highly resolved model forms from creating the high amount of representative values (plausible parameter values) to get a distribution of posterior (Kruschke, 2015).

A Gibbs sampler was also used in comparison to the results of the Metropolis sampler. A Metropolis sampler can be problematic if the proposal distribution is not informed to the posterior distribution. A Gibbs sampler, although a type of Metropolis-Hasting algorithm, does not require the informed proposal distribution, and can apply to a model that has more than one parameter (Kruscke, 2015). The proposal distribution is determined on position in parameter space (Kruschke, 2015). Overall the proposal in a Gibbs sampler is dependent on both the location within the parameter space as well as what component parameter is selected (Kruschke, 2015). A Gibbs sampler will choose parameters from the conditional probability distribution haphazardly. When the parameter is haphazardly chosen by the sampler then the proposal distribution for the randomly chosen parameter has the conditional posterior of the parameter.

**Methods**

**Experimental**

Forty-five 1.6 x 1.7 x 0.6 m (110 gallon Rubbermaid Tuff Tubs) experimental freshwater mesocosms were established at the William H. Miner Agricultural Institute in Chazy, New York (44.891310°N -73.465989°W) in 2012. The mesocosms were buried with 10cm left above ground, located in an open grass field equally exposed to environmental conditions. The mesocosms were filled with approximately 1.5 m of soil and remained saturated with well water throughout the growing season. Refill planting occurred at the beginning of each growing season to maintain a density of 20 individuals per mesocosm. Individual species ranged from 2-10 individuals depending on mesocosm treatment. Of the 40 experimental mesocosms, there were 5 control groups with no plants, each functional group was represented individually 5 times, as well as the possible combinations between 2 and 3 functional groups represented 5 times (Table 1). Layout of diversity was chosen at random throughout the mesocosm configuration (Figure 1).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **A** |  | **B** |  | **C** |  | **D** |  | **E** |  |  |  |
| **1** | **FR** |  | **RT** |  | **RT** |  | **F** |  | **R** |  |  | **5.25'** |
|  |  |  |  |  |  |  |  |  |  |  | **2'** |  |
| **2** | **FRT** |  | **F** |  | **RT** |  | **R** |  | **C** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **3** | **T** |  | **FT** |  | **FT** |  | **T** |  | **FRT** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **4** | **C** |  | **T** |  | **FT** |  | **FRT** |  | **FRT** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **5** | **RT** |  | **R** |  | **FRT** |  | **C** |  | **FR** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **6** | **R** |  | **T** |  | **C** |  | **FR** |  | **RT** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **7** | **C** |  | **FR** |  | **FT** |  | **F** |  | **F** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **8** | **T** |  | **FR** |  | **FT** |  | **R** |  | **NUR** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| **9** | **NUR** |  | **NUR** |  | **NUR** |  | **NUR** |  | **F** |  |  |  |

Table 1. Table 1. Mesocosm experimental design (F-ferns ,R-reeds, T-tussocks). There are 5 replicates for each treatment. Numbers in parentheses represent the amount of mesocosms.

|  |
| --- |
| 1 Functional group F(5) R(5) T(5) |
| 2 Functional group F+R(5) R+T(5) F+T(5) |
| 3 Functional group F+R+T(5) |
| 4 Control No plants (5) |

Figure . Figure 1. Experimental layout of mesocosms. Functional group diversity and composition were randomized for each mesocosm.

**Analysis**

A data frame consisting of both years data was created having column identifications that included the mesocosm site id, the amount of: wetland functional groups present, stored nitrogen, stored carbon, aboveground biomass, and the year in which the data came from. Using R (R Core Team 2017) the maximum likelihood estimates for both 2015 and 2016 data was plotted for the response values of aboveground biomass, stored carbon, and stored nitrogen against the predictor values of amount of wetland plant functional groups present. A linear/MLE model made for 2015 for each of the response variables was plotted to determine the alpha, beta prior, and standard deviation. A normally distributed likelihood function was run as well as a prior function with the priors of alpha and beta. The priors were ran with a normal distribution and standard deviation was ran with an inverse gamma distribution.

Finally, a posterior function was ran to combine the likelihood function and prior with parameters together. Next a Metropolis sampler was used with 10,000 samples for each model parameter. A chain was then created through the library(coda). A summary of the chain was ran to get model parameter estimates. A plot was then fit with the superimposed data including the MLE from 2015, MLE from 2016, Metropolis sampler of 2016 with a 2015 prior, and the random effects from both the year and individual mesocosm. Random effects of both the year and mesocosm site were determined using the function lmer from the library(lme4). Besides a Metropolis sampler a Gibbs sampler was also used to compare the findings. The library(bayesm) was used in the Gibbs sampler. First a response variable was made that consisted of the ecosystem response variable from the 2016 data, this occurred 3 times due to 3 different variables being tested (2016 aboveground biomass, stored carbon, and stored nitrogen). Then a treatment variable was made using the function cbind and the amount of functional groups for 2016 data. The data for the actual Bayesian analysis came from a list function consisting of the response variable of 2016 data and treatment variable. A Gibbs sampler will use a normal prior for the regression coefficients and an inverse chi-square for the variable priors. A linear model/MLE was made of the 2015 data with the ecosystem services as response variables and number of functional groups as the predictor variable. The two coefficients were then made into a variable. A precision matrix was then made for the normal prior, and scale parameters were made for the inverse chi-square prior. There were 10,000 iterations done of the Gibbs sampler. In conclusion each figure for the response variables had a MLE for 2015 data, MLE for 2016 data, Bayesian analysis for 2016 data with the 2015 prior, and the random effects of year and mesocosm for both the Metropolis and Gibbs sampler.

**Results**

The 2015 MLE data line for stored carbon was found to have a slope of 58.7 and an intercept of 267.34. The 2015 MLE data line also began and ended at higher levels of stored carbon in comparison to the other data lines, ranging from approximately 330-450 g/m2 (Table 2). The 2016 MLE data line for carbon was found to have a minimal slope of 2.82 and an intercept of 276.69. The 2016 MLE data line overall was lower than 2015 MLE staying at approximately 280 g/m2 over the entire range of functional group diversity. The Bayesian 2016 data line formed by the Metropolis sampler had a slope of 71.21 and an intercept of 26.01. The Bayesian 2016 data line was the lowest data line of stored carbon ranging from approximately 100-240 g/m2 between 1 to 3 functional groups (Figure 2, top right). The Bayesian 2016 data line formed by the Gibbs sampler was found to have a slope of 43.16 and an intercept of 259.80. Unlike the Bayesian 2016 line formed by the Metropolis sampler, the Bayesian 2016 line formed by the Gibbs sampler fell between the 2105 MLE and the 2016 MLE data line (Figure 2, top left). The Bayesian 2016 data line had a range of approximately 300g/m2 at 1 functional group being present to 385g/m2 at 3 functional groups being present. The mixed effects model, random effects found a standard error of 308.96 for stored carbon intercept, and a standard error of 0.24 was found for the effects of year and mesocosm on stored carbon. The fixed effects results found an intercept of 297.62 and a slope of 20.09 for the effects of year and mesocosm.

The 2015 MLE data line for nitrogen was found to have a slope of 1.79 and an intercept of 3.90 (Table 2). The 2015 MLE data line again began and ended at higher levels of stored nitrogen in comparison to the other data lines, ranging from approximately 5.5-9.0 g/m2. The 2016 MLE data line for nitrogen was found to have a minimal slope, again mimicking the same pattern seen in stored carbon with a slope of 0.67 and an intercept of 4.19. The 2016 MLE data line ranged from roughly 5 g/m2 at 1 functional group diversity to 6.3 g/m2, this data line also fell below all other data lines present. The Bayesian 2016 lined formed by the Metropolis sampler had a slope of 1.71 and an intercept of 3.17. The Bayesian 2016 data line fell between the MLE 2015 and MLE 2016 data line, ranging from 5 g/m2 at 1 functional group to 8.4 g/m2 of stored nitrogen at 3 functional group diversity (Figure 2, middle left). The Bayesian 2016 data line formed by the Gibbs sampler was found to have a slope of 0.66 and an intercept of 4.20. Unlike the Bayesian 2016 line formed by the Metropolis sampler, the Bayesian 2016 line formed by the Gibbs sampler fell almost directly on the 2016 MLE data line (Figure 2, middle right). The Bayesian 2016 data line had a range of approximately 5.0-6.3 g/m2 of stored nitrogen. The mixed effects model, random effects found a standard error of approximately 0.00 for stored nitrogen intercept, and a standard error 0.00 was found for the effects of year and mesocosm on stored nitrogen. The fixed effects results found an intercept of 4.40 and a slope of 1.08 for the effects of year and mesocosm.

The 2015 MLE data line for aboveground biomass was found to have a slope of 17.55 and intercept of 279.64 (Table 2). The 2015 MLE data line had a range of roughly 290-325 g/m2 of biomass from 1 to 3 functional groups being present. The 2016 MLE data line for biomass was found to have a slope of 37.73 and an intercept of 233.74. The 2016 data line ranged from approximately 270 g/m2 at 1 functional group diversity to 345 g/m2 of biomass at 3 functional group diversity. The Bayesian 2016 line formed by the Metropolis sampler had a slope of 37.19 and an intercept of 45.83. The Bayesian 2016 data line fell below all other data lines, ranging from 50 g/m2 at 1 functional group to 125 g/m2 of aboveground biomass at 3 functional group diversity (Figure 2, bottom left). The Bayesian 2016 data line formed by the Gibbs sampler was found to have a slope of 16.71 and an intercept of 278.83. Unlike the Bayesian 2016 line formed by the Metropolis sampler, the Bayesian 2016 line formed by the Gibbs sampler fell almost directly on MLE 2016 and the data line was located relatively in approximation with the other data lines (Figure 2, bottom left). The Bayesian data line for Gibbs sampler had a range of approximately 290-330 g/m2. The mixed effects model, random effects found a standard error of 94.48 for aboveground biomass intercept, and a standard error or 0.59 was found for the effects of year and mesocosm on aboveground biomass. The fixed effects results found an intercept of 287.50 and a slope of 14.80 for the effects of year and mesocosm on aboveground biomass.

Table 2. Estimate, standard error, and p-value of Regression and degrees of freedom, sum of squares and percent contribution of analysis of variance.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable |  | Regression |  |  | Analysis of variance |  |
| **2015 MLE** | **Estimate** | **Standard error** | **p-Value** | **d.f.** | **Sum of squares** | **% contribution** |
| Stored Carbon | 58.70 | 40.62 | 0.16 | 2 | 55348 | 91.72 |
| Stored Carbon Intercept | 267.34 | 76.81 | 0.00 |  |  |  |
| Stored Nitrogen | 1.79 | 0.78 | 0.03 | 2 | 51.593 | 0.09 |
| Stored Nitrogen Intercept | 3.90 | 1.48 | 0.01 |  |  |  |
| Aboveground Biomass | 17.55 | 45.66 | 0.70 | 2 | 4947 | 8.20 |
| Aboveground Biomass Intercept | 279.64 | 86.34 | 0.00 |  |  |  |
| **2016 MLE** |  |  |  |  |  |  |
| Stored Carbon | 2.82 | 35.74 | 0.94 | 2 | 131.9 | 0.55 |
| Stored Carbon Intercept | 276.69 | 66.89 | 0.00 |  |  |  |
| Stored Nitrogen | 0.67 | 0.66 | 0.32 | 2 | 7.362 | 0.03 |
| Stored Nitrogen Intercept | 4.19 | 1.23 | 0.00 |  |  |  |
| Aboveground Biomass | 37.73 | 43.09 | 0.39 | 2 | 23650 | 99.41 |
| Aboveground Biomass Intercept | 233.74 | 80.62 | 0.01 |  |  |  |
| **2016 Bayesian Metropolis** |  |  | **95% Confidence Interval** |  |  |  |
| Stored Carbon | 71.21 | 35.30 | (46.17, 85.18) |  |  |  |
| Stored Carbon Intercept | 26.01 | 119.14 | (1.97, 48.74) |  |  |  |
| Stored Nitrogen | 1.71 | 0.79 | (1.44, 1.98) |  |  |  |
| Stored Nitrogen Intercept | 3.14 | 1.31 | (2.75, 3.61) |  |  |  |
| Aboveground Biomass | 37.19 | 59.01 | (6.94, 76.42) |  |  |  |
| Aboveground Biomass Intercept | 45.83 | 125.42 | (17.22, 53.32) |  |  |  |
| **2016 Bayesian Gibbs** |  |  |  |  |  |  |
| Stored Carbon | 43.16 | 35.75 | (26.67, 60.50) |  |  |  |
| Stored Carbon Intercept | 259.80 | 66.89 | (241.19, 278.60) |  |  |  |
| Stored Nitrogen | 0.66 | 3.87 | (-0.62, 1.98) |  |  |  |
| Stored Nitrogen Intercept | 4.20 | 7.22 | (1.74, 6.61) |  |  |  |
| Aboveground Biomass | 16.71 | 50.01 | (-0.19 33.80) |  |  |  |
| Aboveground Biomass Intercept | 278.83 | 55.47 | (260.41 297.26) |  |  |  |
| **Mixed Effects** |  |  |  |  |  |  |
| **Random Effects** |  | **Standard error** |  |  |  |  |
| Stored Carbon Intercept |  | 308.96 |  |  |  |  |
| Stored Carbon Year & Mesocosm |  | 0.24 |  |  |  |  |
| Stored Nitrogen Intercept |  | 0.00 |  |  |  |  |
| Stored Nitrogen Year & Mesocosm |  | 0.00 |  |  |  |  |
| Aboveground Biomass Intercept |  | 94.48 |  |  |  |  |
| Aboveground Biomass Year & Mesocosm |  | 0.59 |  |  |  |  |
| **Fixed Effects** |  |  |  |  |  |  |
| Stored Carbon Intercept | 297.62 | 481.04 |  |  |  |  |
| Stored Carbon Year & Mesocosm | 20.09 | 5.11 |  |  |  |  |
| Stored Nitrogen Intercept | 4.40 | 1.05 |  |  |  |  |
| Stored Nitrogen Year & Mesocosm | 1.08 | 0.55 |  |  |  |  |
| Aboveground Biomass Intercept | 287.50 | 79.26 |  |  |  |  |
| Aboveground Biomass Year & Mesocosm | 14.80 | 41.45 |  |  |  |  |

****

Figure 2. The Metropolis and Gibbs sampler applied to MLE of 2015 data, MLE of 2016 data, Bayesian 2016 data (2016 data with 2015 prior), and random effects of year and mesocosm site. Predictor variable is the diversity of number of functional groups and the response variables are amount of; stored carbon, nitrogen, and aboveground biomass.

**Discussion**

The goal of this project was to determine how Bayes theorem would affect conclusions drawn on results from wetland plant functional group diversity on ecosystem services of nutrient storage and aboveground biomass. Analysis of random effects caused by both year sampled and mesocosm site also occurred. To determine how Bayes theorem would affect conclusions on functional group diversity and ecosystem services, two MCMC samplers of Metropolis and Gibbs were used. There were two reasons as to why two different samplers were used; 1) The data sets were able to fit both types of samplers 2) Curiosity to see how the outcomes of both samplers would differentiate from one another.

As a whole the Metropolis sampler resulted in the Bayesian 2016 data line almost always falling below the MLE 2016 data line. However, when looking at the variable of nitrogen the Bayesian analysis of the data was located almost directly on the 2015 MLE data line. Overall, it can be seen that with the addition of the 2015 priors on the 2016 data it caused the overall Bayesian 2016 data to fall below rather than fall between the years of data. The Bayesian 2016 data line also had minimal increase in slope for all variables being tested, which was interesting that Bayes would not predict an increase in services when there was an increase in functional group diversity. Although there were slopes, perhaps because there was not significance across the diversity and services this caused the slope to be minimal. That with a MCMC Metropolis Bayesian analysis the amount of functional groups did not have an apparent change in the amount of ecosystem services they provided.

Overall the Gibbs sampler created a much tighter, closer related set of data lines, although producing similar results to that of the Metropolis sampler. Gibbs sampler may be the sampler of choice because there were not outlier data lines that were present when the Metropolis sampler was used. But as mentioned before the Bayesian 2016 analysis even for the Gibbs sampler had minimal increase in slope. This similarity between slope trends for Metropolis and Gibbs sampler makes me believe that although there were increases in ecosystem services with functional group diversity that these effects were not actually significant when basing data on past priors.

Random effects caused by the year the data was sampled and the mesocosm site also had similar results when plotted in both the Metropolis and Gibbs sampler. The results were to be expected, usually falling between the high year MLE and the low year MLE for that ecosystem service. In conclusion Bayes theorem was applied to the data the results did differ from that of the MLE 2015 and the MLE 2016. This was not as expected but did change my overall perspective on what the conclusions of the effect of functional group diversity is on the ecosystem services. Observing the MLE of 2015 and 2016 it was seen that as functional group diversity increased so would ecosystem services. But with a Bayesian approach I believe now that the trends may not be as strong as they seem, that with allocation of probability, the slope may not actually be increasing but remaining stagnant as functional group diversity increase.

Literature Cited

Kruschke, J.K. 2015. Doing Bayesian Data Analysis, A Tutorial with R, JAGS, and Stan: 2nd edition. Elsevier Inc.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical, Computing, Vienna, Austria. URL https://www.R-project.org/.