A Longitudinal Analysis of Ecological Systems Valuation in The Lake Chad Basin

By Nicholas Grabill, Yifei Li, Meiqi Liu, and Heyue Cheng; supervised by Dr. Frederi Viens, Statistics & Probability

Introduction & Project Goals

This research investigates the impact of human and environmental factors on the hydrology of Lake Chad, a great shallow inland body of fresh water in the Eastern Sahel region of North Central Africa. We are specifically interested in the Northern section of the lake and its catchment basin. We would like to know how the rainfall patterns in the Yobe river basin affect the lake, and whether there is any noticeable impact of the water consumption of various crops grown in the region, which can be categorized into **thirsty** and **non-thirsty** crops depending on whether they require systemic irrigation. In addition, in Borno state in Nigeria, which borders the lake, agrarian activities include **"recessional"** agriculture. We investigate these variables' effects on the total volume of water in the northern section of Lake Chad. We hypothesize the amount of water flowing from the southern to the northern section of the Lake to be the most important factor to explain the volume of water in the lake's northern section. We investigate these questions using the methodology of **Bayesian multiple linear regression**. The lack of knowledge of the level of uncertainty in the linear model (standard deviation of its statistical error) implies that a numerical method is required to approximate the **posterior distributions of the regression coefficients**. These coefficients' posteriors can then be analyzed to determine which factors are most statistically significant and which ones have the greatest influence. The broader goals include using these posteriors to build predictive models of the lake's hydrology based on local climate and agricultural production levels, and on long-term evolutions of the global climate. In the case of recessional agriculture, we compute the values of ecosystem services provided by the lake towards this type of activity, linking crop and livestock production levels and prices in the region, to lake levels, with a long-term goal of predicting ecosystem services values under various future climate-change scenarios. With this goal in mind, we hypothesize that crop and animal production levels in Borno state are highly sensitive to lake level variations, far more than in other states in the region.

Materials & Methods

Our Model

For this project, we are building a **Bayesian linear model**:

$$V_t = \alpha + \alpha_0 Q_{0,t} + \alpha_1 R_{YBD,t} + \alpha_2 R_{JBG,t} + \alpha_3 Thirsty_t + \alpha_4 NoThirsty_t + \alpha_5 T_t + \sigma\varepsilon_t$$

It explains the **volume of water** V for the lake's north side, via amount of water flow from south to north (Q_0) , rainfall near the lake (R_{YBD}), rainfall further away in the Yobe basin (R_{JBG}),the water consumption of thirsty crops ($Thirsty$) and of non-thirsty corps ($NoThirsty$). The T is global average temperature. The value ε is standardized statistical error, σ = standard deviation.

Materials & Methods Continued

Gibbs Sampler

The method we use to determine the significance and magnitude of each variable is to read them off of their **coefficients' Bayesian posterior densities, which we approximate using the Gibbs Sampler**. Since σ is unknown and must be estimated, contrary to popular belief, the posteriors are not known explicitly even though we use so-called conjugate priors. Bayes helps us make strong and informative probabilistic statements regarding our findings using credibility levels. The idea of Gibbs sampling is to **generate posterior samples by sweeping through each variable to sample from its conditional posterior distribution with the remaining variables fixed to their current values, and iterating**. At the i-th iteration, the algorithm for some variables is presented below:

$$\alpha^i \sim p(\alpha|\alpha_0 = \alpha_0^{i-1}, \alpha_1 = \alpha_1^{i-1}, \alpha_2 = \alpha_2^{i-1}, \alpha_3 = \alpha_3^{i-1}, \alpha_4 = \alpha_4^{i-1}, \alpha_{51} = \alpha_{51}^{i-1}, \sigma = \sigma^{i-1})$$

$$\alpha_0^i \sim p(\alpha_0|\alpha = \alpha^{i-1}, \alpha_1 = \alpha_1^{i-1}, \alpha_2 = \alpha_2^{i-1}, \alpha_3 = \alpha_3^{i-1}, \alpha_4 = \alpha_4^{i-1}, \alpha_{51} = \alpha_{51}^{i-1}, \sigma = \sigma^{i-1})$$

$$\alpha_1^i \sim p(\alpha_1|\alpha_0 = \alpha_0^{i-1}, \alpha = \alpha^{i-1}, \alpha_2 = \alpha_2^{i-1}, \alpha_3 = \alpha_3^{i-1}, \alpha_4 = \alpha_4^{i-1}, \alpha_{51} = \alpha_{51}^{i-1}, \sigma = \sigma^{i-1})$$

and in the same way for all the other variables in the model. By Bayes' theorem, the posterior densities are proportional to products of the model likelihood and the prior densities. By using the **following conjugate priors:**

- **normal distribution for each α , inverse gamma distribution for σ^2**

we can evaluate posterior densities for the same variables in the same classes of densities for every individual α and σ, and sample from these posteriors using built-in classical sampling at each step of the Gibbs sampler, **looping through all 8 parameters at each step, and repeating the procedure 10,000 times** to ensure that, after a burn-in period, the sampling distribution for each parameter approximates the true posteriors.

Lake Data and Crops data

Our period spans 1995-2010. Our crops data for the region comes from publicly available data from the United Nation's Food and Agricultural Organization, the **CountryStat** service, which contains government-self-reported yearly production data by state for **Nigeria and Niger** among other African countries. The **quality of this data is presumably low**, and is difficult to assess. The water usage data was generated by multiplying this production data by published water footprint values [2]. We have access to good quality high frequency data for Lake Chad from a private dataset provided by Dr. Jacques Lemoalle, which was the basis for his analysis with Bader and Leblanc [1]. **The data for Q_0 and V are highly reliable.**

Ecosystem services modeling and variables

We estimated Ecosystem Services Valuation for **2000-2008** by summing the valuation of crops and livestock. **Commodities prices** were obtained from publicly available **rural market values**. These statistics were used to construct longitudinal representations of the data to view relationships with other aspects of the Lake Chad Basin, such as **lake surface difference**.

MICHIGAN STATE UNIVERSITY | College of Natural Science
Department of Statistics and Probability

Results

Figure 1. Gibbs sampler result **without standardization** on the crop data.

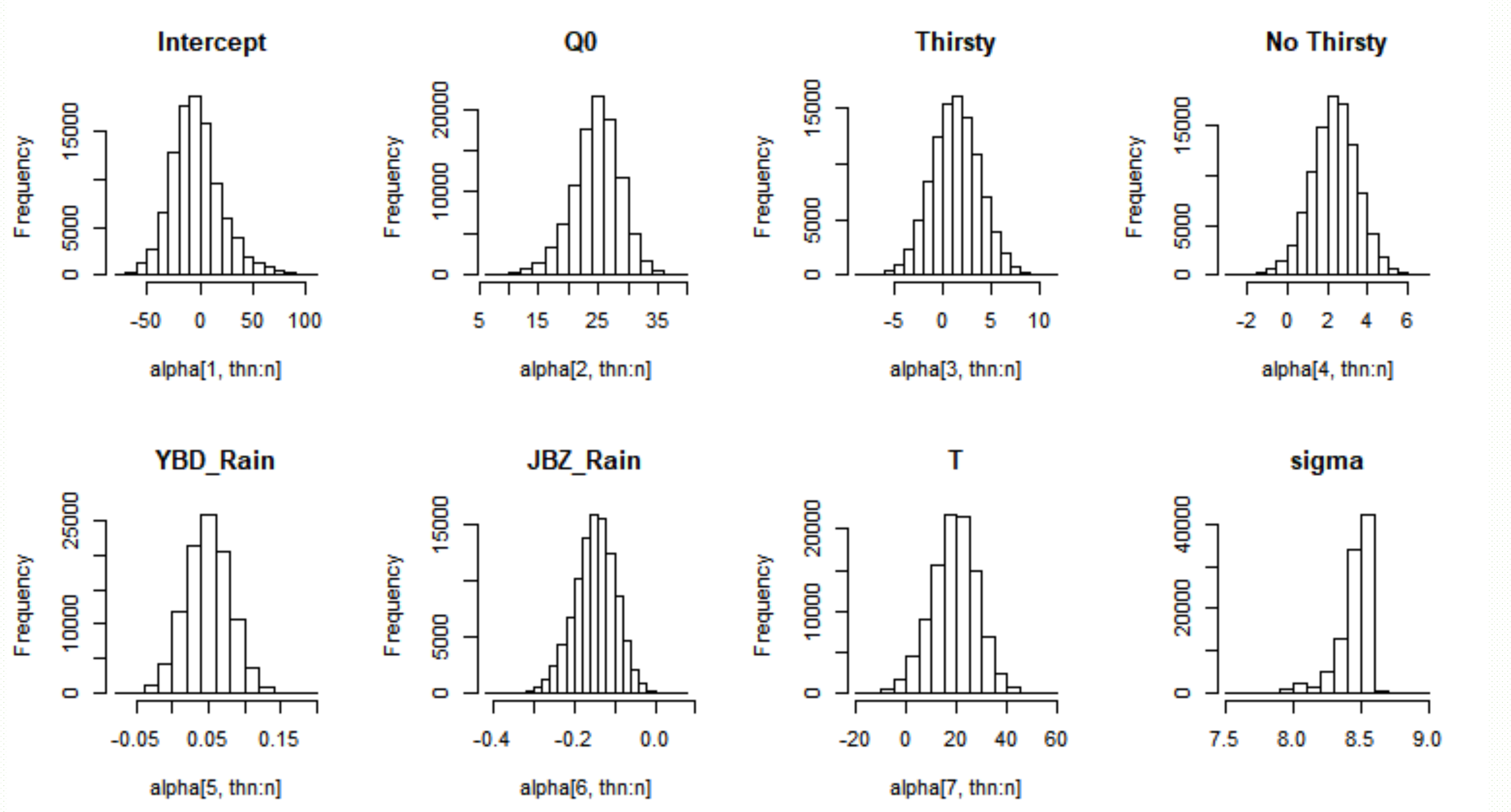
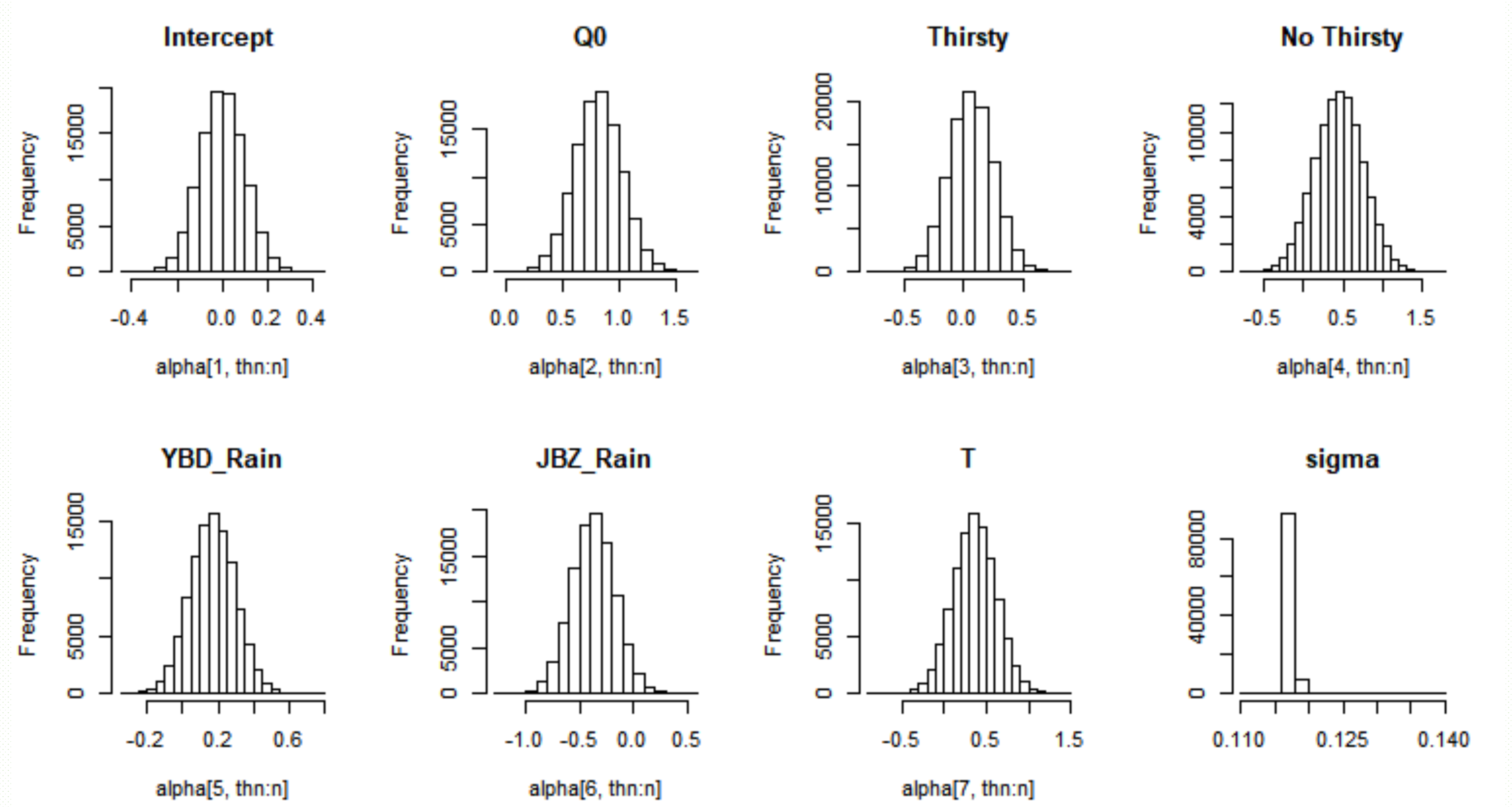


Figure 2. Gibbs sampler result **with standardization** on the crop data.



Conclusions

The graphs in **Fig. 1** and **Fig. 2** represent the **posterior probability densities** of the model's linear regression coefficients (α's and σ), using unstandardized (raw) data and standardized data. The result with standardization (**Fig. 2**) has a **very good fit**, σ is smaller compared with other variables' magnitude. There are two significant variables, Q_0 and R_{YBD}, they exceed 95% credibility level. As we excepted, **Q_0 is the principal contribution to lake volume V.** Rainfall and non-thirsty crops point in non-intuitive directions, indicating that they may be co-linear, both illustrating the effects of rainfed non-irrigated agriculture. Variables R_{YBD}, $Thirsty$, and T are at about the 90% credibility level. The climate indicator T appears positively related to V, a sound **rejection of the notion that global climate change impacts the lake volume negatively**. The rainfall in the nearby region R_{JBZ} is a positive factor, secondary to Q_0 but still significant, which follows our intuition. The $Thirsty$ variable is not at all significant, which **rejects the notion that irrigation activities have a negative influence on the lake volume.**

Conclusions Continued

The result without standardization (**Figure 2**) are **not as stable** as with standardization: indeed, see the very large size and non-significance of the intercept term, the size of σ which is on the same order as V and as the term involving Q_0 (magnitude of 10). This presumably indicates overfitting, much more likely than in the standardized model. However, **this model still rejects the notion that global climate change is causing the lake to shrink, and rejects that the irrigated cropping activities have any influence.** The other variables provide the same conclusions as in the previous model, providing **a robustness check** of the previous conclusions: **Q_0 is the most significant variable** influencing north lake volume, rainfall near the lake plays an important positive role.

In the ecosystem services analysis we found an **association** between the value of livestock traded on local markets, and the difference in lake coverage from the wet season to the dry season, indicating that **recessional pastoralism is of great economic importance** to the lake's agrarian societies

Future Work & Limitations

- A further analysis of the raw data overfitting (**Fig. 2**) could include PCA and other methods for variable selection.
- In both results, indications of possible collinearity for R_{YBD} and $NoThirsty$ could be tied to rainfed agriculture in the JBZ region to explain its influence on the lake.
- Limitations of the research involve confidence of agronomical measurements, and the paucity of measurement years. This is why our ecosystem services study was limited to 2000-2008.
- The use of remote-sensed data could help overcome the difficulties encountered in securing enough years of good quality agronomic measurements.

References

Thanks to Food and Agricultural Organization CountryStat service (http://www.fao.org/home/en/) for crops/livestock data, Open Data for Africa by the African Development Bank Group (https://dataportal.opendataforafrica.org/) for crops/livestock data prices, and Jacques Lemoalle for his hydrological data on the Lake Chad basin and discussions.

[1] Bader, JC, Lemoalle, J, Leblanc, M, "Modèle Hydrologique Du Lac Tchad." *Hydrological Sciences Journal*, vol. 56, no. 3, 2011, pp. 411–425., doi:10.1080/02626667.2011.560853.
[2] Brouwer, C., and M. Heilbloem. *Irrigation Water Needs.* Food and Agricultural Organization of the United, 1986.

Presenters' Contact Information:
Nicholas Grabill : grabilln@msu.edu
Yifei Li : liyifei8@msu.edu