STAT 6570 Final Exam

April 18th, 2019

Instructions:

Carefully read this document, as it contains all the instructions for this exam (no additional instructions will be provided to anyone). **All** aspects of this final exam must be done independently and cannot be discussed with anyone other than the instructor. You are responsible for ensuring that other students do not have access to your exam. Any violation of these instructions constitutes academic misconduct and will be reported to the Committee on Academic Misconduct (COAM). Being found in violation of COAM rules can have a negative impact on a student's progress through their degree.

The final exam must be submitted (in hard copy and electronically) by Monday, April 29th at noon. No late submissions will be accepted excepting cases of prior official justification. A hard copy must be submitted (either to me in person or placed under my office door; please do not use any binders or fancy clips, simply staple the report together). An electronic copy of the report and all computer code used must be submitted via Carmen (two drop-boxes will be provided). The report cannot exceed five pages (two-sided), excluding appendix and references. Supporting materials such as tables and figures can be provided in the appendix if they are clearly referenced in the main report. The appendix must contain your computer code, which you will also submit into a separate Carmen drop-box. Your submitted files should include your last name, first name, and course number. For example, my report submissions would be called "Chkrebtii-Oksana-STAT6570-final-project.pdf" and my software submission would be called "Chkrebtii-Oksana-STAT6570-software.zip". The online report must be submitted in pdf format. The software must run without any additional changes: this means that all relevant data files must be provided in the folder and that the code should be clearly named. If I am unable to run your code to reproduce all relevant results and figures, your grade will be impacted.

Evaluation:

The majority of the grade will be assigned for developing a reasonable Bayesian model for this data that can be used to address the research problem. You may consider multiple models, but due to space constraints you will need to choose one (in which case you should discuss why you chose this model over other candidates). Please note that I am not looking for any specific model or type of model. Instead, you will be evaluated on providing a well justified model structure without needless complexity. Obviously, the model should take into account the ecological mechanism that is being studied. You must explain your model as clearly as possible, justify any modeling choices and discuss the assumptions. You will also be evaluated on fitting the model to the data and then analyzing and interpreting the results. Style will be evaluated only to the extent that it clearly conveys your model and analysis, so it is in your best interest to submit a clear, well organized, and legible report. Figures are often helpful, but they must be carefully chosen/curated/constructed to make your point (do not provide dozens of figures in the appendix and expect me to go through them and analyze them without being prompted).

Background:

Wild salmon fisheries are valuable natural resources. Their successful management requires understanding of salmon population dynamics. Pacific Northwest salmon spawn at the site in which they were born. Spawning salmon leave the ocean to travel upstream of a river to lay their eggs (after which they die, having spent all of their energy on migration). The offspring hatch and begin a long migration towards the ocean, where they spend a fixed amount of time. Their whereabouts during this time are still largely mysterious. After a fixed time (depending on the species), they return to the original spawning site to reproduce. In this project, we will consider the spawning and recruitment behaviour of Pacific Northwest pink salmon which has a two-year life-cycle. Note that the pink salmon population is divided into two *genetically distinct species*: the odd-year pink salmon and the even-year pink salmon. These two populations lay their eggs in odd and even numbered years respectively. If you are curious about the salmon life-cycle, more information and references can be found at the National Park Service website.

Data:

The data comes from a public database on salmon spawning which contains the name and location of each spawning site (and therefore of each stock, since salmon always return to their place of birth) and its along-shore distance in kilometers from south to north (the southern-most spawning site has an along-shore distance of 0). The dataset also provides the number or “spawners” (S, adults who reproduce) and “recruits” (R, their offspring) in thousands at each site are measured for a number of years. I have shared with you a paper by Dorner et al. (2009) which describes this data. To give you a head start, I have included some R code to read the data and make some basic plots of the spawning locations etc. You may start by reading and running this code.

Research question:

The goal of this work is to study the patterns of spawning and recruitment of salmon at spawning sites along the Pacific Northwest coast and make predictions on any variables of interest to you in this context. Obviously this goal is very broad, but we must begin somewhere. Students must choose a specific aspect of this question, clearly explaining this in the introduction, and investigate it. Because of this freedom, you must be very clear about how your specific research question relates to the above stated goal. Here is an example of a thought process one might follow (you do not need to answer these questions necessarily): from the above brief description of the pink salmon life-cycle, it seems reasonable that spawning and recruitment should be correlated between subsequent salmon runs (separately for even and odd years): what are the features of this relationship? Do they make biological sense? Is location related to spawning abundance? Are there relationships between nearby sites that do not exist between sites that are further away from one another?

I did some applied work on this topic some years ago (Chkrebtii and Cao, 2013): you are welcome to read this paper if you have extra time, but note that it will most likely not be helpful to you as the analysis is not related to the research question posed in this project. In the past, some students have tried to re-use the summary statistics considered in the above paper without justification, losing many points as a result.

Scope:

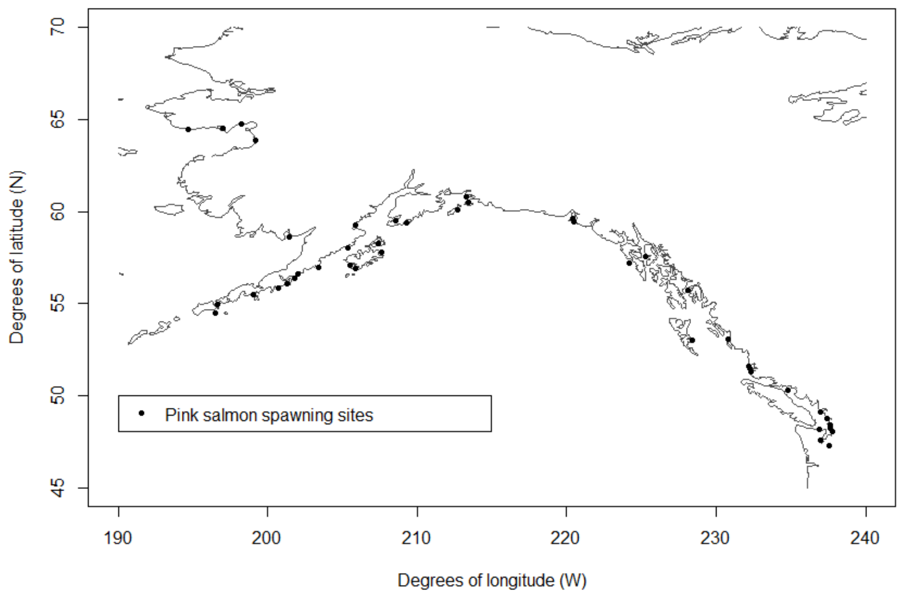
Due to the limited amount of time to complete this project, you must be careful to clearly define a manageable scope/research question. A complex model is not necessarily better than a simple, well-justified model that is appropriate to the question. Also, try to begin as early as possible so that you will have time to work out any coding issues in time for the due date. The grading rubric has been provided on Carmen. Note that no extra points are provided for the number/complexity of models considered, but points will be taken off for failing to appropriately motivate, justify, analyze, and diagnose your model.

Resources:

I have provided some R-code to begin visualizing the data. Please read this code and do not just use what is provided without question. Feel free to supplement this with your own visualizations. Also, if any figures appear in the report that are produced using my code, you must cite it in the report.

References

Oksana Chkrebtii and Jiguo Cao. Modeling spatiotemporal trends in the productivity of North Pacific salmon. *Environmetrics*, 24(1):31–40, 2013. ISSN 1099-095X. doi: 10.1002/env.2183.



Location of pink salmon spawning sites.

URL http://dx.doi.org/10.1002/env.2183.

Brigitte Dorner, Randall M. Peterman, and Zhenming Su. Evaluation of performance of alternative management models of pacific salmon (oncorhynchus spp.) in the presence of climatic change and outcome uncertainty using monte carlo simulations. *Canadian Journal of Fisheries and Aquatic Sciences*, 66(12):2199–2221, 2009.