Review of M. Sc. Thesis in Data Science

Matthew Horn: Building a Bayesian Survival Model to Estimate Left-censoring Effects and Correct Phenological Shift Estimates in Ecological Long Term Monitoring Programs

Examiners: Professor Jarno Vanhatalo / PhD Elina Numminen

	1	2	3	4	5
1. Research topic and purpose					X
2. Knowledge of the research field and related theories as well as use of literature				x	
3. Material, acquisition of material and analyses					x
4. Research results and reporting				x	
5. Examination of results (discussion) and conclusions				x	
6. Structure, clarity and general polish of the thesis					x
7. Work during the thesis process					x

Research topic and purpose

The research theme is timely and well-justified from both biological and methodological points-of-view. The research problem was independently developed by the candidate based on candidates own justifiable hypothesis that some of the earlier published results on trends in phenological events in the Finnish nature may have overlooked the effects of sampling and may actually be wrong due to that. The motivating hypothesis of the thesis is novel and very significant for ecology. However, there were no methods to answer this hypothesis. Hence, the aim of the thesis was to development these methods, validate their performance in a simulation study, and test them with some real data.

Knowledge of the research field and related theories as well as use of literature

The thesis shows good knowledge on the statistical and mathematical theory behind the developed methods. The work is based on timely and high-quality scientific source material, and shows insight and maturity. Relevant studies to the topic are cited, considered and contemplated in the work. However, these could have been re-discussed in the synthesis part slightly more. Discussion is easy to read and shows good understanding of the theoretical background of the work and its research context.

Material, acquisition of material and analyses

The datasets used were selected based on good knowledge of the litterature and research question. The chosen analysis methods were very appropriate for the task at hand. At times the used datasets could have been even better described, same holding partially for the methods. The analysis method was used in a critical and evaluative way, as e.g. the feature selection process was developed thoughtfully to the species and modeling question at hand, similarly the simulation study was utilized in a careful way to gain insights on the final model fitted.

Research results and reporting

The research question (model development and assesment) was thoroughly addressed in the results as a whole, and in general the presentation of the results is clear and logical. At times some figures and the interplay between the figures and the text could have been improved sligthly, and some figures may have improved from sligthly more explanation.

Examination of results (discussion) and conclusions

The candidate has really immersed into the developed results and contemplated them both critically and creatively, suggesting several routes for future research, and ways to further enhance the developed framework. The results have been investigated thoroughfully from several angles, and uncertainty and possible weakenesses are presented honestly along with the more general conclusions.

Structure, clarity and general polish of the thesis

The text flows well, is grammatically correct and structured in a logical way. The references are timely and used across the thesis and they are accompanied with plentiful interpretations, and the two are clearly distinguished

from each other. The bibliography could have been ordered more meaningfully and extended at times, but it does cover the references most relevant for the work.

Work during the thesis process

The candidate's work on the thesis has been very creative and independent – from setting up the research question, choosing the ways to tackle it methodologically, the ways to resolve the emerging challenges, getting to know the relevant litterature, and interpreting and validating the results. The work progressed at a constant fast pace according to the scheduled plan, The supervision was mostly contemplating together the ideas developed and results developed by Matthew.

Summary

The candidate has completed a multifaceted project which involved biology/ecology, statistical modeling, programming and data science. All the pieces in the puzzle have been thoughfully developed for the exact modeling question at hand and, thus, required understanding on both the ecological phenomena that motivated the methods development, as well as the statistical model. The results have been carefully investigated and everything critically contemplated in the end. The entire work progressed very flluently and promptly, and the candidate was able to independently come up with ideas and solutions needed along the work.

Suggested grade (1-5)*: **4***

In Helsinki on the 29th of May, 2024

Jarno Vanhatalo
Associate professor

Elina Numminen
University lecturer

Datatieteen maisteriohjelma
Matemaattis-luonnontieteellinen tiedekunta
PL 68 (Pietari Kalmin katu 5), 00014 Helsingin yliopisto
Puhelin (09) 1911, faksi (09) 1915 1120

Magisterprogrammet i Data science
Matematisk-naturvetenskapliga fakulteten
PB 68 (Pehr Kalms gata 5), FIN-00014 Helsingfors universitet
Telefon +358 9 1911, fax +358 9 1915 1120

Master's Programme in Data Science
Faculty of Science
P.O. Box 68 (Pietari Kalmin katu 5), FIN-00014 University of Helsinki
Telephone +358 9 19111, fax +358 9 1915 1120