Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Master Thesis: Point Processes for Species Modelling with Active Learning Citizen Science

Mojmír Mutný Andreas Krause
ETH

June 17, 2024

1 Motivation

Climate change, change in environment and diseases are one of the main factors influencing the size of species population as well as size of their habitat. In order to infer these changes specientist need to monitor these population by catch-release methods, using remote sensing or time-tested surveillance in the nature. The data obtained from such surveillance application have spatio-temporal nature. They are events recording of location, time and nature of the event. Such information is used to elucidate habitat of species, their dynamics and or source of problems. The data is used to fit event models which are then used to elucidate probability of certain events, most commonly using Poisson point process models (Heikkinen and Arjas, 1999; Kingman, 1993). An example of such fit is represented in Fig. 1. Example for recent large scale studies of habitat using event observations from satellite are seal and whale population monitoring: Gonçalves et al. (2020); Guirado et al. (2019).

2 Problem Statement

In this project, you will use observational data of plant species in Switzerland, among other data sources, to find a dynamic spatial model of their distribution. You will extend prior works on the subject Mutný and Krause (2021, 2022); Flaxman et al. (2017); Adams et al. (2009), and create an interactive maps of the habitat akin to the map in Fig. 1.

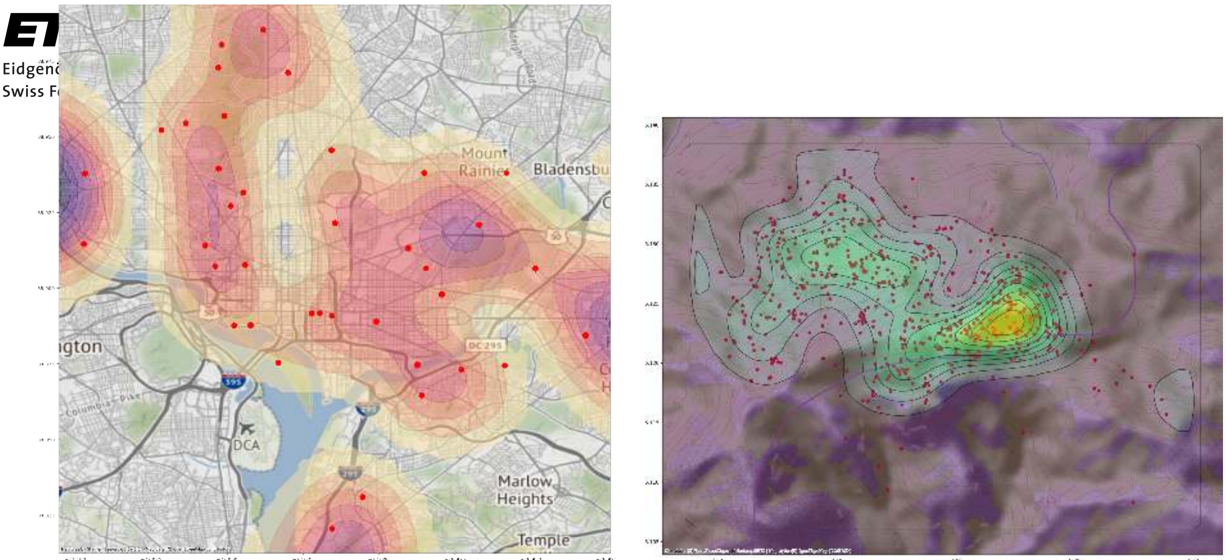
2.1 Active Learning

As the data is not perfect and there are dynamic processes influencing the habitat, the core challenge is to suggest further observational studies. By this, we mean designing an intervention algorithm that suggests where and when further monitoring/measuring of the species' presence should take place. The space of possible interventions is much larger than the available resources that can be devoted to surveillance, and hence there is a need for adaptive allocation of resources. This is where this project intersects with experiment design or *active learning* as it is more commonly referred to.

Additionally, the surveillance does not need to be done by professionals in many cases. A hiker equipped with a photo camera with GPS is often sufficient evidence for monitoring. This opens the possibility to suggest to users in a local environment to spot a certain species or perhaps record these dynamically. This trend is known as *citizen science* and is widely explored in Switzerland within the Swiss Data Science Center.

2.2 Challenges

- Analyzing large spatio-temporal datasets of plant species with point process mathematical models
- Implementation of novel spatio-temporal ML models with the goal of capturing different resolutions of data
- Design of informative data collection schemes (Experiment Design) for improving point process models



(a) Mosquito occurence in Washington DC area. (b) Habitat Gorillas in the Cameroon rainforest.

Figure 1: Images of fitted Poisson point process rate from suriveillance data.

3 Background

We are seeking a student passionate about spatial data and willing to learn about experiment design in the context of point processes. This thesis is well suited for data science, statistics, computer science, or applied mathematics master's program.

References

Adams, R. P., Murray, I., and MacKay, D. J. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 9–16.

Flaxman, S., Yee, W. T., and Sejdinovic, D. (2017). Poisson intensity estimation with reproducing kernels. *AISTATS 2017*, pages 5081–5104.

Gonçalves, B. C., Spitzbart, B., and Lynch, H. J. (2020). SealNet: A fully-automated pack-ice seal detection pipeline for sub-meter satellite imagery. *Remote Sensing of Environment*, 239:111617.

Guirado, E., Tabik, S., Rivas, M. L., Alcaraz-Segura, D., and Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. *Scientific reports*, 9(1):1–12.

Heikkinen, J. and Arjas, E. (1999). Modeling a Poisson forest in variable elevations: a nonparametric Bayesian approach. *Biometrics*, 55(3):738–745.

Kingman, J. F. C. (1993). *Poisson Processes*. Claredon Press.

Mutný, M. and Krause, A. (2021). No-regret Algorithms for Capturing Events in Poisson Point Processes. In *Proc. International Conference for Machine Learning (ICML)*.

Mutný, M. and Krause, A. (2022). Sensing Cox Processes via Posterior Sampling and Positive Bases. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*.