Title: Species distribution models with point process models (and extensions!)

Abstract:
Species distribution models (SDMs) are built to predict the intensity of a species (or group of species) as a function of environmental variables. SDMs rely on information about the environment as well as information about species presence, which can come in a variety of forms. While systematic survey data in which the detection and non-detection of a species is recorded over a set of predetermined sites are preferable, often the best available data are “presence-only data”, which consist of a list of locations where the species has been reported to have been observed. While such “citizen science” data is cheap and abundant and hence is generally widely available, it presents unique statistical challenges to estimate the true surface of the intensity of the species.

The most popular methods for fitting SDMs to presence-only data include Maxent and pseudo-absence logistic regression, yet both of these approaches are subject to challenges in model implementation, interpretation, and assumption checking which can make ecological inference very difficult. For example, both of these methods require a number of pseudo-absences to be introduced, and there is a lot of literature exploring questions of how these should be chosen, with often contradictory recommendations. Moreover, neither method can accommodate spatial dependence in the points, which is commonplace for a number of reasons in SDMs. Finally, presence-only data are subject to “observer bias”, as the list of reported locations reflects the distribution of observers as well as the distribution of the target species.

In this talk, I will discuss PPM-LASSO, a method I have developed for fitting SDMs to presence-only data. This method fits point process models to presence-only data along with a Lasso-type penalty to improve predictive performance. I will discuss links that point process models have with Maxent and pseudo-absence logistic regression, and demonstrate the advances that this flexible framework offers to presence-only analysis, including the choice of pseudo-absences, diagnosing inherent assumptions of the model, and accounting for observer bias, among others. Furthermore, I will discuss extensions of PPM-LASSO to contexts in which species data are available across multiple sources, such as fitting a model to both presence-only data and repeated survey data with a combined point process and occupancy model.