Bayesian inference for categorical and censored regression models for phenological data

The study of phenology has allowed for a better understanding of ecosystem function and the impacts of climate change on animal and plant populations, as well as informing applications such as crop breeding or pest management. Phenological analyses usually focus on the timing of particular events, e.g. dates of first appearance. However, for many phenomena exact dates of particular events are more difficult to observe than the state of the system itself.

For example, surveys may record whether a plant is in flower or not, but not the exact date of flowering. Such observations categorize an organism's state and are often encoded as discrete, ordered classes or (semi-)continuous progression scores. A plethora of statistical models have been developed by ecologists to infer dates of interest from such data, often framed around taxon-specific applications. Many of these models are special cases of more general categorical and/or censored regression models, which can be estimated with general purpose modelling software.

I describe parallels between existing modelling approaches and apply Bayesian inference to this class of models. This inference framework allows the inclusion of hierarchical model structures to accommodate e.g. the integration of data sets using different modes of recording, or individual heterogeneity in phenological progression. I show that Bayesian hierarchical models can provide less biased estimates and more accurate uncertainty quantification than existing estimation methods, in particular when sampling assumptions are not fully met. This is a great advantage when applying this class of models to real world data, e.g. from citizen science observation schemes.

I present applications with a focus on avian breeding and feather moult phenology, where observational data often comprise a mixture of categorical and continuous scores, as well as state-specific sampling biases. I provide an R package 'moultmcmc' which implements fast inference for these models using Hamiltonian Monte Carlo samplers.