**Spatiotemporal Models for Ecologists**

**Homework #4 – Project ideas, data sets, and initial approaches**

Goal: Individual projects provide an opportunity to develop models that are specific to your area of study. Typically, through the process of developing a project idea, obtaining the data, and developing the model, one gains a greater understanding of the modelling framework.

So, far we have covered mixed-effects models (varying intercepts) and time-series models (linear, nonlinear, and multivariate state-space models). Over the next few weeks, we will be presenting models that have spatial and temporal dynamics as well as spatial autocorrelation in the random effects. These will provide the framework by which you will be able to address your question of interest for your project.

We would like each student to identify their project idea now, so that we can provide some guidance on it and hopefully ensure that it is a reasonable project for the class.

Files to turn in:

1-2 page written description of class project listing at a minimum:

(a) the data set to be used,

(b) the question to be addressed, and

(c) the structure of the spatial and/or temporal processes. How are spatial or temporal dynamics hypothesized to affect the state portion of the model? How is spatial or temporal autocorrelation hypothesized to affect the random effects portion of the model?

For the class project I am interested in modelling spatiotemporal trends of human fertility in Africa. I am planning on using data from the Demographic and Health Surveys (DHS) that are conducted in low-middle income countries. These surveys are considered to be of very high quality and measure important indicators of population health. I am specifically interested in trends in total fertility rate (TFR) and age specific fertility rates (ASFR). Fertility in Africa has decreased substantially over the past 50 years but is still well above replacement level; because of this, Africa is seen as the region in the world that will continue to drive population growth for the considerable future. In addition to impacting population growth, these indicators are important because fertility is seen as a component of a country’s development and there is a strong focus on reducing teen pregnancy rates.

In high income countries where complete birth registries exist, it is relatively easy to measure fertility trends over time and space. Measurements of fertility in low income countries depends on complete birth histories (CBH) or summary birth histories (SBH) collected as part of a survey. A CBH records separately each of a woman’s births and when they occurred; these can be used directly to calculate ASFR (births / person-time). SBHs record in total how many births a woman has had so far in her lifetime and require separate demographic models to infer ASFR. The DHS surveys groupings of households, known as a cluster. These clusters are geolocated so that a longitude and latitude are associated with each cluster but for privacy purposes they jitter the locations by a couple of kilometres. As input into the modelling process I will need to estimate ASFR at each of these survey clusters.

The model that I will initial try fitting to this data will be fit separately for each five-year age group of the mothers between 15 and 45. It is a generalized linear regression model for Poisson data with a log link function.

Here is the count of monthly exposures in each survey cluster for each mother’s age group, is the count of mother’s age specific births and is the estimated age-specific fertility rate. The log ASFR will be modelled as a linear combination of a global intercept, cluster level covariates (possibly including maternal education, income) and an error term. This error term should have autocorrelation over both space and time which will probably be modelled using a Gaussian Process (GP).

The obvious question I want to investigate is how fertility varies sub nationally in parts of Africa. I think the model hypothesized above is a start towards investigating that question. As described above, I will initially fit this model separately by age but I think it would be interesting to see if the GP could be extended to add another dimension and induce correlation over maternal age groups. One interesting aspect of the data process is the jittering of the cluster positions that the DHS folks do. I think it would be really interesting to also investigate the effect this has on estimation by doing some simulation experiments.