# Mixed models analysis: biodiversity and agriculture

*Team Agriculture*

*4 August 2014*

## Context

Model forms part of a multi-level analysis, looking at the effects of rural production variables on biodiversity over time. Level of replication is county.

## Questions

1. Incorporation of repeated measures time variable - necessary to have as fixed and random, or just fixed? (see potential random effects spec below)
2. Spatial autocorrelation beyond a (1|State/County) approach. Abutting county neighbours nested within state?
3. Sensitivity of the model to predictors: is this a separate question better dealt with by say, boosted regression trees? NB by sensitivity I mean, if we change each parameter (in turn) by 10%, which has the biggest change on the model, and what is the relevant changes. I am **pretty** sure I mean something different to effect size here, but I'm also borrowing the term from biodiversity offsetting.
4. General optimisation/suggestions?

## Prediction

We would like to predict whether the response slope is positive or negative in the following factual scenarios:

1. Pesticide usage risk decreases by c(10, 50, 80)%; land use change continues on current trajectory; and
2. Pesticide usage risk decreases by c(10, 50, 80)%; cropping yield and area of fertilised land increase; land use change plateaus; and
3. Pesticide usage risk continues as now; cropping types shift towards c("x", "y") crops; average farmsize decreases; and
4. [suggestions? - **Kelly** any specific questions that lead on from the Mineau paper?]

### Predictors (details)

```
predictor.matrix <- expand.grid(
  pesticide = c("continued_trajectory", "plateau", "decreasing", "increasing"), #numeric
  cropping_area = numeric; ha/county ha or ha/total non-urban ha,
  forest_cover = numeric, potential fragmentation multiplier,
  average_farmsize * yield = #numeric,
  time = 1997 - 2012 (at minimum))

where:
- pesticide = application/ha * risk * ha * crop (or some measure thereof) (**)
- cropping_area = proportion of county in crops (+/- split into fert/unfert),
and crops are split into functional groups (so yes,
either multiple separate predictors or a ??multivariate predictor)
- average_farmsize*yield (+/- fert) is aimed at getting a measure of intensity of the cropping

(**) Unecessary level of detail: Another way of trying to nabble out the
```

```
yield-vs-intensity could be to add a multiplier for fertilised land to the 'intensity'
scale (easy enough, I think it's in NASS). Or lumped as pesticides, fungicides and herbicides.
Or look at historical application of since-banned pesticides and their rates and change
from 'historic' (pre pesticide) application to 'modern' (so only two time points)
```

**Random effects**

We'd like to take account of:

- Repeated measures
- Differing rates of population growth
- State-level regulation (i.e. nesting within state)
- Urban area (where not accounted for in the fixed effects)
- Spatial autocorrelation

Proposing to do this using the following syntax:

```
+ (1|State/County); or (State/County|time_period)  #these are all getting at pesky
urbanisation effects we may choose to incorporate as fixed effects later
+ (1|size_urban_area); or (size_urban_area/time_period)
+ (1|populationsize)
+ Abutting neighbour analysis nested within state (**)

(**) I know how to do a distance matrix for use in (partial) mantels but I have no
experience in its implementation in mixed models, or the use of nearest neighbour-type
specs (yet).
```

## Response variables

Suggest separate models for the following responses:

- Birds (within and among functional groups)
- Pollinators (bees)
- Water quality

Table 1: Conceptual layout of functional groups bird data

| | species | habitat_functional_group | food | disturbance |
|---|---|---|---|---|
| 1 | birdspecies_a | farm | insect | intolerant |
| 2 | birdspecies_b | farm | grain | tolerant |
| 3 | birdspecies_c | forest | fruit | intolerant |
| 4 | birdspecies_d | forest | insect | tolerant |