

# Mixed models analysis: biodiversity and agriculture

*Team Agriculture*

*4 August 2014*

## Context

This 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.x'

## Questions

1. In some cases we will have time in as a fixed effect, where we want to predict based on current trends. In others we will only have it in as a random effect. Where we have it in as a fixed effect, is there a clear best option between raw pop size; proportion pop size over time; or scaling and centering popsize?
2. Spatial autocorrelation beyond a (1|State/County) approach. Abutting county neighbours nested within state? Or full distance matrix based on central lat/lon of county?
3. Any suggestions for accounting for the fact biodiversity measures are point estimates of a random size (citizen science data) (and annual) whereas the rest of our variables are by county, and predominantly 5 yearly?

## Prediction

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

1. Broad scale - land use change
  - a. Key disperser/pollinator species (univariate) and bird functional group (univariate) and among bird functional groups (multivariate; NMDS and permanova [adonis]) as a response to land use change
  - b. Same as above as a response to metrics of land use intensity
2. Finer grained - land management (examples only, using univariate responses above)
  - a. Projection:
    - i. pesticide usage risk decreases by c(10, 50, 80)%; land use change continues on current trajectory; and
    - ii. pesticide usage risk decreases by c(10, 50, 80)%; cropping yield and area of fertilised land increase; land use change plateaus.
  - b. Prediction:
    - i. Predicting effect of land use changes on function group diversity (etc)

## Predictors (details, examples for finer scale analysis)

```
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

## 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
- Soil types (probably, main soil type in state)

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
+ (1|soil_type)
```

## Response variables

Suggest separate models for the following responses:

- Birds (within and among functional groups)
- Grassland
- Forest
- Anthropogenic
- 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

## Family for response variables

1. Poisson for count data (presence only obviously)
2. Binomial to look at present bird levels as a proportion of historical levels (later)