DOM Manuscript framework

**Introduction**

* Occupancy has long been an important concept in ecology
* We’ve come up with ways to measure and model occupancy
* However, there are challenges
* Imperfect detection is near-universal, and it introduces bias
* Non-equilibrium is also a challenge
* Occupancy models address each of these problems
  + Initial form allowed for estimating detection-corrected occupancy
  + Dynamic extension gives change through time/accounts for non-equilibrium

*Dynamic occupancy model form*

* Ecological process is described by occupancy, colonization, extinction
* Observation process is described by detection probability
* <form figure>
* This is all jointly estimated via maximum likelihood
* All this requires a hierarchical sampling design
* <sampling figure>
* Multiple observations, multiple seasons. Not particularly strict, missing data okay
* This is the basic form – multiple extensions exist
* Three key assumptions
  + Closure
  + No false-positives
  + Heterogeniety – requires covariates

*Model building*

* To meet these assumptions, users have several key inflection points
* Must collect data in a way which makes false positives unlikely
* Must split data into seasons/surveys in a way that makes closure reasonable
* Must decide what covariates to consider to reflect heterogeneity
* After all that, should probably assess fit/performance somehow.. Goodness of fit is available.

**Review methods**

* Full WOS sample from 2004-2023
* All articles either citing Mackenzie or matching key search terms
* Split into 5, 4-year strata
* 20 pulled from each strata

**Results and discussion**

We processed a total of 100 articles, out of an estimated XXX based on hit rates.

<hit rate figure>

*Applications of DOMs*

* Used widely for variety of species
  + Some amount of articles used multi-season implementations; both joint and hierarchical.
* Lots of flexibility in study scale
  + Short term, long term
  + Small areas, big areas
  + Few sites, lots of sites…
* Has scale changed over time?
  + Figure of Primary\*Secondary\*Sites boxplot across strata
* Lots of variation in detection methods. Cameras increasingly popular (?)
  + Some cool ones too – whale airplanes, CWD assays, ranger records of poachers..
* Much data manipulated *post hoc* – many articles starting well before DOMs even existed.
  + Grinnell’s surveys?
* Variation in how the models were used. Parameters, Inference, Drivers, Predictions.

*Implementation of DOMS*

* Variation in degree of complexity here
* Number of covariates used in models varied. Most on which parameters? Environmental vs Structural covariates..
* Types of covariates really depend too. Most popular were things like.. Land cover?
* Some covariates more/less likely to be available without direct measurement via remote sensing, or to be dynamic through time. Some, like biotic interactions, often collected alongside survey data. Implications for prediction.
* Regardless of type, responses were relatively simple linear terms. In environment covariates in X% of articles, or in structural covariates for Y% of articles.
* Covariate selection shows preference for non-exhaustive candidate sets, and occasionally sequential protocols.
* Bayesian more likely to be a priori, but there is occasionally more advanced usage. Posterior predictive check; dropping non-informative terms. Or cross validation.
* Across the board, evaluation was quite low. GOF or out of sample
* Things are different based on objective (?)
  + Number of covariates, probability of complex responses, probability of evaluation.

**Synthesis**

* Complexity is an important discussion
  + We use DOMs because they account for known complexity; non-equilibrium, imperfect detection…
  + But are our responses appropriate?
  + Non-linear responses are abundant in nature. Particularly at large scales
  + This certainly affects things in SDMs.
  + Number of terms needed to capture heterogeneity is also important
  + If there is more data available, maybe more complex models are feasible
* Model selection may be important
  + It affects things in SDMs
  + Morin found ad hoc methods, not always the best…
  + More work needed on Bayesian methods in particular
* And what about model evaluation?
  + Contingent on the objective of the project, but…
  + GOF should probably be looked at for inference.. No guarantee of a ‘good’ model in small candidate sets.
  + And predictions should find value in validation.
* Lots of promising opportunities, too
  + Spatial prediction is still under-explored and hugely promising
    - Ties in with earlier thoughts on model selection, though
  + DOMs are very well placed for handling autonomous detection data.
    - Some questions around closure, of course.
    - And false positives, for auto-classifiers.
    - But also some interesting ideas about continuous detection (Gardner paper).

**Conclusion**

DOMs are important and really good tools for important issues. Well-suited for a lot of pressing problems and newly available data types.

But we need to talk more about complexity, model building and evaluation, and test how that can affect performance.