Contents

General notes

Link to Ecological Forecasting Initiative workshop website.

1 Characterizing uncertainty

Presenter: Shannon LaDeau

Classic assumptions of the linear model:

- Homoskedasticity
- No error in X variables
- Error in Y is measurement error
- Normally distributed error
- Pbservations are independent
- No missing data

These assumptions are for data from agricultural experiments, but these assumptions often don't hold for a lot of ecological data. Great for getting a statistically significant slope but that might not be great for making a forecast.

$$y_i \sim \beta_0 + \beta(x_i) + \epsilon_i \tag{1}$$

Data model:

$$y_i \sim N(\mu_i, \epsilon_i) \tag{2}$$

Process model:

$$\mu_i = \beta_0 + \beta(x_i) \tag{3}$$

Parameter model:

$$\epsilon_i \sim N(0, \tau) \tag{4}$$
$$\beta \sim N(\beta_{bar}, \nu) \tag{5}$$

Can also write this model in graph notation. In this case we break the model into a data model, process model, and parameter model. Three layers where the data model is on top, the process model is second, the parameter model is third.

Beyond the classic assumptions, what are you going to do with real ecological data? What are you actually doing when you're working with data? Assume that the data are random samples from a true population. We describe the population as a distribution because we can't sample everyone so we are using a sample. The distribution let's us make some assumptions about how the sample is related to the whole population. The distribution also helps us estimate what we didn't measure. The expected relationship between the population and the sample is described by a probability distribution.

For non-normal data (e.g. a 0/1), we have a data model, a process model that contains a link and a linear model, and then a parameter model. Another assumption that might be violated is constant variance, so could model the variance as a function of data. Shannon presents an example in which the data are modeled with a constant versus a changing variance, which shows that the credible interval changes if using a model with variance changing. Capturing uncertainty in this case does not mean improving your estimate of the response but rather actually means doing a better job of capturing the increase in variance.

Now we move on to observation error. A regression model assumes that all error is in the Y [$Y \sim N(\beta_1 + \beta_2 x_i, \sigma^2)$. But sometimes we also don't measure X very well either, and so there is error in the predictor variables. Imagine that you are getting error in X but you are not measuring that. One way of modeling that is considering errors in variables. Here we are modeling X as a random variable.

$$Y \sim N(\beta_1 + \beta_2 x_i, \sigma^2) \tag{6}$$

$$X^{(o)} \sim N(\chi, \tau^2) \tag{7}$$

Now we move on to latent variables. These are any variable that is not directly observed such as: missing data, variables that are measured with error (biased or random), and proxy measures. Ignoring variable latency (e.g. modeling a .... ???)

With a good model, you might be able to make predictions of missing data. For example, you might be able to use regression to fill it in. In a Bayesian framework, the model can predict missing Ys but you can do the same thing and estimate missing predictors. However, to do this you need to develop a model for the missing X and a model for how those missing X influence the Ys. To make this work, you need to assume that data is missing at random. Data that is missing systematically can not be estimated but you need a better model if it's missing systematically. It's a powerful approach but you're not going to do magic with the data.

An example comes from a forest ecology FACE experiment at a pine plantation in Duke Forest. The experiment measured the seed response to CO2 fumigation. So to get fecundity of trees, there's often a relationship between tree size and seed number (seed number as a function of diameter). In Shannon's data on cone counts, there is a lot of individual variability. So they used data on cones and seeds to inform (latent) fecundity estimates.

They were able to combine data from seed baskets to constrain what is observation error and what is natural variability. The study is described in **?**.

2 Hierarchical Bayes

Presenter: Mike Dietze

2.1 Hierarchical models

Hierarchical models can help us identify as-yet uncharacterized variability. They can help us deal with the fact that we can never deal with all variability.

Imagine that you're confronted with making the same measurement over different observational units. For example, it could be years or sites. One way of dealing with this would be lumping all data together and estimating a single mean (μ from Y_1, Y_2, Y_3) - a common mean. Another would be estimating independent means (μ_1 for Y_1, μ_2 for Y_2, μ_3 for Y_3). In reality, there is often some pattern in between. They provide an intermediate case to represent the continuum where there is some sharing. In a simple hierarchical model for a mean, there would be a mean that you're estimating at each site where you estimate a mean at each site (θs) and then an across-site mean (μ).

Partially pooled models allow you to partition the variability within each site, and the variability across sites. Partition into across-site, and within-site. This represents a continuum because it represents what you see in the data. If the data tells you there is a lot of variability from site to site, it will behave more like the independent means model. If the data tells you there is little variability from site to site, it will behave more like the common means model. The data will tell you how this works.

The model here represents the idea of borrowing strength. The inference that you make at each site borrows strength from all other sites. This is because the across-site mean is the prior for each site mean.

A model with independent means and common variance will look like this:

$$y_i \sim N(\mu_k, \sigma^2) \tag{8}$$

$$\mu_k \sim N(\mu, \tau^2) \tag{9}$$

$$\sigma^2 \sim IG(s_1, s_2) \tag{10}$$

Here, the model is fitting each data set independently but assuming the mean for each has the same prior.

For a hierarchical model, we want the mean μ to be an unknown that we're fitting based on the data.

$$y_i \sim N(\mu_k, \sigma^2) \tag{11}$$
$$\mu_k \sim N(\mu, \tau^2) \tag{12}$$
$$\sigma^2 \sim IG(s_1, s_2) \tag{13}$$
$$\mu_k \sim N(\mu_0, V_\mu) \tag{14}$$
$$\tau^2 \sim IG(t_1, t_2) \tag{15}$$

Here are some of the key take home messages about HM models.

- We can model variability in the parameters of a model. This is helpful if we are (for example) interested in variability among sites but there is a common process that constrains the process.
- We can partition variability more explicitly into multiple terms. In the simple example, we partitioned variability in the mean. But this can extend to more complicated cases.
- We can borrow strength across data sets. Data-rich parts of analysis can help constrain data-poor parts of the analysis. To the extreme, is when there is no data. Hierarchical models formally let you distinguish between within and out of sample inference.
- Details of hierarchical models are usually in the subscripts.
- Models are typically hierarchical with respect to the parameters.

In relation to the third point, out of sample predictions are possible with HM models. Formally, we think that out of sample predictions should be more uncertain than in-sample predictions. For example, consider the idea of fitting models to data from 3 sites and then trying to predict for a fourth site. If all sites are fit independently, then we have no way of making predictions about the fourth site.

2.2 Random effects

Random effects are a common *special case* of hierarchical models. Here's the model with partial pooling:

$$y_i \sim N(\mu_k, \sigma^2) \tag{16}$$
$$\mu_k \sim N(\mu, \tau^2) \tag{17}$$
$$\sigma^2 \sim IG(s_1, s_2) \tag{18}$$
$$\mu_k \sim N(\mu_0, V_\mu) \tag{19}$$
$$\tau^2 \sim IG(t_1, t_2) \tag{20}$$

Here's how the above model is related to random effects

$$y_i \sim N(\mu_g + \alpha_k, \sigma^2) \tag{21}$$
$$\alpha_k \sim N(0, \tau^2) \tag{22}$$
$$\sigma^2 \sim IG(s_1, s_2) \tag{23}$$
$$\mu_g \sim N(\mu_0, V_\mu) \tag{24}$$
$$\tau^2 \sim IG(t_1, t_2) \tag{25}$$

Some important things to note are that random effects always have a mean 0 (see α_k). This is the same as saying that the difference from the global mean is unbiased. Random effects variance attributes a portion of uncertainty to a specific source (subscripted by k).

Random effects are generally applied to aspects of a study that would not be the same if replicated. For example, this might include things like plots, block, year, individuals, etc. There is also often some amount of replication for the random effect to be identifiably different from the residual "noise" term $J \sim N(0, \sigma^2)$. In practice, random effects are often used to account for a lack of independence among these units. In contrast, fixed effects include things like treatments and covariates of interest.

We are using random effects to account for unexplained variance associated with groups. This may point to scales (temporal, spatial) that may need additional explanation. Adding covariates or fixed effects may explain some of the variance but there's always something unmeasured and sometimes additional fixed effects are not justified.

Can use these random effect estimates to look for correlation between variability and covariates without having to rerun the whole model.

Big points: properly accounting for uncertainty can be just as important for making valid inference from your model as the process model and covariates. Random effects are used to account for the impacts of unmeasured/unmeasurable covariates.

An example of this approach can be found in **?**.

2.2.1 Model 1: Global mean

```
model{

mu ~ dnorm (0, 0.001)
sigma ~ dgamma(0.001,0.001)

for(t in 1:nt){
   for(b in 1:nb){
      for(i in 1:nrep){
      x[t,b,i] ~ dnorm(mu, sigma)
      }
   }
```

```
}
}
```

2.2.2 Model 2: Random temporal effect

```
model{

mu ~ dnorm (0, 0.001)
sigma ~ dgamma(0.001,0.001)

for(t in 1:nt) { alpha.t[t] ~ dnorm (0, tau.t) }
tau.t ~ dgamma(0.001,0.001)

for(t in 1:nt){
   ## process model
   Ex[t] <- mu + alpha.t[t]
   for(b in 1:nb){
      for(i in 1:nrep){
      x[t,b,i] ~ dnorm(Ex[t], sigma)
      }
   }
}
}
```

Could take the posterior of random year effects and look at what explains that variability. Plot those posteriors against covariates and then add those into the model.

2.2.3 Model 3: Random block effect

```
model{

mu ~ dnorm (0, 0.001)
sigma ~ dgamma(0.001,0.001)

for(b in 1:nb) { alpha.b[t] ~ dnorm (0, tau.b) }
tau.b ~ dgamma(0.001,0.001)

for(b in 1:nb){
   ## process model
   Ex[b] <- mu + alpha.b[b]
   for(t in 1:nt){
      for(i in 1:nrep){
      x[t,b,i] ~ dnorm(Ex[b], sigma)
```

```
      }
   }
}
}
```

2.2.4 Model 4: Random block and time effect

```
model{

mu ~ dnorm (0, 0.001)
sigma ~ dgamma(0.001,0.001)

for(t in 1:nt) { alpha.t[t] ~ dnorm (0, tau.t) }
for(b in 1:nb) { alpha.b[t] ~ dnorm (0, tau.b) }
tau.b ~ dgamma(0.001,0.001)
tau.t ~ dgamma(0.001,0.001)

for(t in 1:nt){
   ## process model
   for(b in 1:nb){
   Ex[t,b] <- mu + alpha.b[b] + alpha.t[t]
      for(i in 1:nrep){
      x[t,b,i] ~ dnorm(Ex[t,b], sigma)
      }
   }
}
}
```

It's really challenging to identify the number of parameters in a hierarchical model. The number of parameters is ill-defined by the structure of the model; instead the number of parameters in the model depends on what data it's fit to.

2.3 Nonlinear hierarchical models

These are possible but often take a bit more thought for deciding which parameters are considered random or fixed. One issue is that setting all parameters to random can often result in unidentifiability. Inclusion of covariates is also challenging

2.4 Prediction

Hierarchical models facilitate prediction about an unobserved species, site, years, etc. Formally, out-of-sample predictions integrate over random effects and are more uncertain than in-sample. The posterior for a new species, site, year could be updated with a relatively

small number of observations. Can improve predictions because they have this prior already from the HM model.

3 Expert elicitation

Presenter: Melissa Kenney and Mike Dietze

Develop probability distributions on priors. Develop probability distributions on likelihoods. Elicit structural aspects of the model.

The literature is often full of uninformative priors. For those who have prior knowledge about phenomena that we are interested in and have information about, expert elicitation helps construct meaningful priors. Expert elicitation is particularly useful when working in data limited situations.

An example where expert elicitation and model based forecasting have been combined is weather forecasting. For example, in the 1960s weather forecasts didn't do so well. But local weather forecasters did. In some cases, experts actually outperformed models initially. The experts were an important part of the forecast process and helped inform models.

Iterative: repeated updating of forecasts as new data become available. All of this data helped experts learn how they were doing with their forecasts and to improve their own forecasts over time. Models had the same benefits of making forecasts and seeing their forecasts improve over time.

Thinking about formalized expert judgment and model forecasts combine is a key way of improving ecological forecasts.

Problems that occur when constructing an expert elicitation. It's not trivial. We can think about our own judgment but if you are formally incorporating probabilistic elicitations, it's hard! This requires a lot of forethought and collaboration with social scientists.

Grainger-Morgan PNAS 2014 **?** identifies a few problems/issues. First, you need to focus on a topic where an expert can make an informative, predictive judgement. This isn't always the case. Need to set up the problem so someone can answer a probabilistic judgment. Second, terms like "likely" "about normal" are imprecise and instead need to focus on precise quantitative probabilities. Third, when working with experts are often over-confident and may construct over-confident probabilities due to mental shortcuts/heuristics. Fourth, want to get the right experts. You want to ask the right people. Capture a diversity of expert opinion. Fifth, you want to spend time on developing, revising, and implementing your elicitation protocol. Sixth, the situation needs to be realistic (e.g. give experts space/time to give answers). Finally, you would need to consider whether you want to aggregate the expert opinions into a single probability distribution or keep them separate.

Expert elicitation forces you to think about your priors.

Mike Dietze also not a fan of uniform priors.

Elicitation often happens with Cumulative Distribution Functions. Also, elicitation also works from outside in. For example, by guessing min and max rather than mean!

4 State-space

Presenter: Shannon LaDeau John Foster'

Broad framework for thinking about models is signal and noise. Confidence intervals get tighter by adding in more data. Predictive intervals - you add in un certainty. Width is based on observation error and how well you characterized the model.

When we're forecasting we often think beyond the range of the data that we have. The confidence and predictive intervals get bigger. Many of the issues with uncertainty come into play here: we're outside of the range of our data, there is uncertainty in the covariates.

Ecologists have not had a lot of success in forecasting 5-10 years out.

4.1 When to use state space models

State-space models are useful when latent variables are connected in time or space. State-space models are also dynamic models, where the current state depends on the past state. State in this vocabulary is the variable of interest.

$$X_{t+1} = X_t \tag{26}$$

A Bayesian state-space model has the process model and data model.

$$X_t = f(X_{t-1}) \tag{27}$$

Missing

State-space models are flexible frameworks for understanding how states of interest change through time. These models are quite general:

- ....

4.2 Random walk state space model

$$X_t = f(X_{t-1}) + \epsilon_t \tag{28}$$
$$Y_t = g(X_t) + \omega_t \tag{29}$$

...

Graphically, this can be broken into a process model that links X_{t-1}, X_t, X_{t+1} that are dependent on each other. There are parameters that are connected to all of the Xs. The Ys are conditionally independent on the Xs.

```
RandomWalk = model{

## data model

## process model

## parameters
```

The further you get from data, the more uncertainty there is. Shannon presents an example with flu data.

4.3 Dynamic linear state space model

$$X_t = f(X_{t-1}) + \beta_0 + \beta_1 X + \epsilon_t \tag{30}$$
$$Y_t = g(X_t) + \omega_t \tag{31}$$

Parts of the model are:

X = latent time series Y = observed dataq β = parameters in linear model ϵ = process error ω = observation error

4.4 Nonlinear state space model: capture-recapture

One example of this is mark-recapture data. Individuals are capture, marked, recaptured...

Data in this case look like this. An individual record consists of $Y_i = [1, 0, 1, 0, 0]$. This is compatible with several true states:

$$Z_i = [1, 1, 1, 0, 0]$$
$$Z_i = [1, 1, 1, 1, 0]$$
$$Z_i = [1, 1, 1, 1, 1]$$

The different possibilities are that it died, that it survived one more sampling period, or survived the entire sampling period. We don't know the exact time of death. We also do know that we didn't capture it in the second census. In the example that was put up, the parameter model also has variance that changes throughout time.

Survival is modeled with a Bernoulli. The process model has four possibilities (survival porbabilities):

$$P(X_t = 1|X_{t-1} = 1) = s_t \tag{32}$$
$$P(X_t = 1|X_{t-1} =) = 0 \tag{33}$$
$$P(X_t = 0|X_{t-1} = 1) = 1 - s_t \tag{34}$$
$$P(X_t = 0|X_{t-1} = 1) = 1 \tag{35}$$

The observation model has four possibilities (detection probabilities) [missing]

$$P(X_t = 1|X_{t-1} = 1) = \tag{36}$$
$$P(X_t = 1|X_{t-1} =) = \tag{37}$$
$$P(X_t = 0|X_{t-1} = 1) = \tag{38}$$
$$P(X_t = 0|X_{t-1} = 1) = \tag{39}$$

There's also a parameter component.

4.5 Assessment

1. 1. The ?latent state? is our variable of interest, which we infer from observations, but cannot measure perfectly. TRUE. The goal here was to make sure that we understand what the latent state is.

2. Using state-space models, we separate observation error and process error. TRUE. This is one of the strengths of state-space models. It's also challenging because there's a tradeoff. If there's any way of constraining observation error, you can get a better estimate of process error. The state-space framework focuses on these two sources of error.

3. State-space temporal models are also called dynamic models or hidden Markov models. TRUE. Dynamic models refer to the time series component. State-space models and hidden Markov models are interchangeabley used. You can write dynamic models that are not state-space models. You can also write Markov models that are not state-space models. It's the hidden component that connects it to latency.

4. State-space models are limited by common statistical assumptions, such as independent observations or Normally-distributed data. FALSE

5. Random walk models use the current state, plus error, to make forecasts. TRUE

6. A random walk model is often a reasonable null model. TRUE

7. Error terms must be static over time. FALSE

5 Dynamic models

Presenter: Mike Dietze

Dynamic models explain change over time. These are models that describe how the state of the system is changing over time: the future state is a function of the current state (and lags). Dynamic models force us to focus on the process that causes change. These models behave differently from regression.

Dynamic models are different from regression models, and behave differently.

5.1 Examples

Persistence/random walk: $X_{t+1} = X_t$

Exponential growth/decay: $X_{t+1} = \rho X_t$

[name?]: $X_{t+1} = \beta_0 + \rho X_t$ In this model the equilibrium is $\beta_0/(1-\rho)$. ρ controls the rate of approach to equilibrium. This model is a classic first order autoregressive model (??). A ρ close to 1 would imply high autocorrelation, whereas close to 0 would imply independence.

Quadratic: $X_{t+1} = \rho X_t + \beta_1 X_t^2$ This is equivalent to the logistic growth model. These are connected if $\rho = 1 + r$ and $\beta_1 = -r/K$.

Compatibility of hypothesis and forecasting. All of these different models are hypotheses about a dynamic system, and if you make forecasts with multiple models they are alternative hypotheses about the system works.

Environmental covariate: $X_{t+1} = X_t + \beta_1 Z_t$ This is a population growth model with an environmental driver. The effect this has is to add some variability around the mean persistence.

Something that happens a lot with shorter timescales, there are often environmental drivers that have seasonal or diurnal cycles as well as data that have cycles. While a seasonally varying Z may drive a seasonally varying X, they may not be in sync. There may be a lag/shift out of phase.

These can become more and more complicated. So, for example, you might consider the change in Z: Z(t+1)-Z(t) as the covariate.

What if, instead, the temperature is setting the carrying capacity. In this case, the carrying capacity becomes a linear function of the driver variable.

Can also consider an AR[2] autoregressive.

In addition to including current state into the system, you can also include previous states and that have memory (AR(2)). These can also help capture trends in systems.

5.2 Building a process based model

The best place to start is a classic box and arrow diagram. Boxes represent states, solid arrows represent fluxes, and dashed arrows represent influences. This can help you develop an understanding of your system without worrying about the specifics of equations. This can then help you develop the equations.

Developing the balance equations comes next. This helps you conserve mass. You can also include terms that are functions of covariates.

The next step would be to develop your flux equations. For example, what describes the transfer processes.

5.3 When are dynamic models appropriate for forecasts?

The distinction between a statistical and dynamic model is a subtle one. From the perspective of forecasting, there's not a big difference between phenomenological versus process-

based models. Both use the same statistical frameworks. The process models are slower to fit because there are more parts. (???)

If we want to predict tomorrow, knowing what's happening today/knowing where you are right now really should be one of the covariates for where you are going to be tomorrow. Step back and think about how you're treating the model because you are describing dynamics rather than fitting a curve through an x-y scatterplot.

You can also take a process-free and equation-free approach to predicting the future. Some of this comes out in machine learning approaches (see lecture this afternoon). Maybe return to this for Thursday afternoon roundtable.

6 Machine learning

7 PROACT

8 Propagating uncertainty

9 Analytical data assimilation

10 Ensemble data assimilation

11 Social science

12 Model assessment

13 Forecast infrastructure/FLARE case study

-
-
-

$$\begin{aligned} \dot{x_1} &= u(t)x_1 \\ \dot{x_2} &= (1-u(t))x_1 \end{aligned} \tag{40}$$

$$A_{11} = \begin{bmatrix} -1 & 0 & \dots & 0 \\ 0 & -1 & \ddots & \vdots \\ \vdots & \ddots & -1 & 0 \\ 0 & \dots & 0 & -1 \end{bmatrix}_{t \times t} \tag{41}$$

```
## Function that computes values of derivatives in the ODE system
## Takes the time step, system of differential equations, parameters
derivs = numeric(2);
control <- function(times0,y,parms,f1,...) {

  # x1 and x2 are the two entries in y (ode)
  x1=y[1];

  # values calculated by the interpolated function at different time points
  u <- f1(times0);

  derivs = c(u*x1,(1-u)*x1)
  return(list(derivs));
}
## Compiles the function control()
control=cmpfun(control);
```

Index

References

Dietze, M. 2017. Ecological Forecasting. Princeton University Press, Princeton, NJ.

Hobday, A. J., J. R. Hartog, J. P. Manderson, K. E. Mills, M. J. Oliver, A. J. Pershing, and S. Siedlecki. 2019. Ethical considerations and unanticipated consequences associated with ecological forecasting for marine resources. ICES Journal of Marine Science, **76**:1244–1256. Publisher: Oxford Academic.

Ladeau, S. L. and J. S. Clark. 2006. Elevated CO2 and tree fecundity: the role of tree size, interannual variability, and population heterogeneity. Global Change Biology, **12**:822–833. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2486.2006.01137.x.