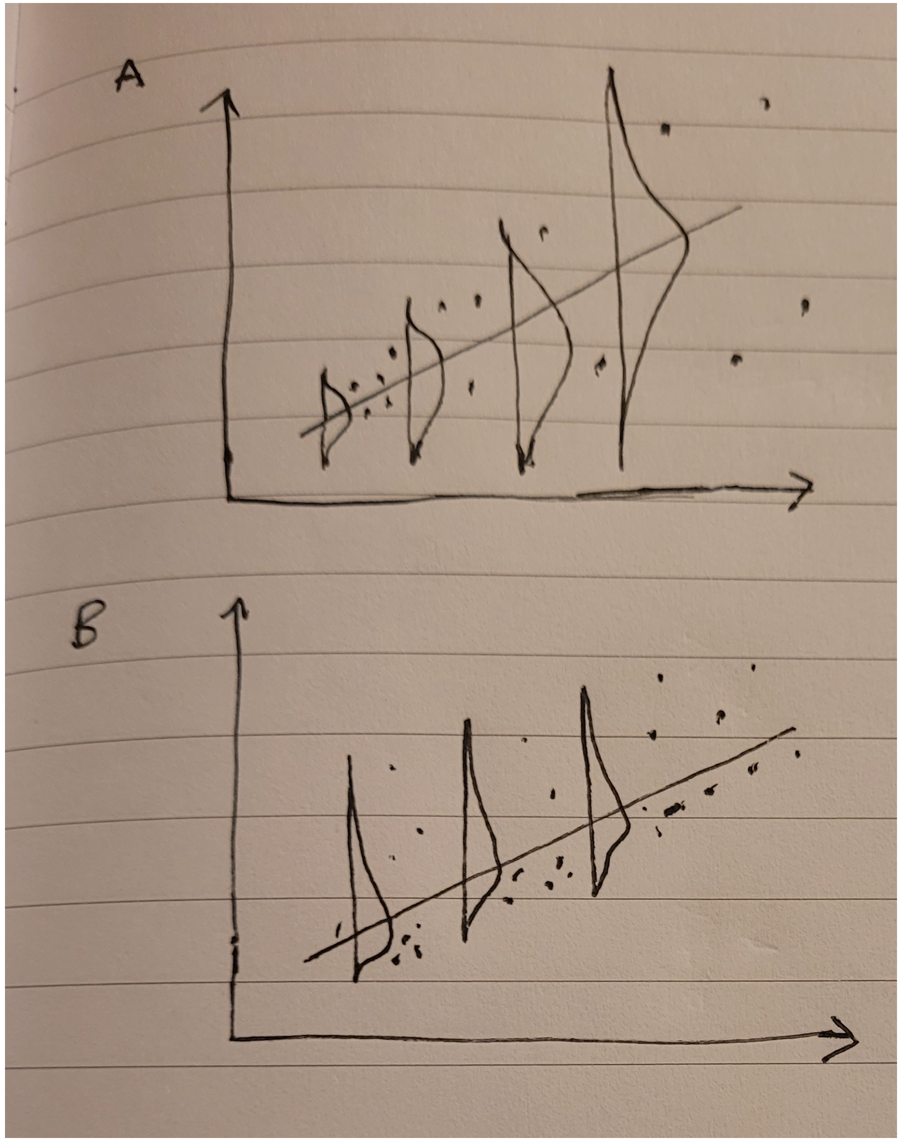
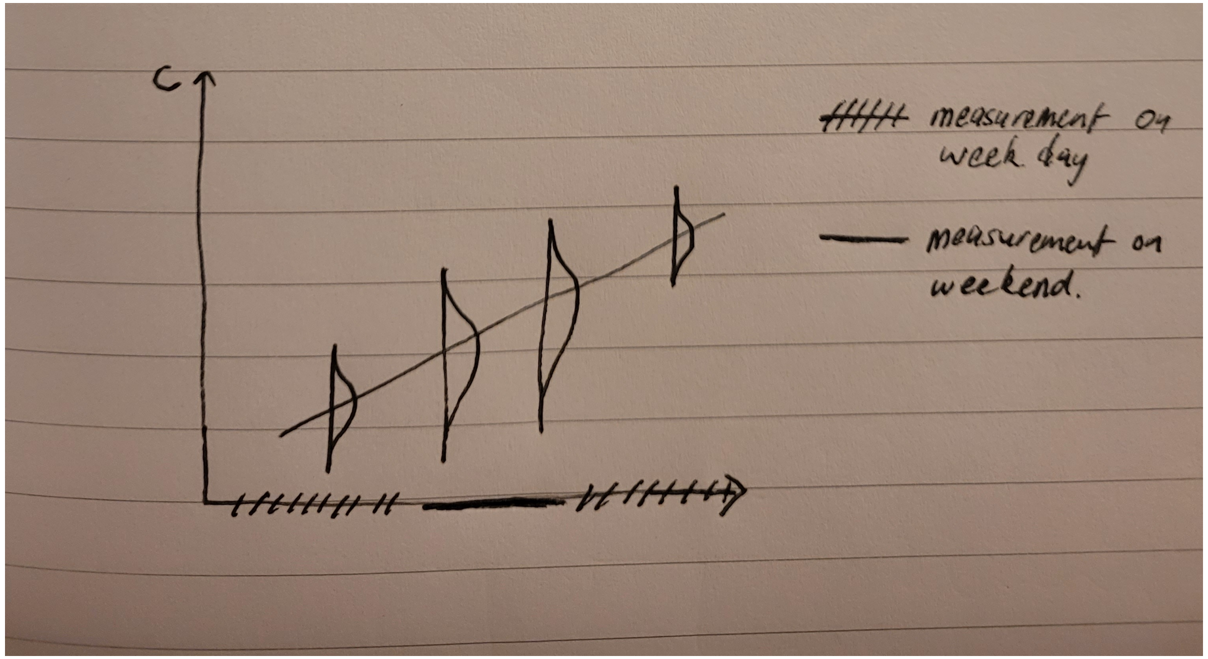
Ch 6. Characterizing uncertainty

Amy Hurford
Memorial University

Traditional statistical assumptions



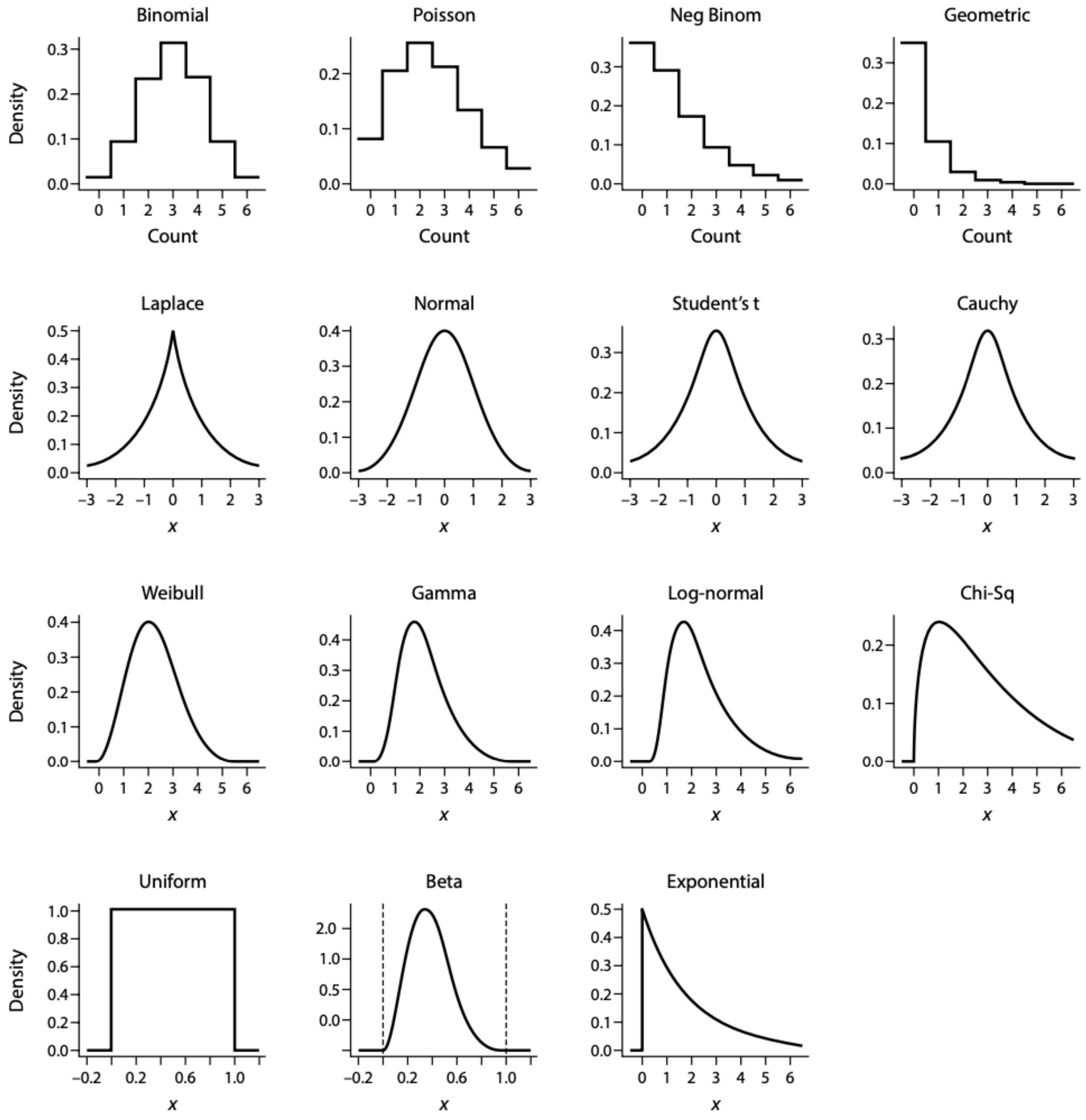


Probability distributions

Discrete			
Bernoulli(p)	$p^x(1-p)^{1-x}$	[0,1]	Success ($x = 1$) with probability p. Binomial with $n = 1$.
Binomial(n,p)	$\binom{n}{x}p^x(1-p)^{n-x}$	$[0 \ldots N]$	Number of successes, with probability p, given n trials.
Poisson(λ)	$\frac{\lambda^x}{x!}e^{-\lambda}$	$0,\infty$	Number of events, occurring at rate λ, to occur over a fixed interval.
Negative Binomial(n,p)	$\binom{x+n-1}{x}p^n(1-p)^x$	$0,\infty$	Number of trials, with probability p, before n successes occur. Also a Poisson-Gamma mixture.
Geometric(p)	$p(1-p)^x$	$0,\infty$	Number of trials needed before a success occurs. Special case of negative binomial with $n = 1$.

Dietz 2017. Ecological forecasting.

Probability distributions



Dietz 2017. Ecological forecasting.

Generalized linear models

Logistic regression:

$$y_i = Bern(p_i)$$ Data model

Boolean: TRUE/FALSE or 0/1

$$p_i = logit(\theta_i) = \frac{1}{1 + e^{-\theta_i}}$$ Link function

Converts real number to domain of the distributions mean

$$\theta_i = \beta_0 + \beta_1 x_i$$ Linear model

Dietz 2017. Ecological forecasting.

Observation error

A discussion on modeling observation error is straightforward if we believe that all of the observed residual error is measurement error and that there is no uncertainty in the model structure, covariates, or drivers.

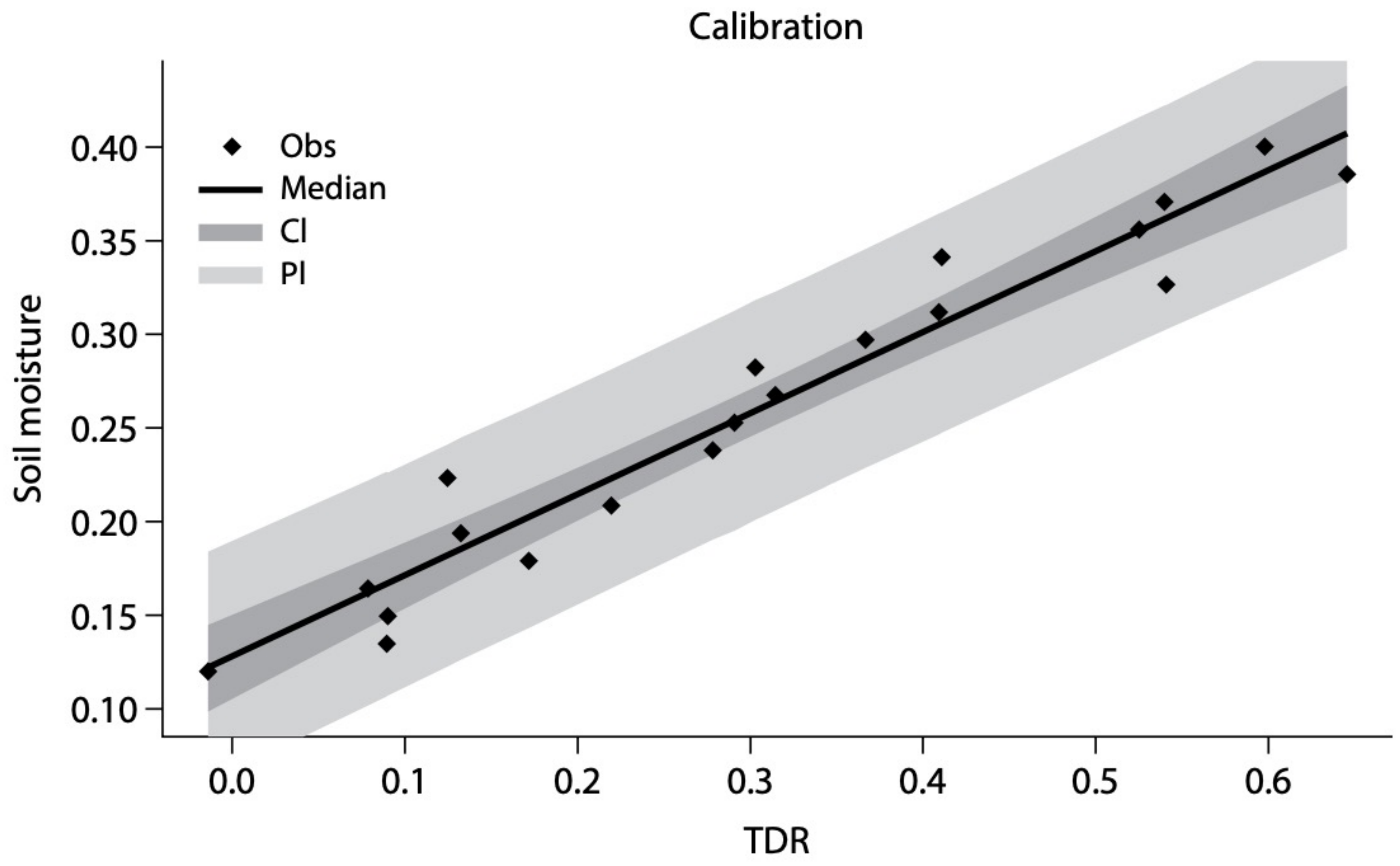
$$y_{i,obs} \sim g(y_i) \qquad \text{Data model}$$

$$y_i = \beta_0 + \beta_1 x + \varepsilon_{i,add} \qquad \text{Process model}$$

$$\varepsilon_{i,add} \sim N(0, \tau_{add}) \qquad \text{Process error}$$

Dietz 2017. Ecological forecasting.

Observation error



<u>Confidence interval</u>: uncertainty in parameter estimates

<u>Predictive interval</u>: all sources of model uncertainty

Based on how these are constructed in Exercise05B_Regression.Rmd (EF_Activities)

Random effects

- A fixed factor has categories, or levels, that we set to certain values in an experiment or levels that we choose in an observational study. We infer only to those levels.
- A random factor has categories that we have not chosen, or that vary even after we make them as uniform as possible.
- The same factor might be treated as fixed or random depending on the population of inference.

By Dave Scheinder

Fixed or random effect?

- treated versus untreated (control) units
- before versus after treatment of an experimental unit.
- day versus night
- habitat types
- insect stages (larval, adult)
- tanks in aquaculture
- plots in agriculture
- individual organisms

By Dave Scheinder

Hierarchical models

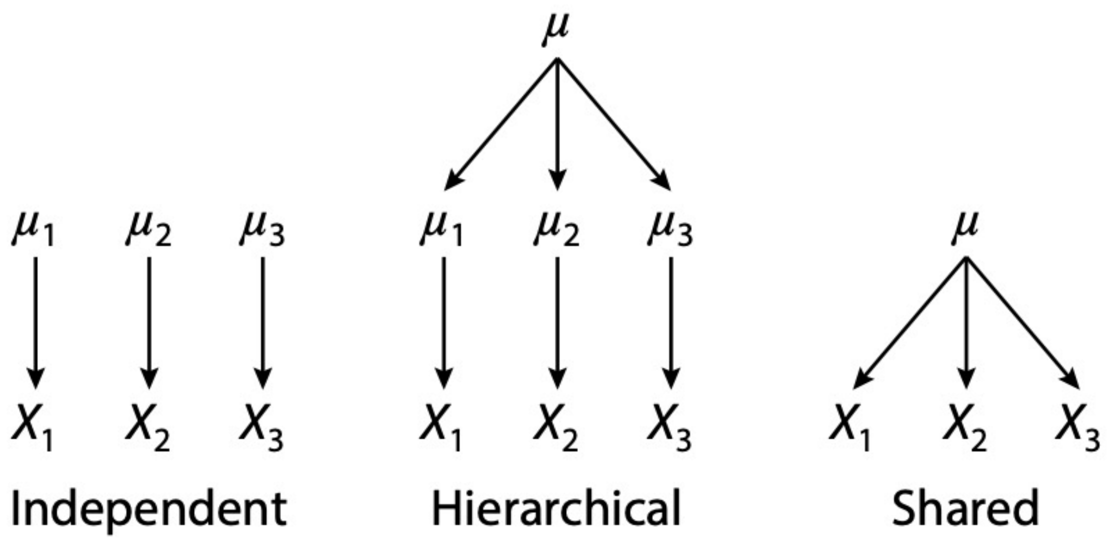
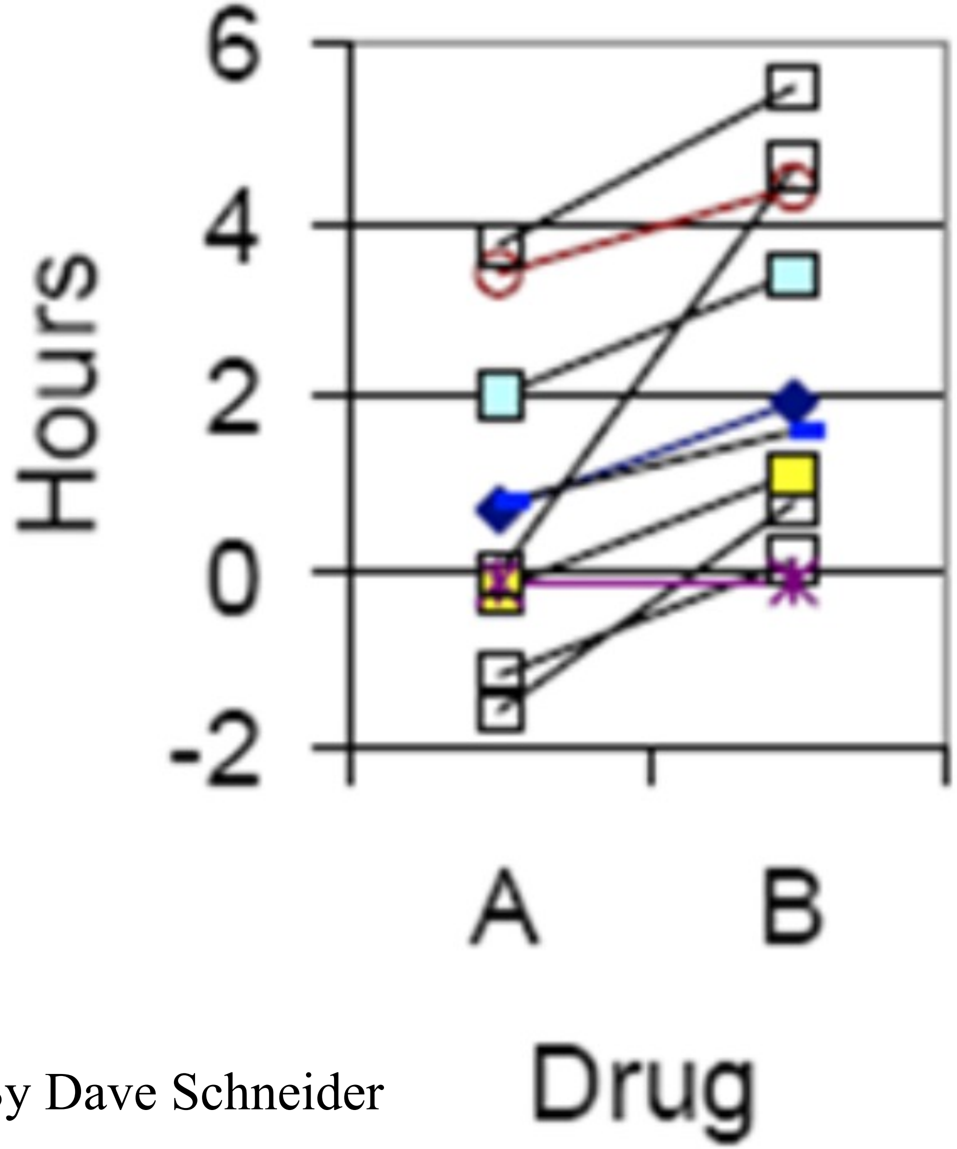


FIGURE 6.6. Hierarchical models represent the continuum of possibilities between treating data sets as independent versus treating them as identical. As a result, they partition process variability between the different levels of the hierarchy.

Dietz 2017. Ecological forecasting.

Hierarchical models



Account for variability in process or scale even if we cannot yet explain that variability in our process model

Dietz 2017. Ecological forecasting.

By Dave Schneider

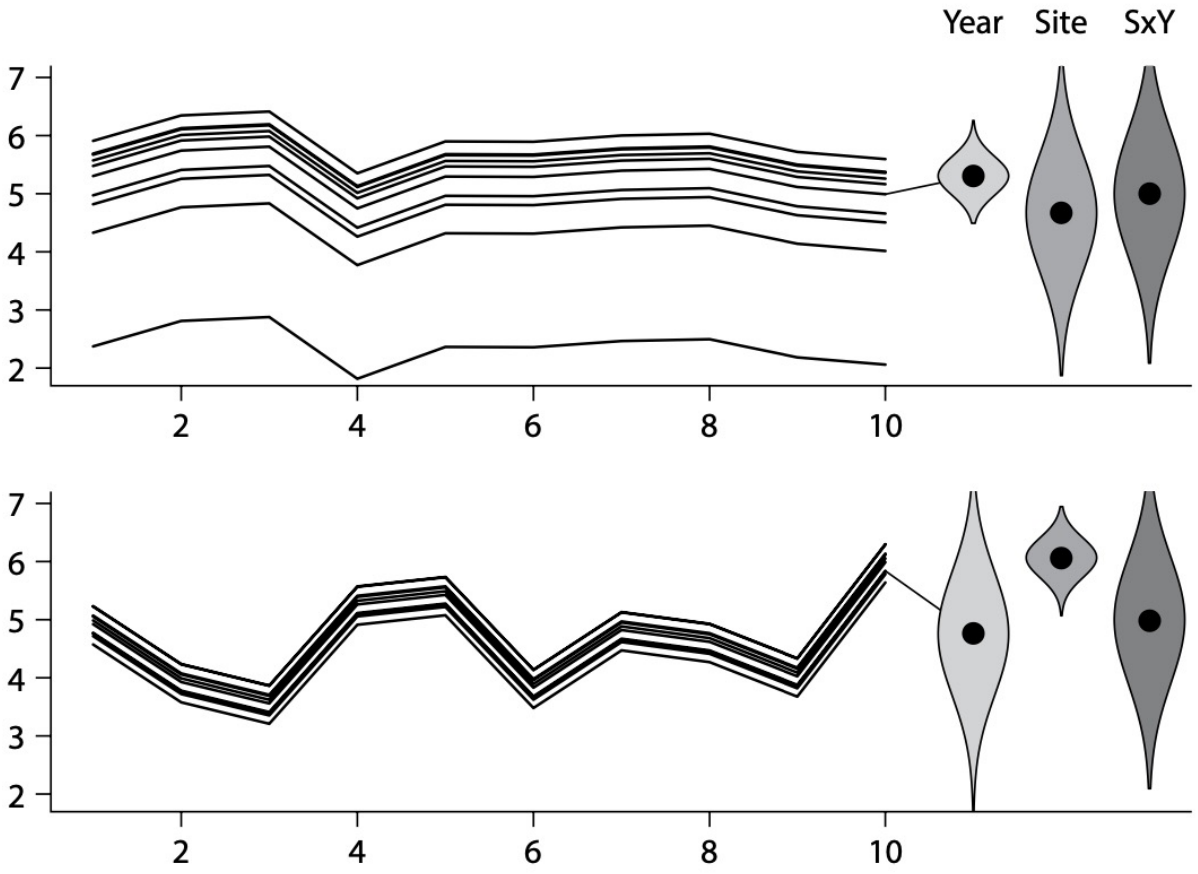


FIGURE 6.7. Impact of random effects on forecasts. Both panels show the trajectories of 10 sites over 10 years and have the same total variance. In the top panel 75% of the variability is site-to-site, while in the bottom panel 75% is year-to-year. On the right-hand side the violin plots under "Year" depict the prediction for a previously known site for the next new year. In this case the top example (low year-to-year variability) is more predictable. "Site" depicts the prediction for a new site in year 10, and in this case the bottom example (low site-to-site variability) is more predictable. Finally, "SxY" depicts the prediction for a new site in a new year, which is identical for both panels since they have the same total variability.

Hierarchical models are a powerful tool for estimating and partitioning process error

Dietz 2017. Ecological forecasting.

Conclusion

3. It is better to thoughtfully choose the appropriate data model than transform data to meet normality assumptions

6. Ignoring errors in variables (covariates, inputs) leads to falsely overconfident and potentially biased conclusions

7. Derived data products should be used with caution, especially if they lack a rigorous partitioning of uncertainties

11. Partitioning process errors improves forecasts because different errors do not propagate the same ways in space and time