Fitting growth models

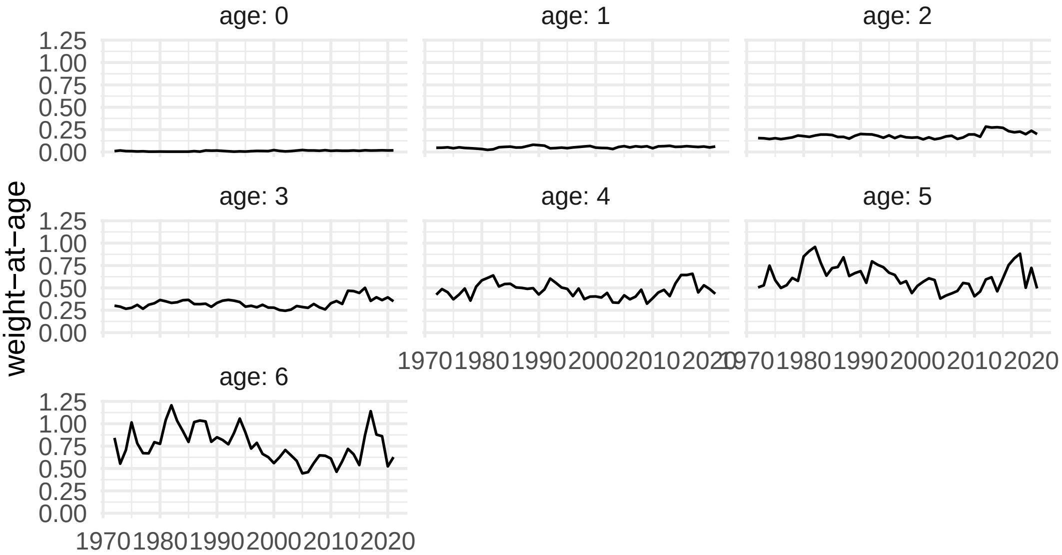
Mollie Brooks
DTU Aqua

November 2024

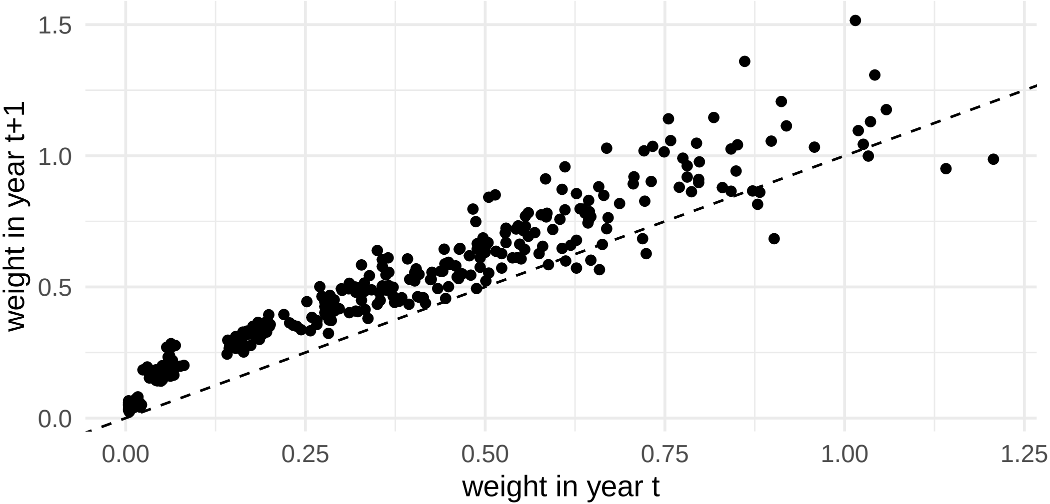
Topics today

- annual variation in growth increments
- scaling environmental covariates for model estimation
- linear mixed models of weight-at-age
- mechanistic models of length-at-age
- AICc for model selection

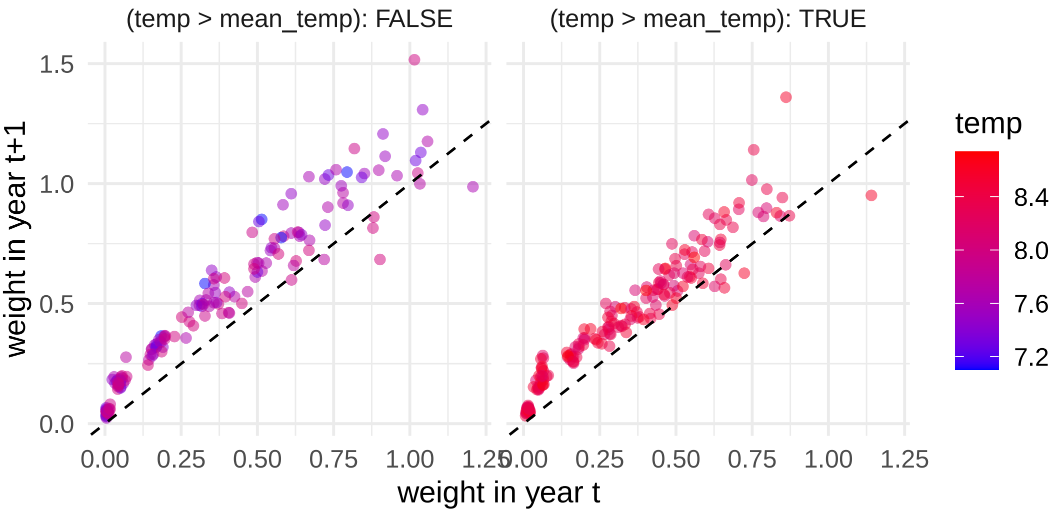
Weight-at-age is a component of productivity



Ford-Walford plots



Ford-Walford plots with extra dimensions



Modelling annual growth incremeмts

- the environment varies, so growth rates vary
- model L_t as a function of L_{t-1} or $L_{t-\Delta t}$ and the environment
- easier if you scale variables to the average environment (mean=0 and sd=1, Schielzeth 2010)
  - allows clear interpretation of 0 coefficient estimate => 0 effect
  - better for some estimation algorithyms
- we omitted the plus-age-group because it is a time-varying mixture of ages

Schielzeth, H., 2010. Simple means to improve the interpretability of regression coefficients. Methods Ecol. Evol. https://doi.org/10.1111/j.2041-210X.2010.00012.x

The simplest way is with a linear mixed model (LMM)

An overly simple formula could be

```
log(Wnext) ~ log(Wnow) + temperature + (1|year) + (1|cohort)
```

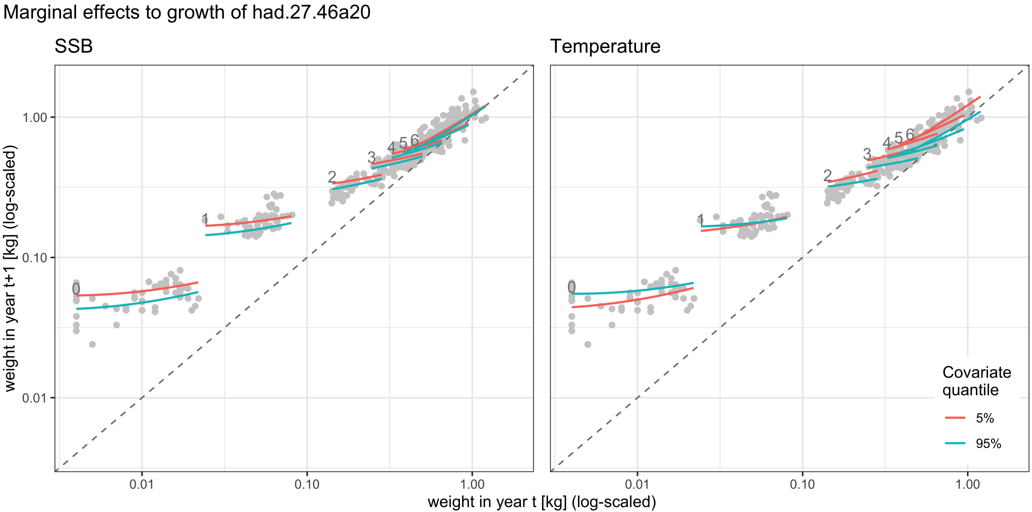
- "log-weight next year depends on log-weight now and the temperature this year"
- `(1|year)` means that the intercept varies by year (i.e. good and bad years)
- `(1|cohort)` means that the intercept varies by cohort (i.e. some cohorts have a lifelong advantage)
- assumes random year effects are 0 on average and have a normal distributions
- same for cohorts

For more flexibility, we could make the formula more complicated

```
logw2 ~ logw + age+ logw:age +  I(age^2) + I(logw^2) +
logw:I(age^2)+ sal.s + ssb.s + temp.s +I(temp.s^2) + logw:sal.s +
logw:ssb.s + logw:temp.s +(1|cohortf) + (1|yearf)
```

- `sal.s` salinity (mean=0, sd=1)
- `ssb.s` SSB (mean=0, sd=1)
- `temp.s` temperature (mean=0, sd=1)

A Ford-Walford plot with model predictions including age, SSB, and temperature



Simplify the model by doing model selection on the fixed effects

```
logw2 ~ logw + age+ logw:age +  I(age^2) + I(logw^2) +
logw:I(age^2)+ sal.s + ssb.s + temp.s +I(temp.s^2) + logw:sal.s +
logw:ssb.s + logw:temp.s + (1|cohortf) + (1|yearf)
```

Keep the random effects to avoid pseudoreplication (Hurlbert 1984).

- AIC should select the best predictive model (assuming all correlations are accounted for).
- BIC is sometimes used to get an even simpler model, but is intended for cases where you think the true process is represented by one of the models your selecting from.
- AICc is AIC corrected for small sample size.

Hurlbert (1984), Pseudoreplication and the Design of Ecological Field Experiments. Ecological Monographs, 54: 187-211. https://doi.org/10.2307/1942661

Select from all subsets of global model

- `MuMIn::dredge()` will do it automatically
- you choose the infomration criteria (AIC, AICc, BIC)
- `broom.mixed::tidy` will extract estimated coefficients with CI
- beware that CI will always be too narrow and p-values too low after model selection

See code in `LMM_had_example.R`

Mechanistic growth models

- converted weights to lengths
- modeled annual growth increments using
  - Gompertz $L_{t+\Delta t} = e^{log(L_\infty)(1-e^{-K\Delta t})+log(L_t)e^{-K\Delta t}}$
  - Specialised von Bertalanffy $L_{t+\Delta t} = L_t + (L_\infty - L_t)(1 - e^{(-K\Delta t)})$
  - Generalised von Bertalanffy $L_{t+\Delta t} = ((L_\infty)^{1/D}(1 - e^{-K\Delta t}) + L_t^{1/D} e^{-K\Delta t})^D$
- in each of the above $L_\infty = exp(X\beta + \varepsilon_{cohort} + \alpha_{year})$
- $X\beta$ selected from global model formula `~1 + ssb.s + sal.s + temp.s` via AICc
- ε_{cohort} and α_{year} are normal random intercepts with mean 0
- $L_{t,observed} \sim N(L_{t,predicted}, \sigma_{residual})$
- converted predicted lengths back to weights

Example stock "had.27.46a20"

Generalised von Bertalanffy $L_{t+\Delta t} = ((L_\infty)^{1/D}(1 - e^{-K\Delta t}) + L_t^{1/D} e^{-K\Delta t})^D$

See code in `laaGrowth_had_example.R`

Predictions from a lognormal model

- If $x \sim N(\mu, \sigma^2)$ then e^x has a lognormal distribution
- with expectation $e^{(\mu+\sigma^2/2)}$
- This model `glmmTMB(logw2 ~ logw + age + sal.s + ssb.s + temp.s+ (1|cohortf) + (1|yearf), data=dat)` has 3 components to the variance $\sigma^2_{total} = \sigma^2_{cohort} + \sigma^2_{year} + \sigma^2_{residual}$

See code in `LMM_had_example.R` line 64

The mechanistic model code doesn't require this step because it fits the response on the natural scale.

Redo the analyses and discuss in subgroups

- Use "sol.27.7a.Rdata"
- Look closely at the AICc table from LMMs.
- What environmental covariates are in most of the top models?
- What does the top model say (with overly-narrow CI)?

The bottom line

Of the 27 stocks that I applied these methods to, for 70% of them, a 3-year average of the most recent observations did a better job of predicting weight-at-age than any model I could come up with.

You'll see some of the forecasting tests in Marc's lessons later on.

Acknowledgments