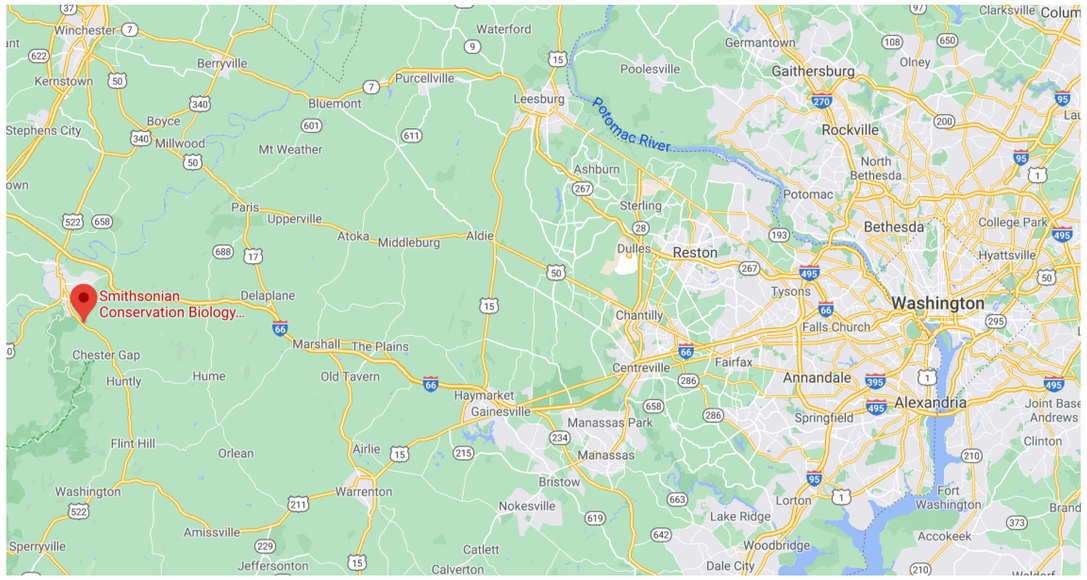
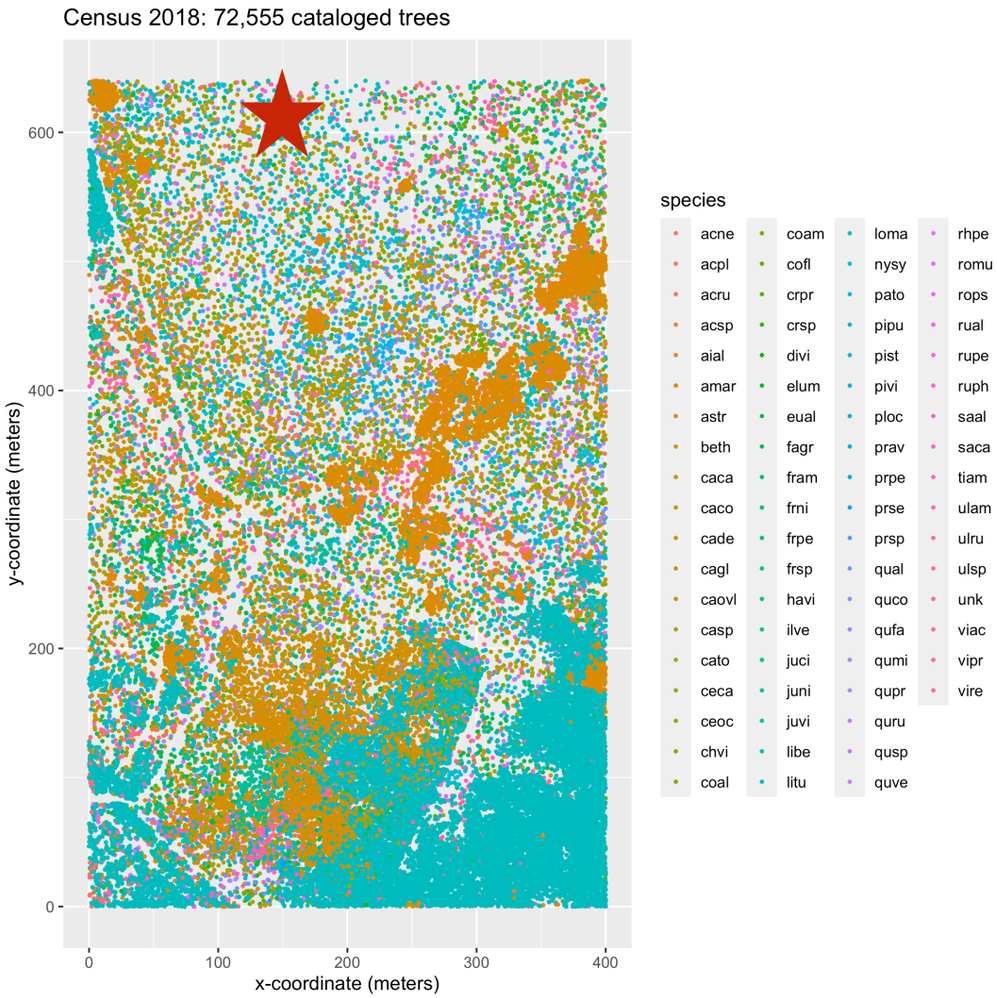
Fusing disparate measurement data for forecasting the growth of trees via Hidden Markov Models

Prof. Albert Y. Kim akim04@smith.edu
US Forest Service
Wednesday, January 12 2022

One particular tulip poplar at SCBI

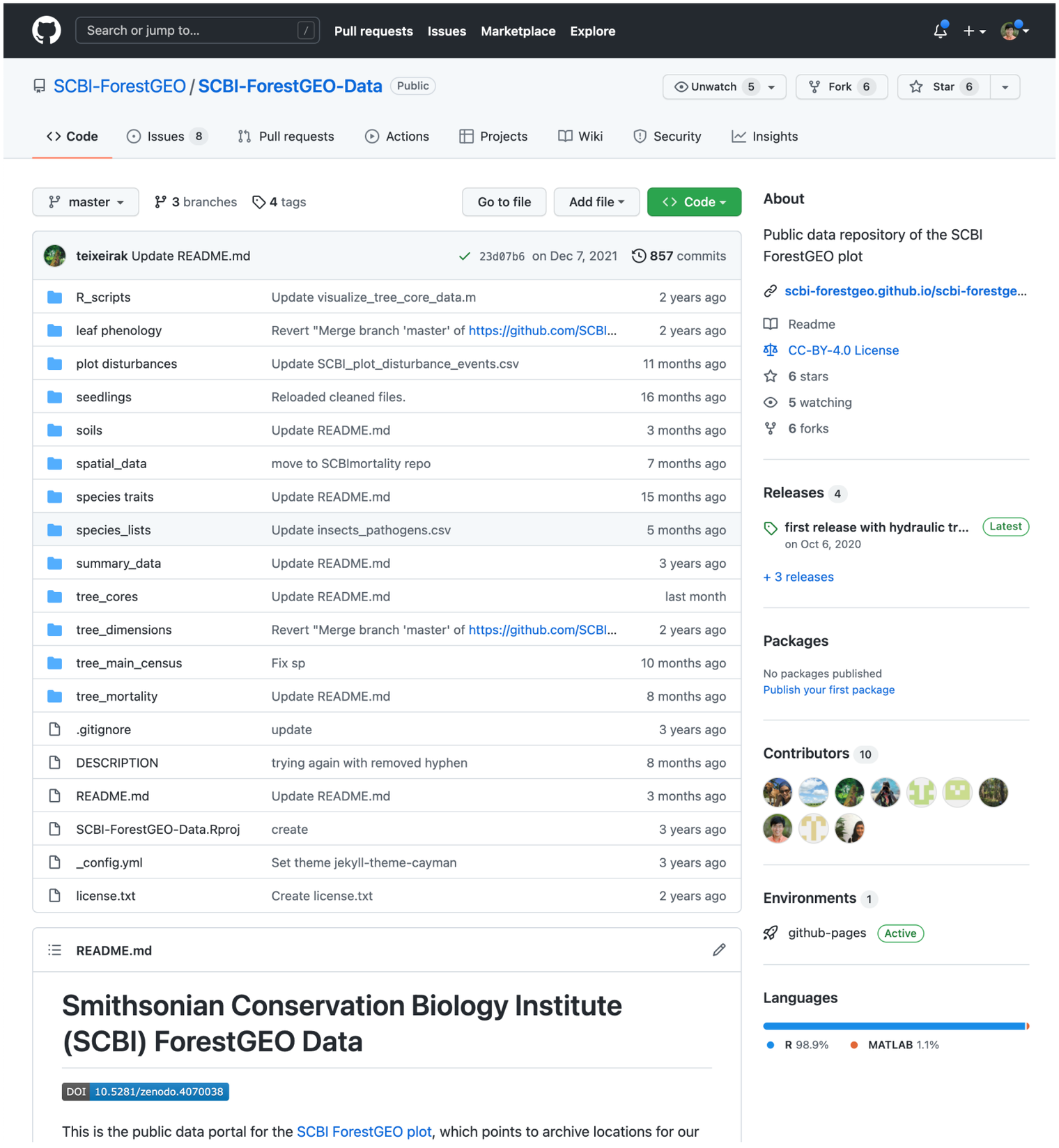


25.6 ha = 35.85 soccer fields



Tag 082422

Data on GitHub

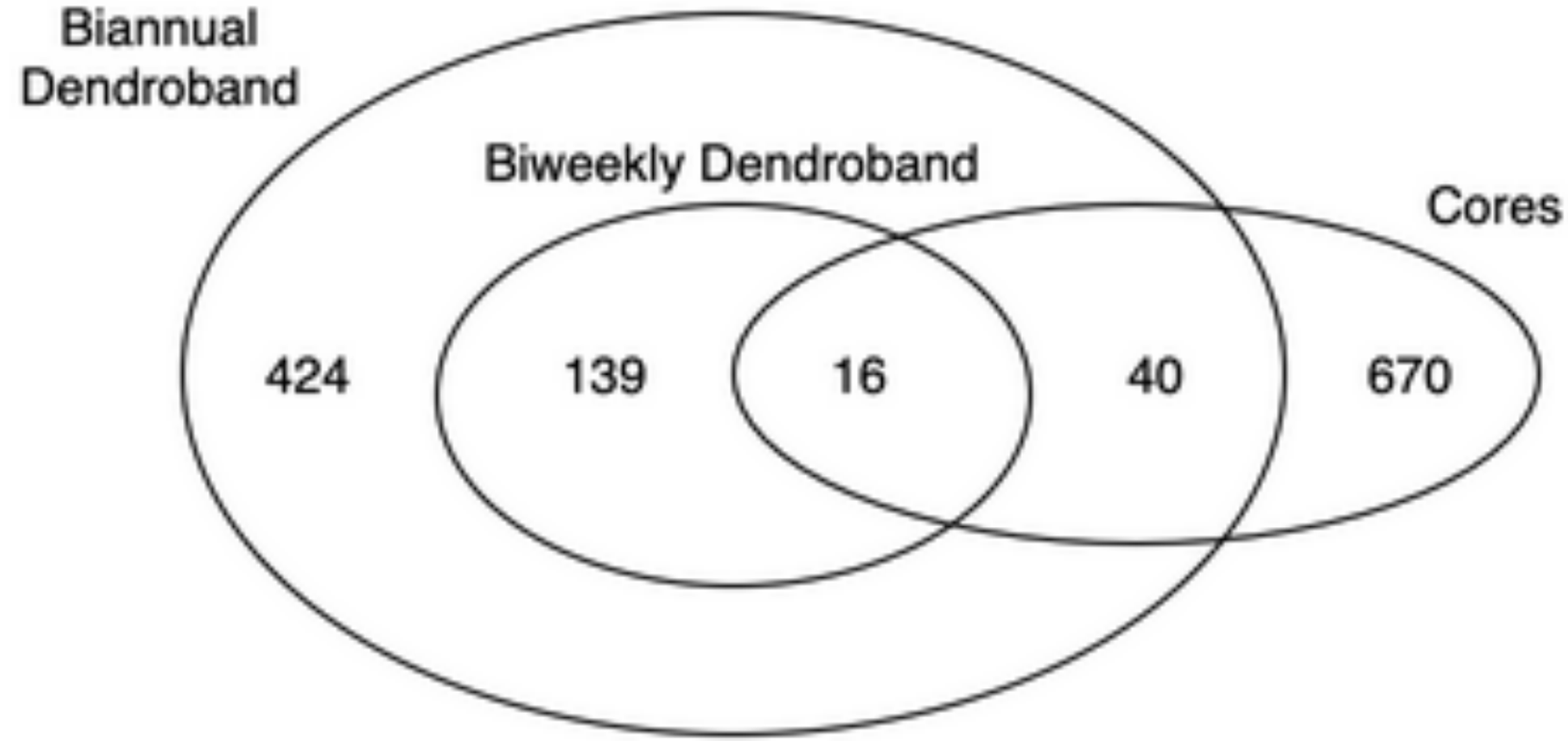


Comparison Chart

	Data source	Measurement	Cost	Sources of Error?
	**Census via tape**	Diameter	Cheap	Large variation in dbh 📏 technique
	**Tree coring**	Ring width increment	Expensive	Standardized, cores are dried, no bark effects
	**Intraannual dendroband (every 2 weeks)**	Diameter (baseline + gap size)	High setup, rapid follow-up	Climate induced variation in bark & device (-'ve growth)
	**Biannual dendroband (start & end of year)**	Diameter (baseline + gap size)	High setup, rapid follow-up	

Sample sizes

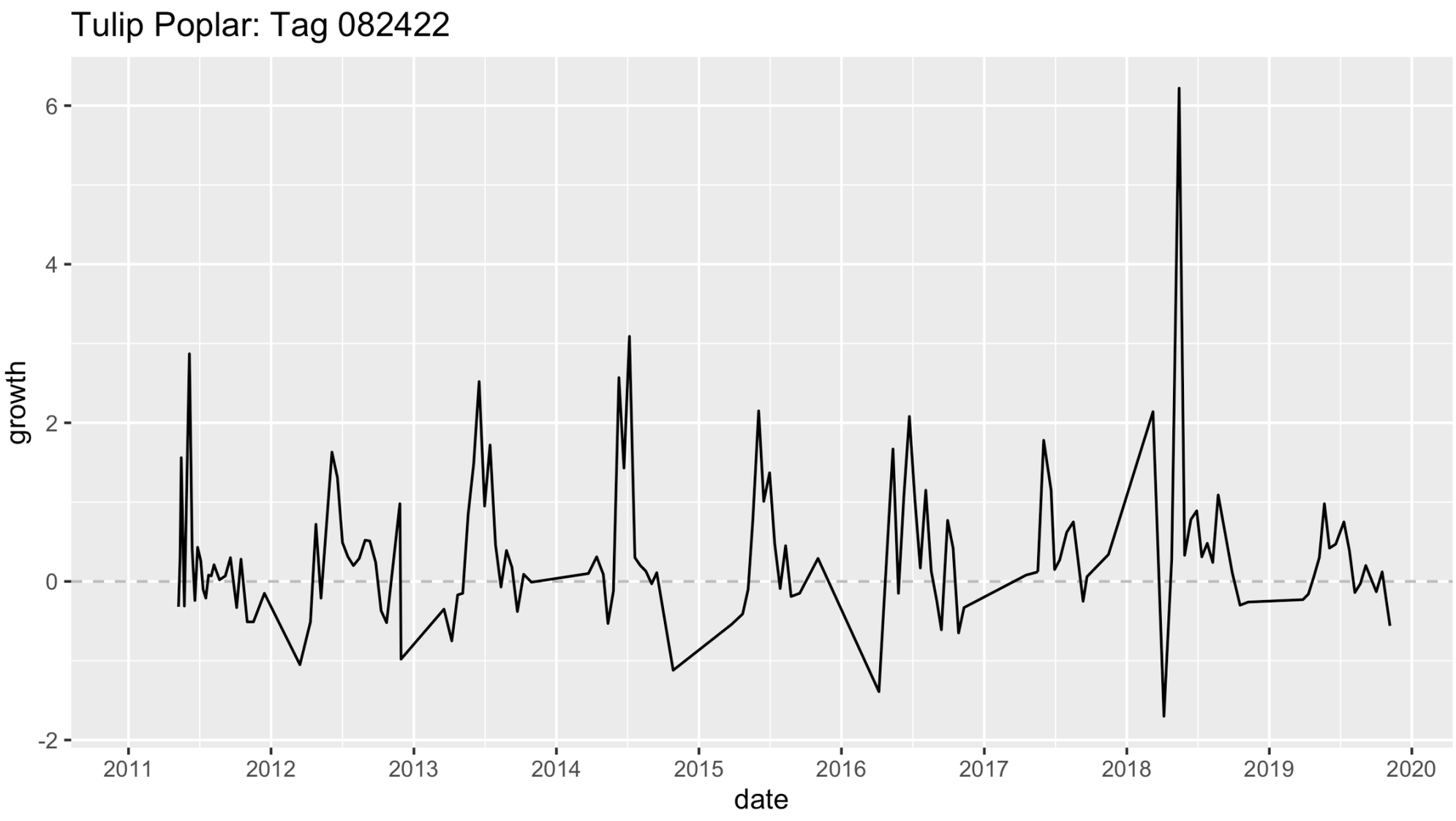
1289 stems total (all have census observations)



3 trees had 3 dendrobanded stems & 8 trees had 2 dendrobanded stems;
we assume these stems are statistically independent

Because dendroband measurements…

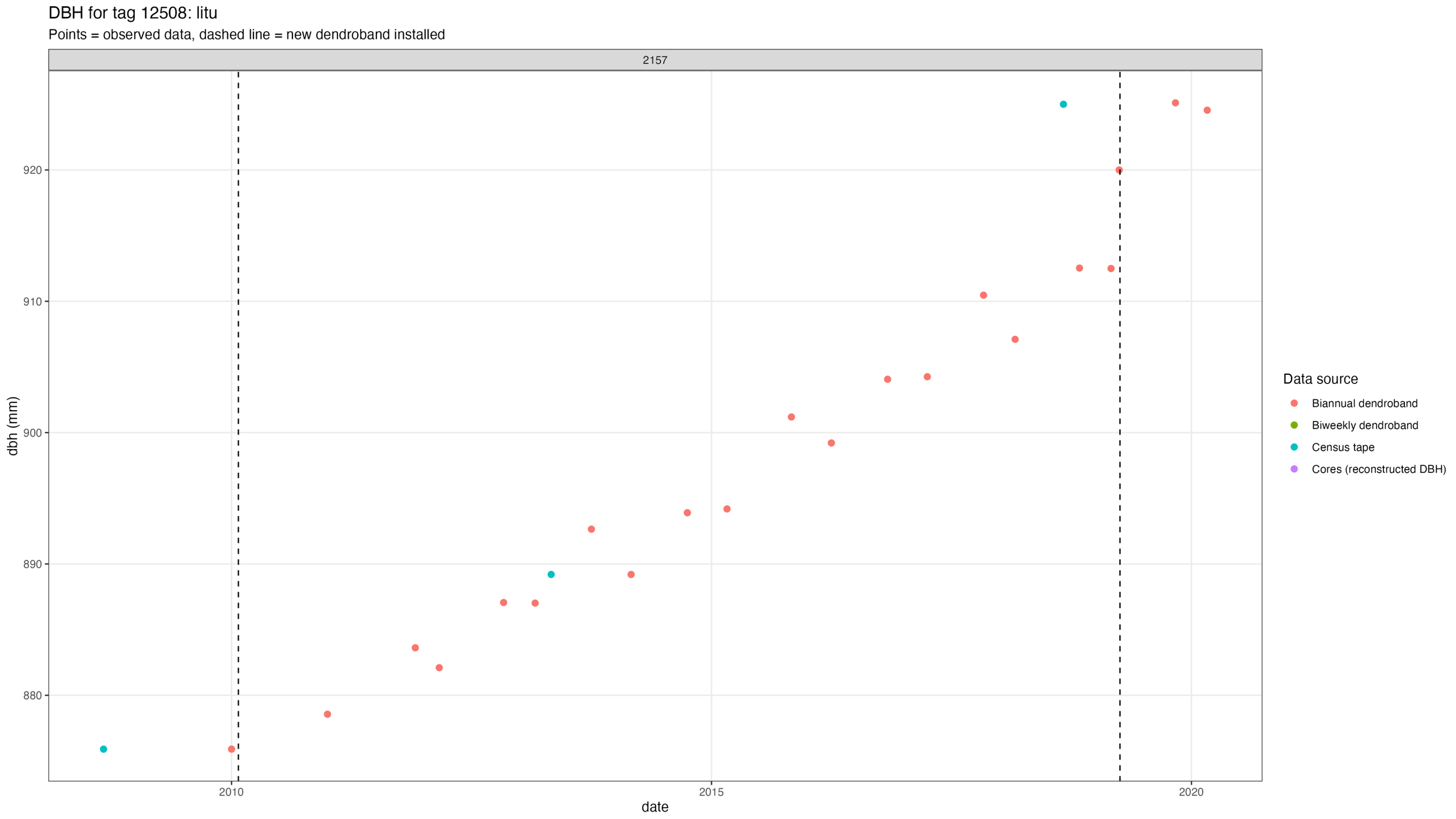
… are taken many times within-year



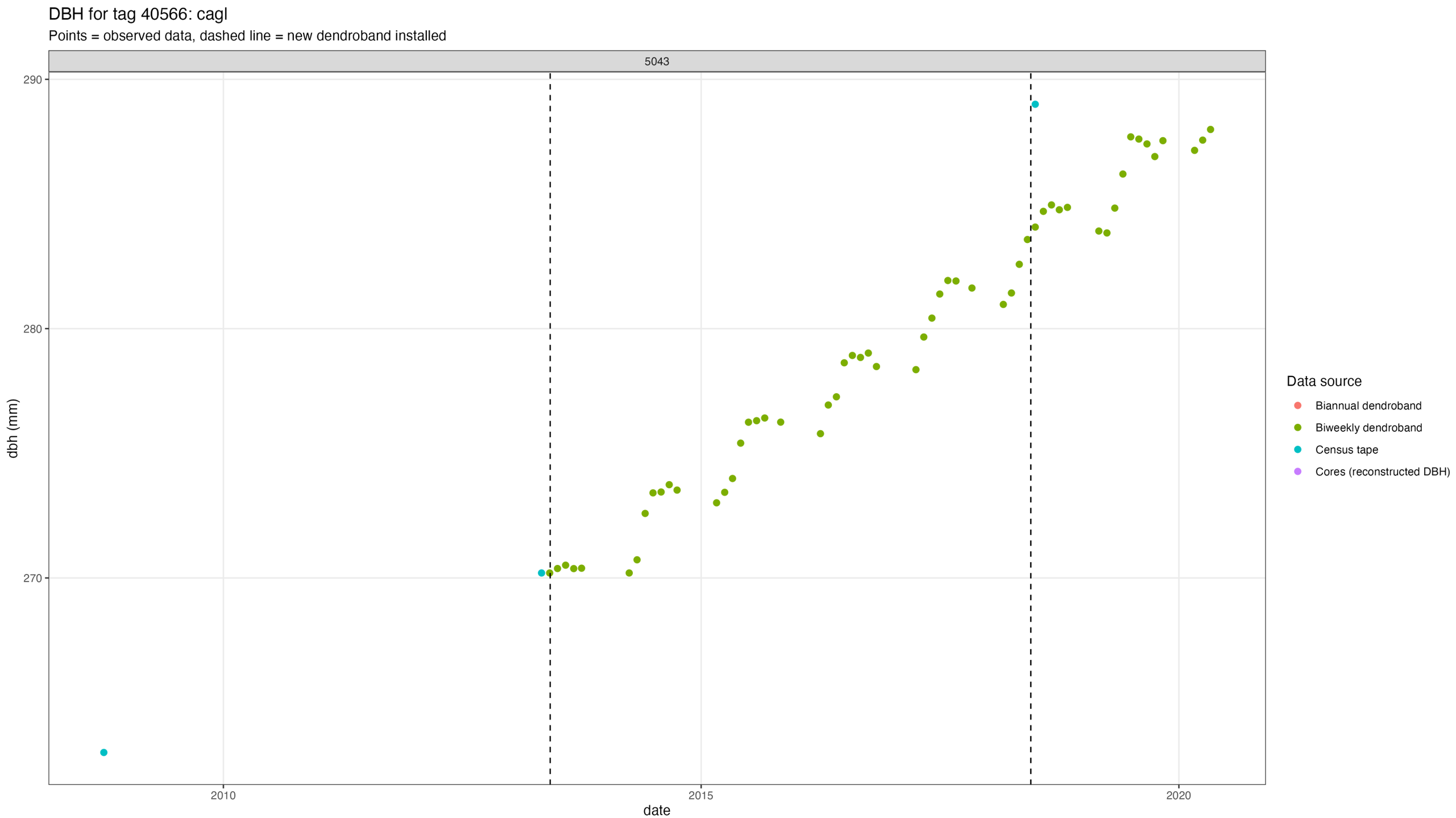
… we can link within-year variation in growth to within-year variation in climate!

Our Data

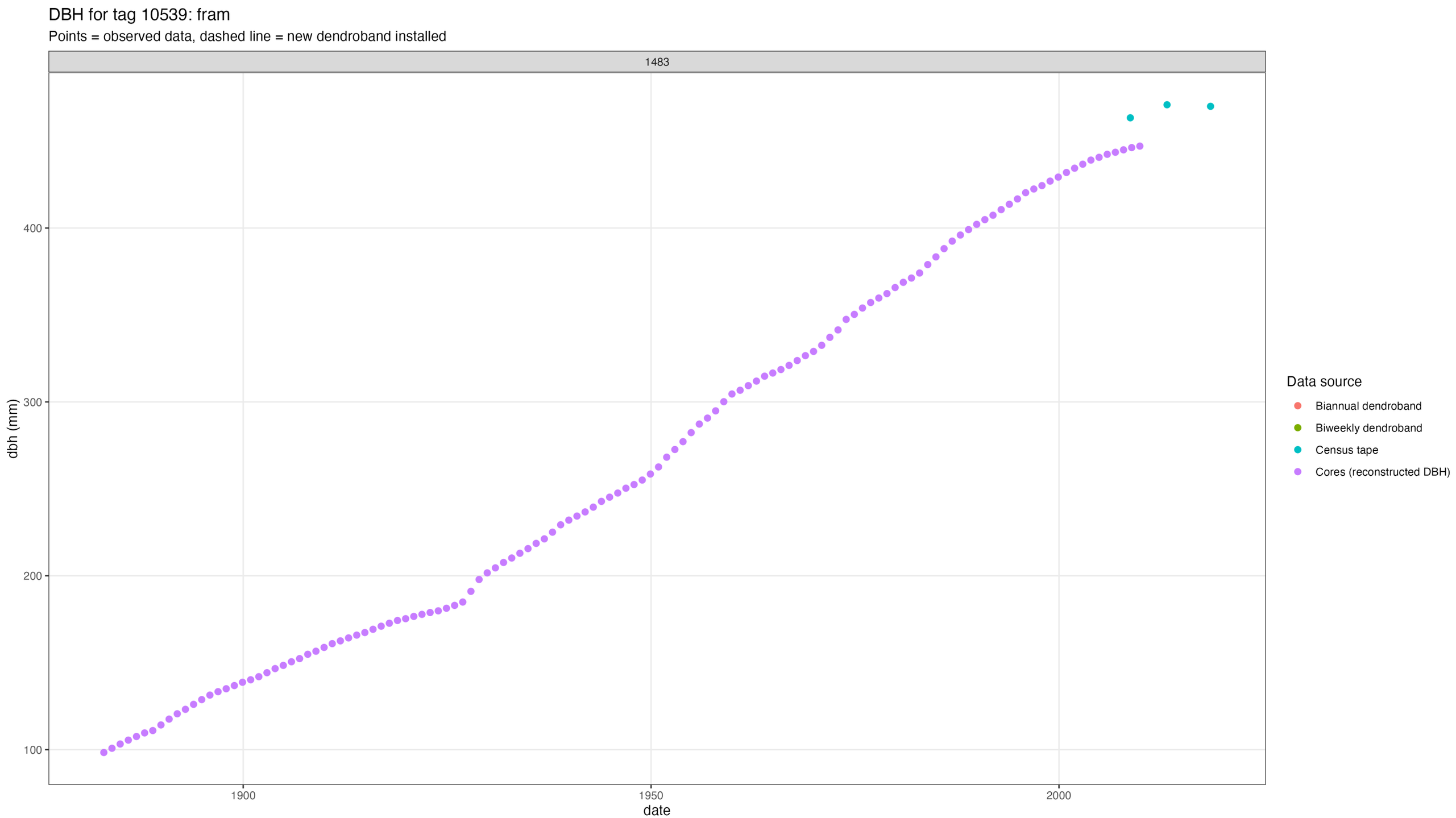
dbh trajectories: biannual dendro only



dbh trajectories: biweekly



dbh trajectories: cores (as dbh, not increments)



Goal 🎯🎯🎯

Can we fuse these disparate data sources into a single model to forecast the growth of trees?

Hidden Markov Models

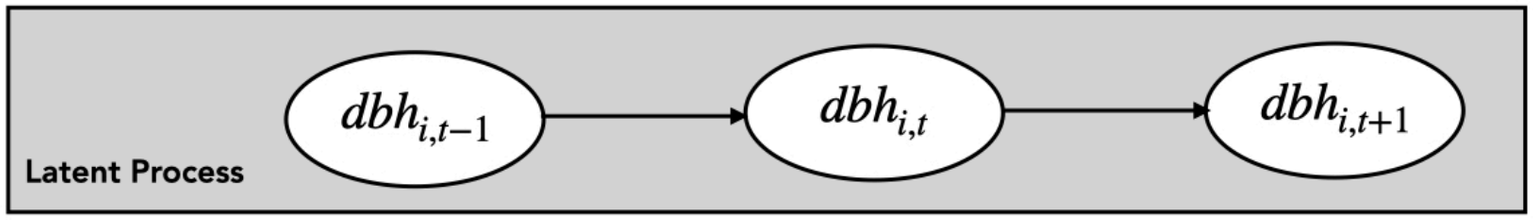
- Hidden: "Data fusion" via latent variables
  In our case: "true" dbh
- Main "process model" driving y_t
- Markov property: y_t depends on y_{t-1}
- Partition sources of error into those that
  - Are not of direct interest
  - Are "one and done" i.e. measurement error
  - Propagate when forecasting

Other implementations

- Clark 2007
- Schleip 2014
- Dawson 2021
- Heilman (forthcoming)

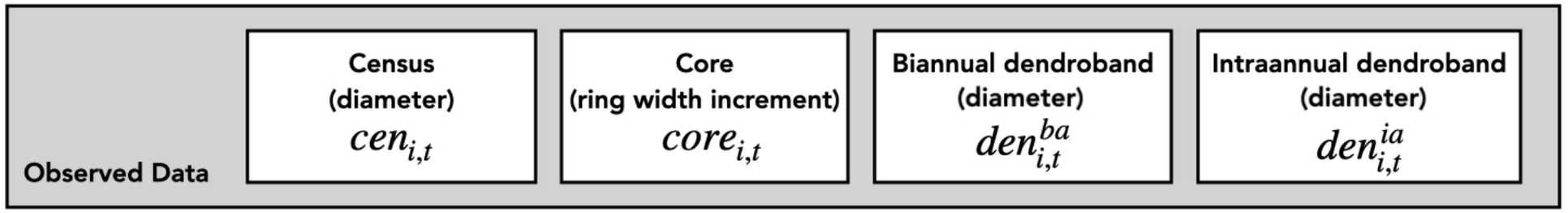
Spirit of Hidden Markov Model

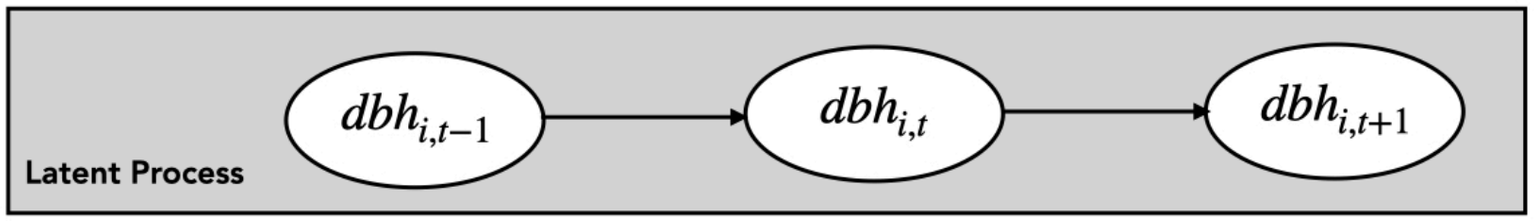
Latent process model



Clark 2007

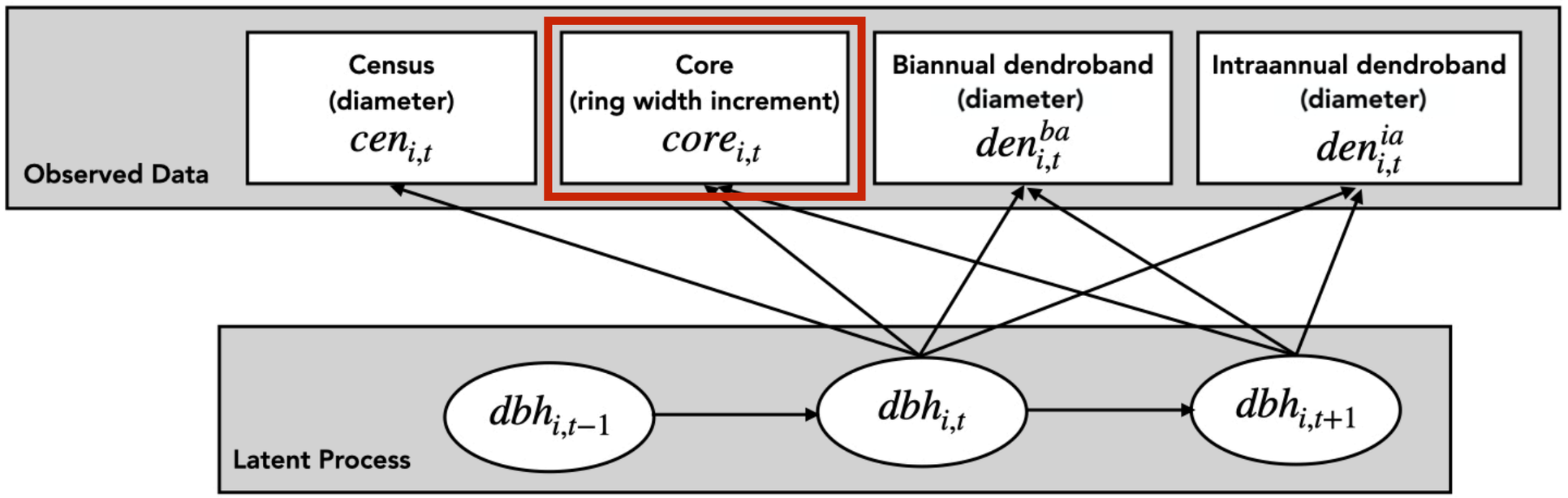
+ field collected data



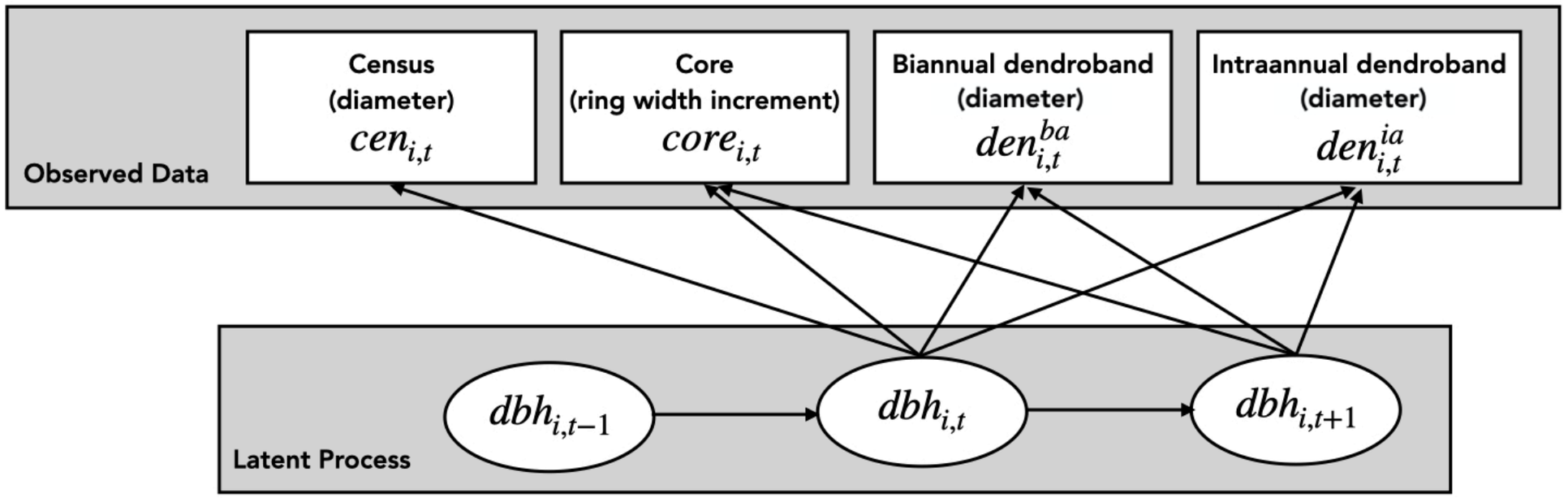


+ field collected data

Increments = $dbh_{i,t+1} - dbh_{i,t}$

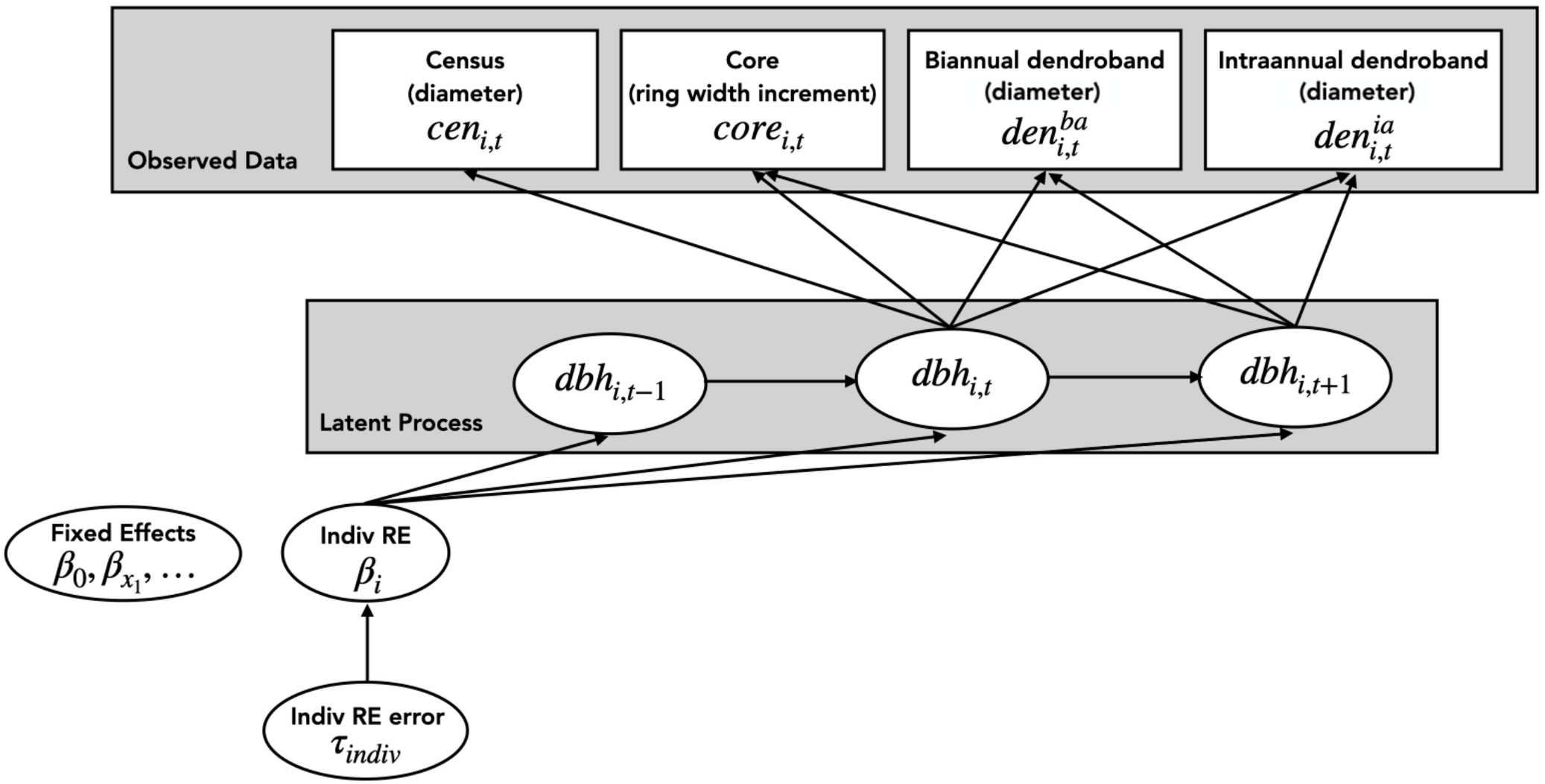


+ fixed effects (Ex: climate)

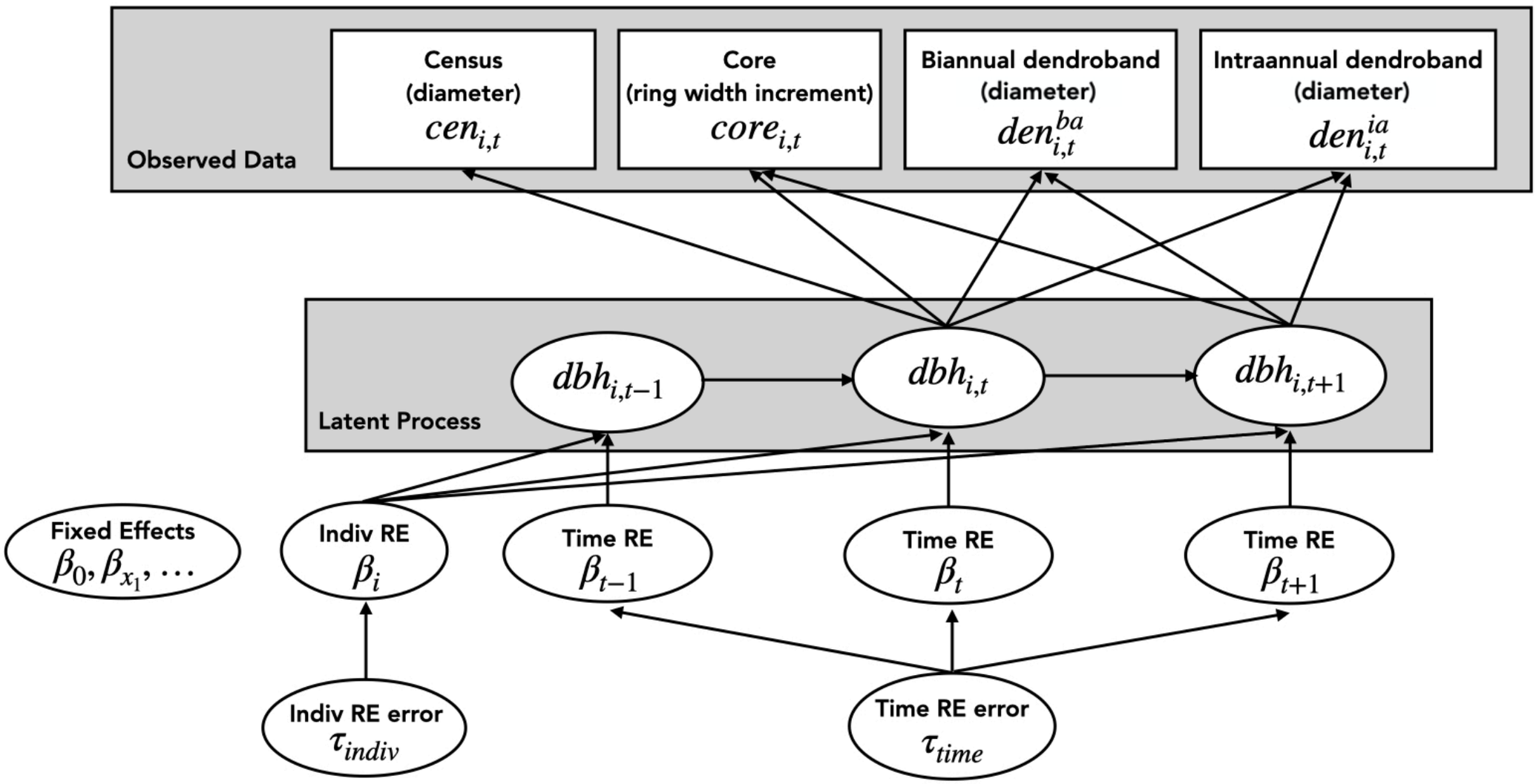


Fixed Effects
$\beta_0, \beta_{x_1}, \ldots$

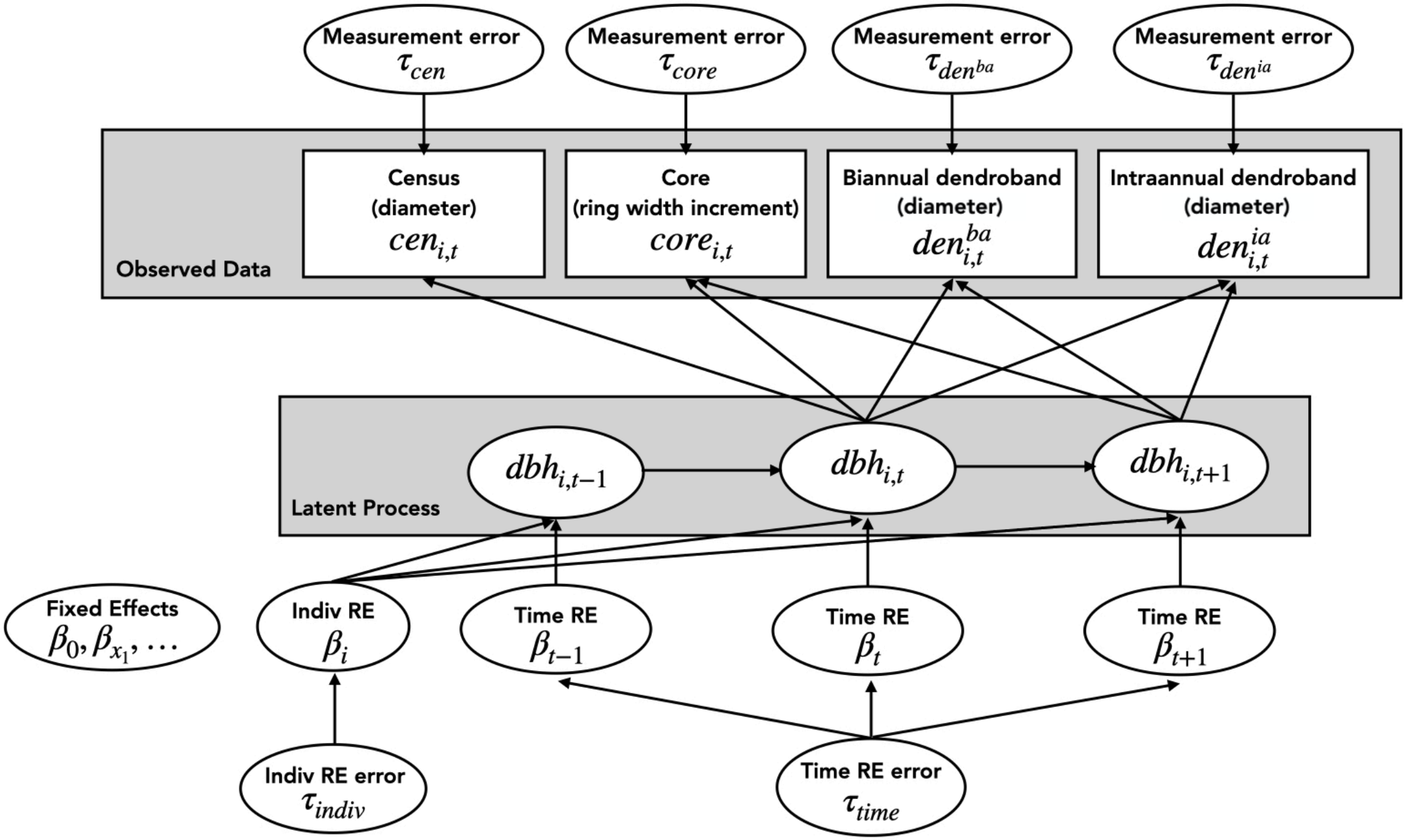
+ random effects for individual trees



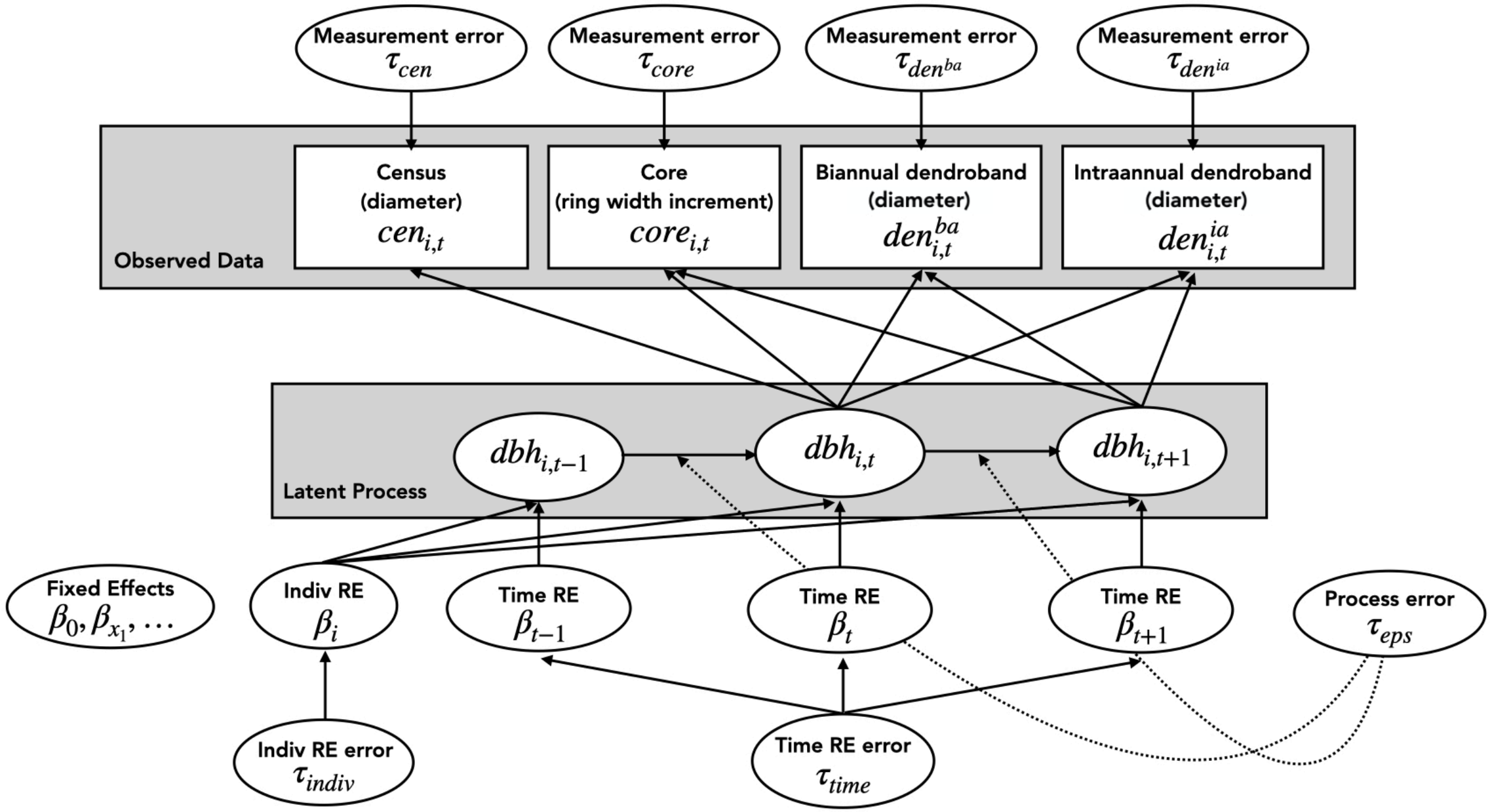
+ random effects for time points



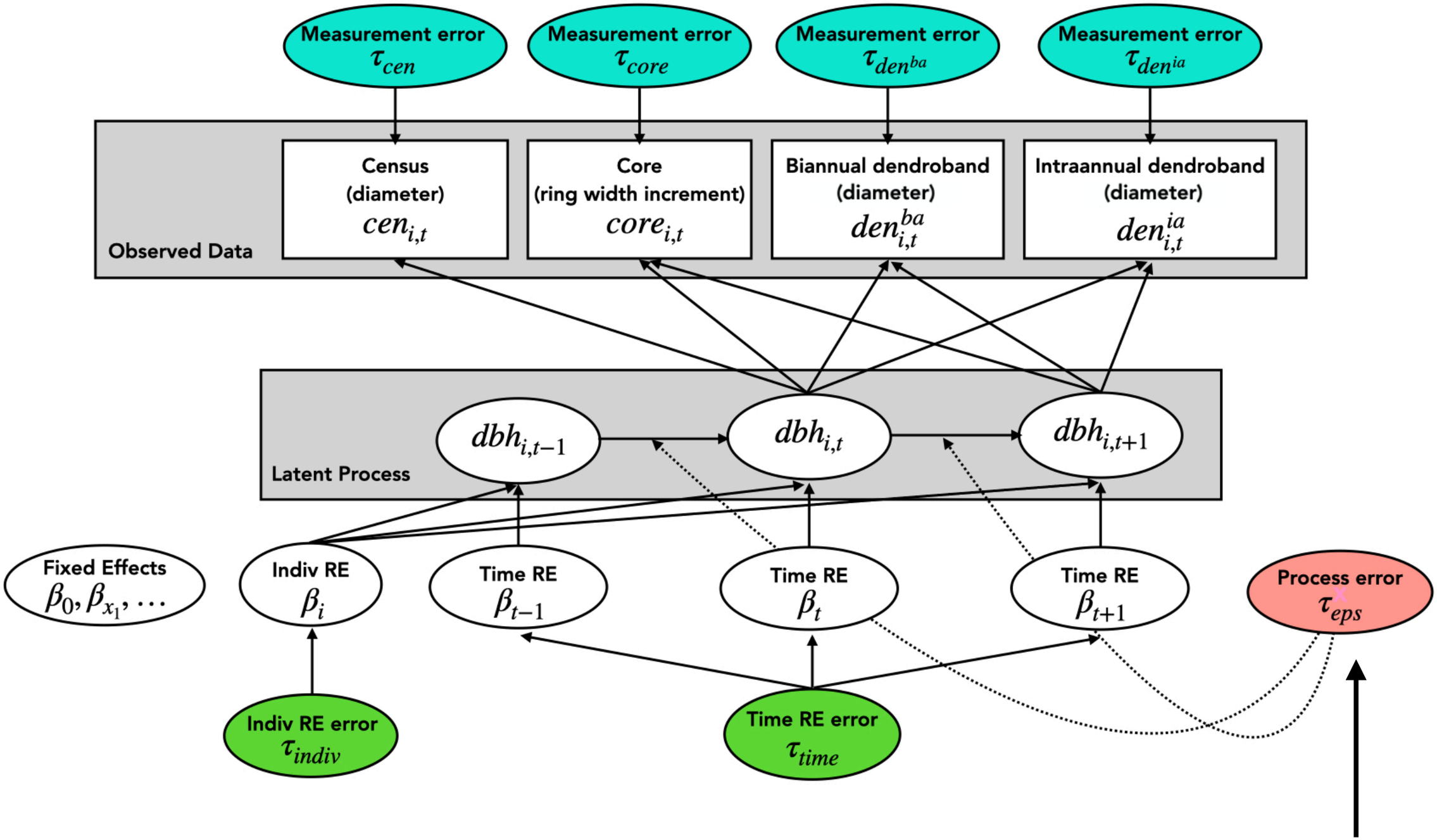
+ measurement error



+ process error



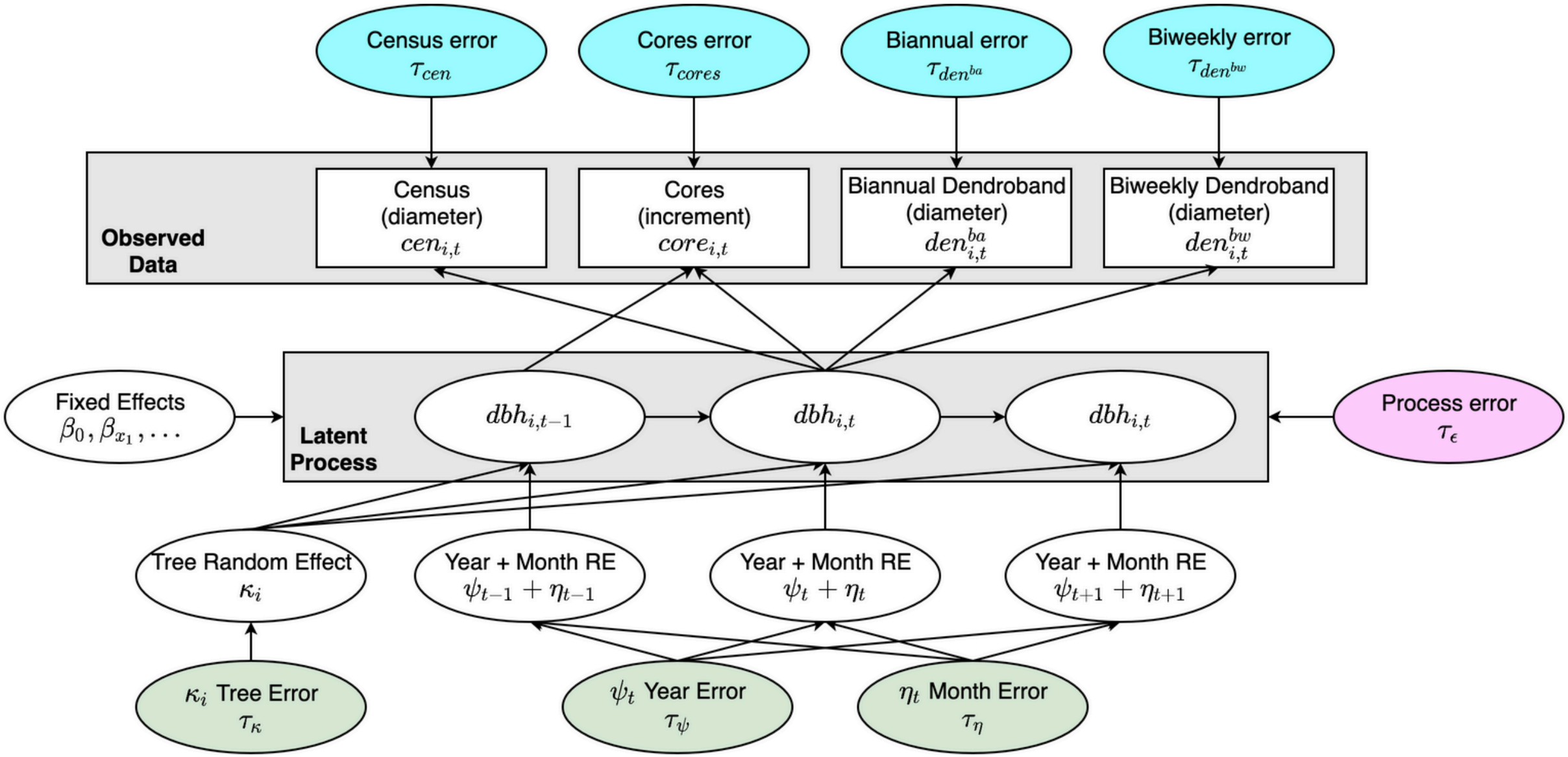
= Hidden Markov Model



Moral: only error this propagates across time in forecasts

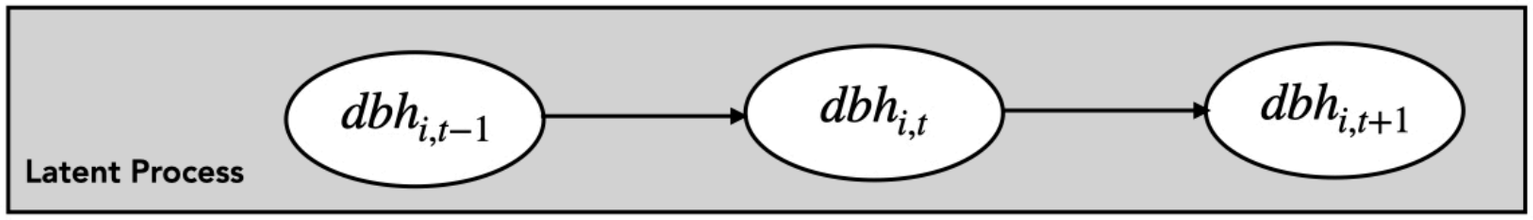
Actual Model

Model as of 2021/8/20



Made with draw.io

Need to specify latent process model…



Hidden Markov latent process model

$$dbh_{i,t} - dbh_{i,t-1} = \alpha + \beta \cdot dbh_{i,t-1} + \kappa_i + \psi_t + \eta_t + \epsilon_{i,t}$$

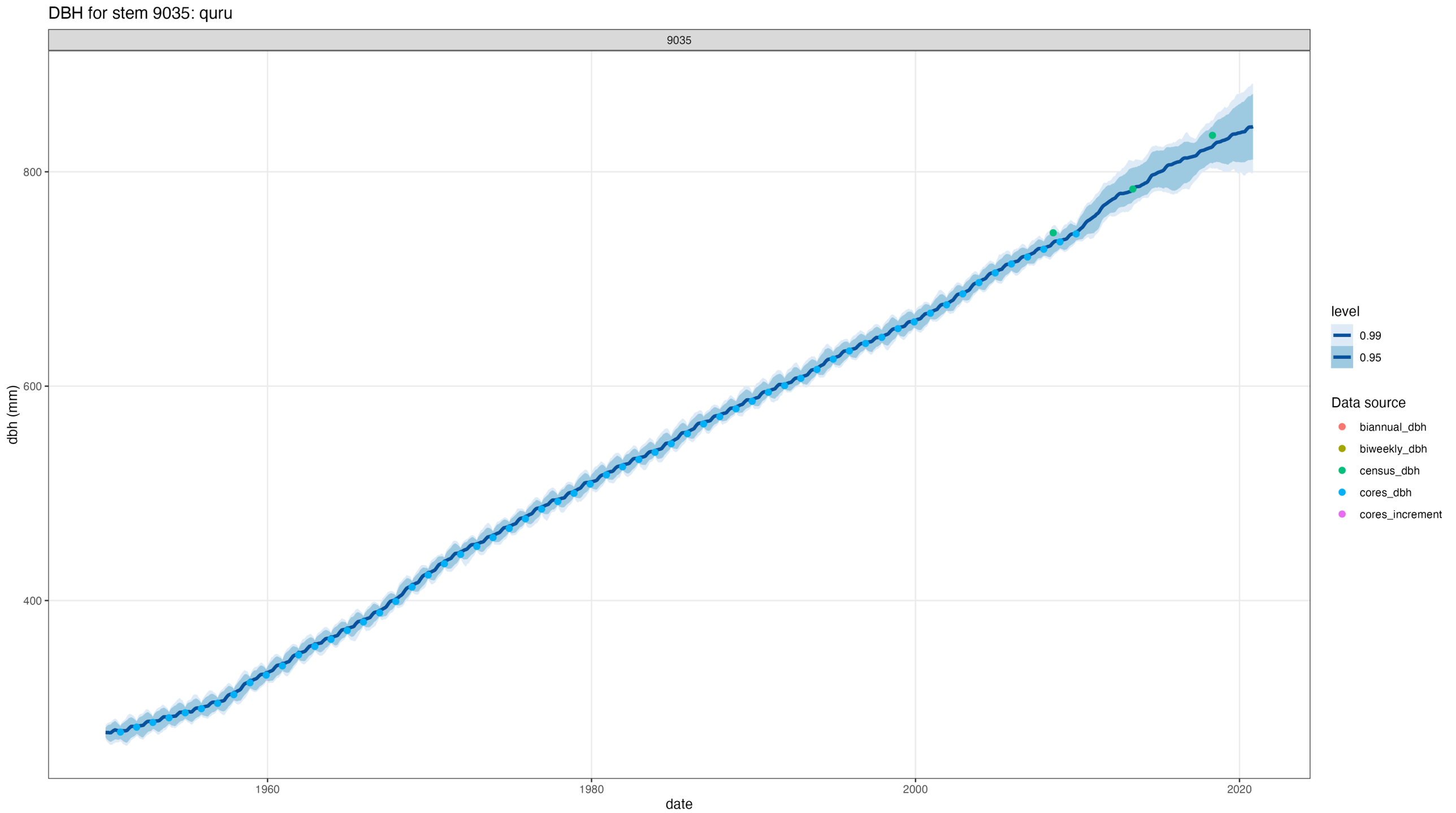
- $dbh_{i,t}$: "True" latent dbh for individual i at time t
- α : Baseline growth
- β : Slope for DBH
- κ_i : Individual tree i random effect
- ψ_t and η_t : Year & month time t random effects
- $\epsilon_{i,t} \sim$ Normal $\left(0, \sigma^2_\epsilon\right)$

Results

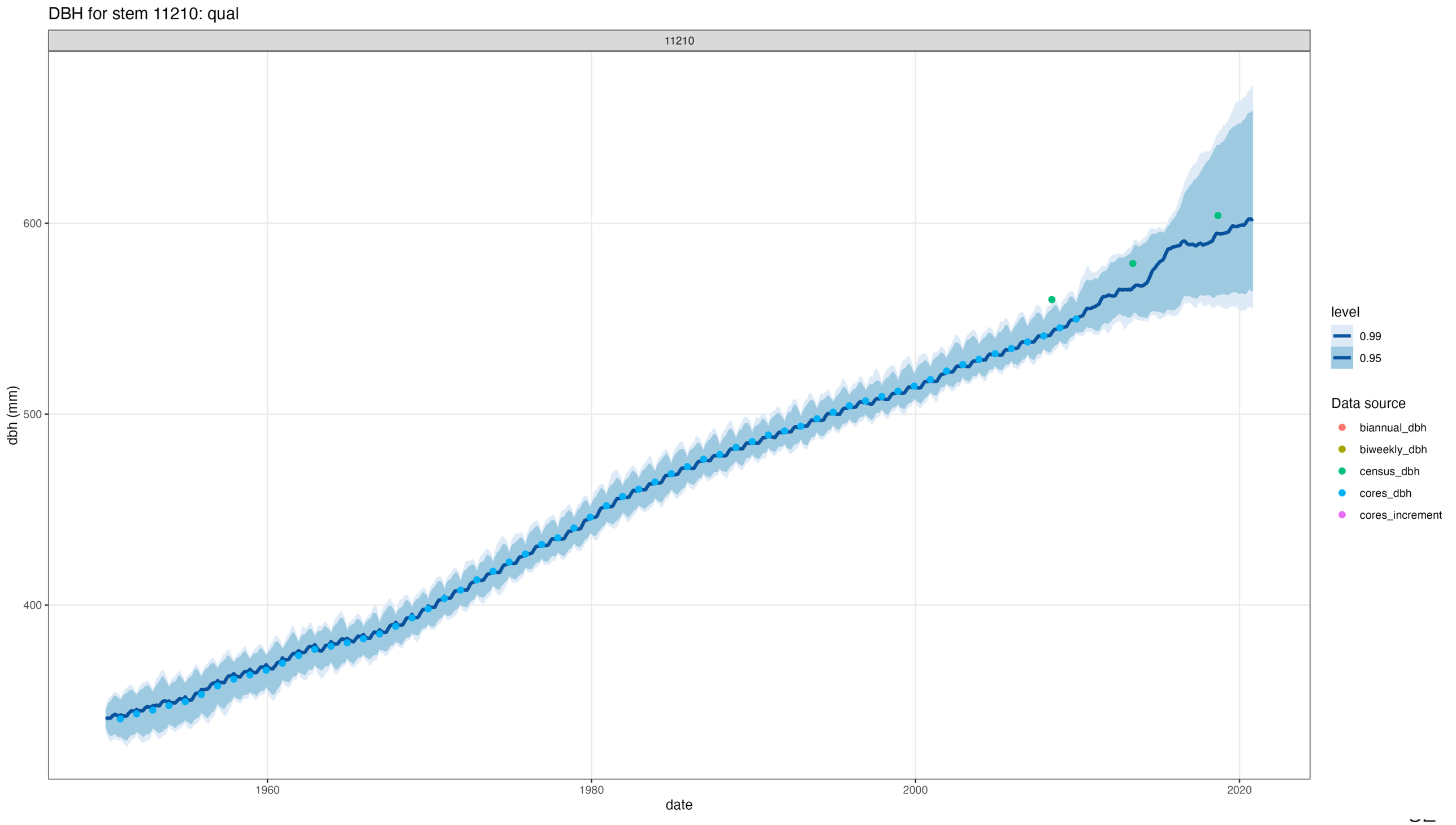
MCMC specifications

- Only on random sample of 100 out of 1288 stems with core, dendro biweekly, or dendro biannual data only for values of cores > 1950
- 3 chains of 1k draws from posterior minus 10% burn-in
- 10min run-time
- Following results should be taken with a massive grain of salt 🤔❌🧂

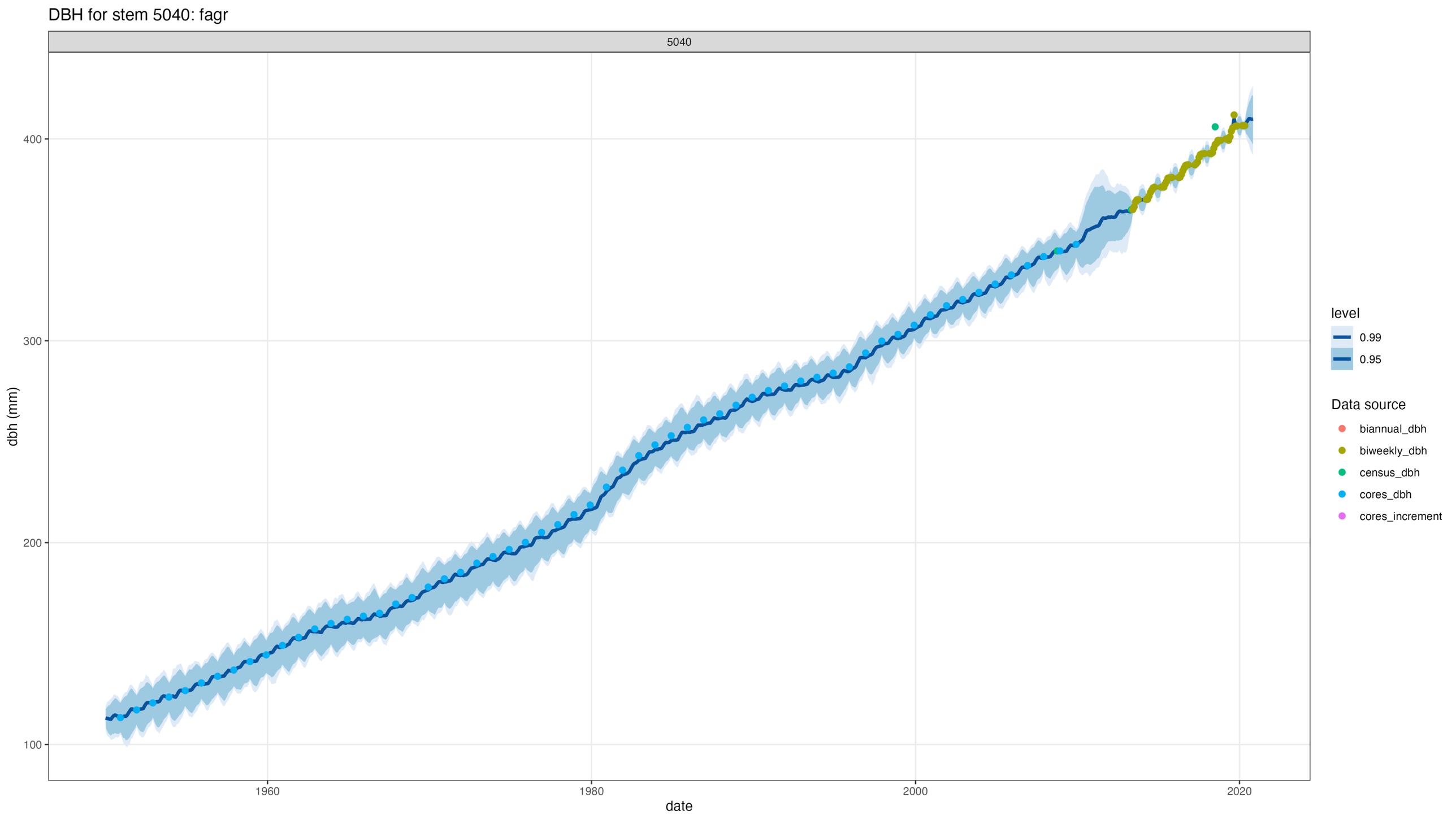
dbh trajectories



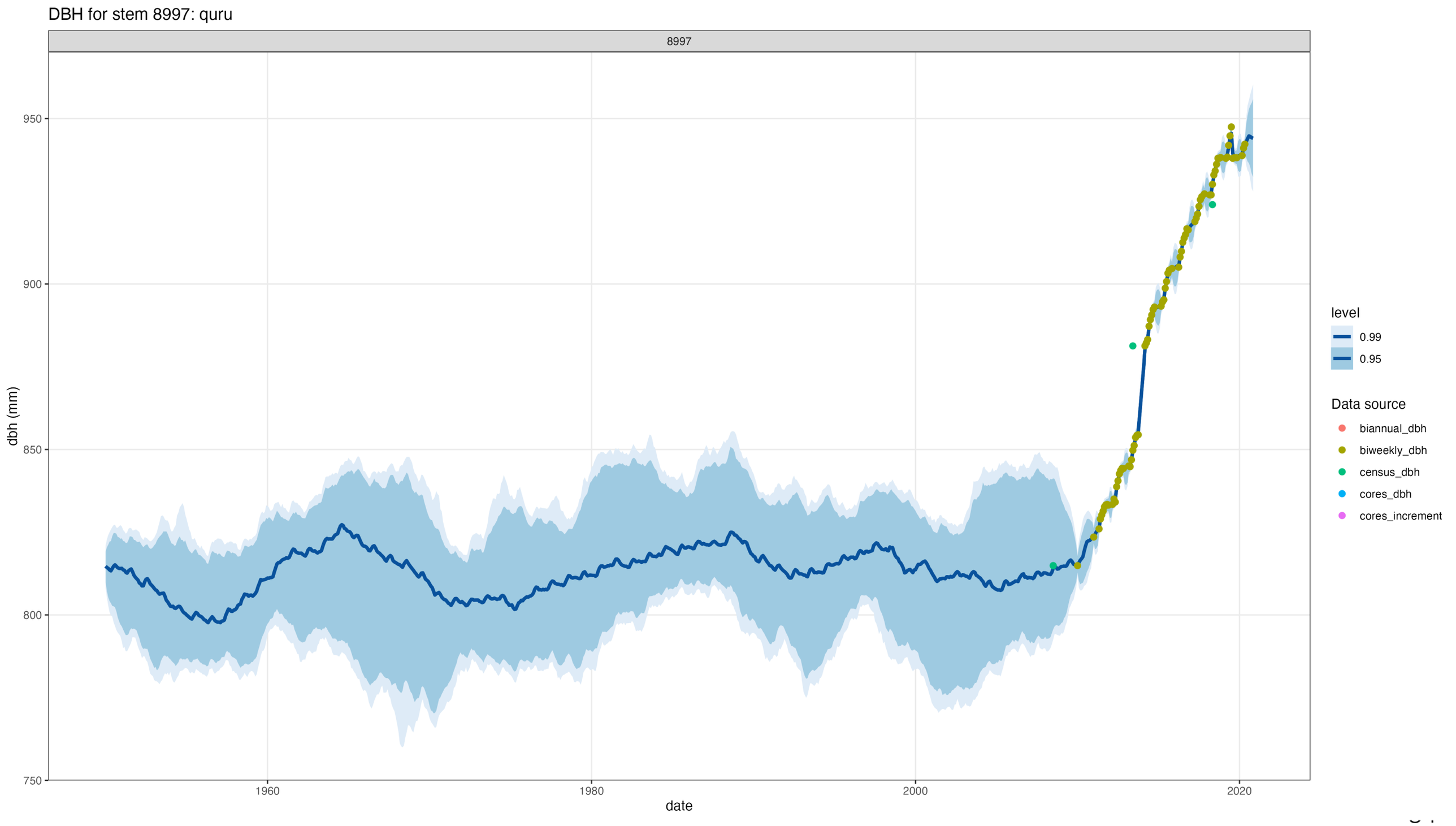
dbh trajectories



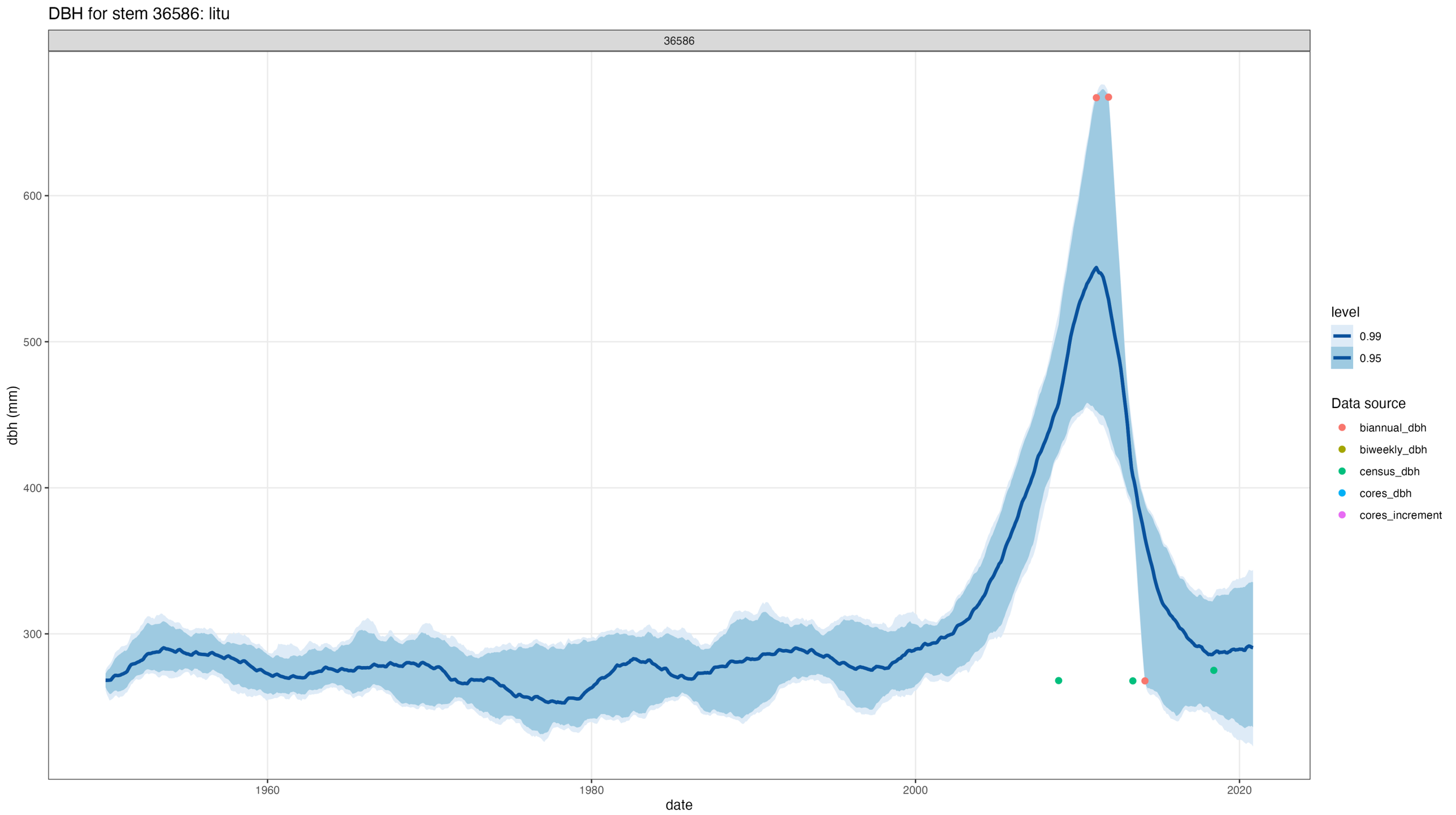
dbh trajectories



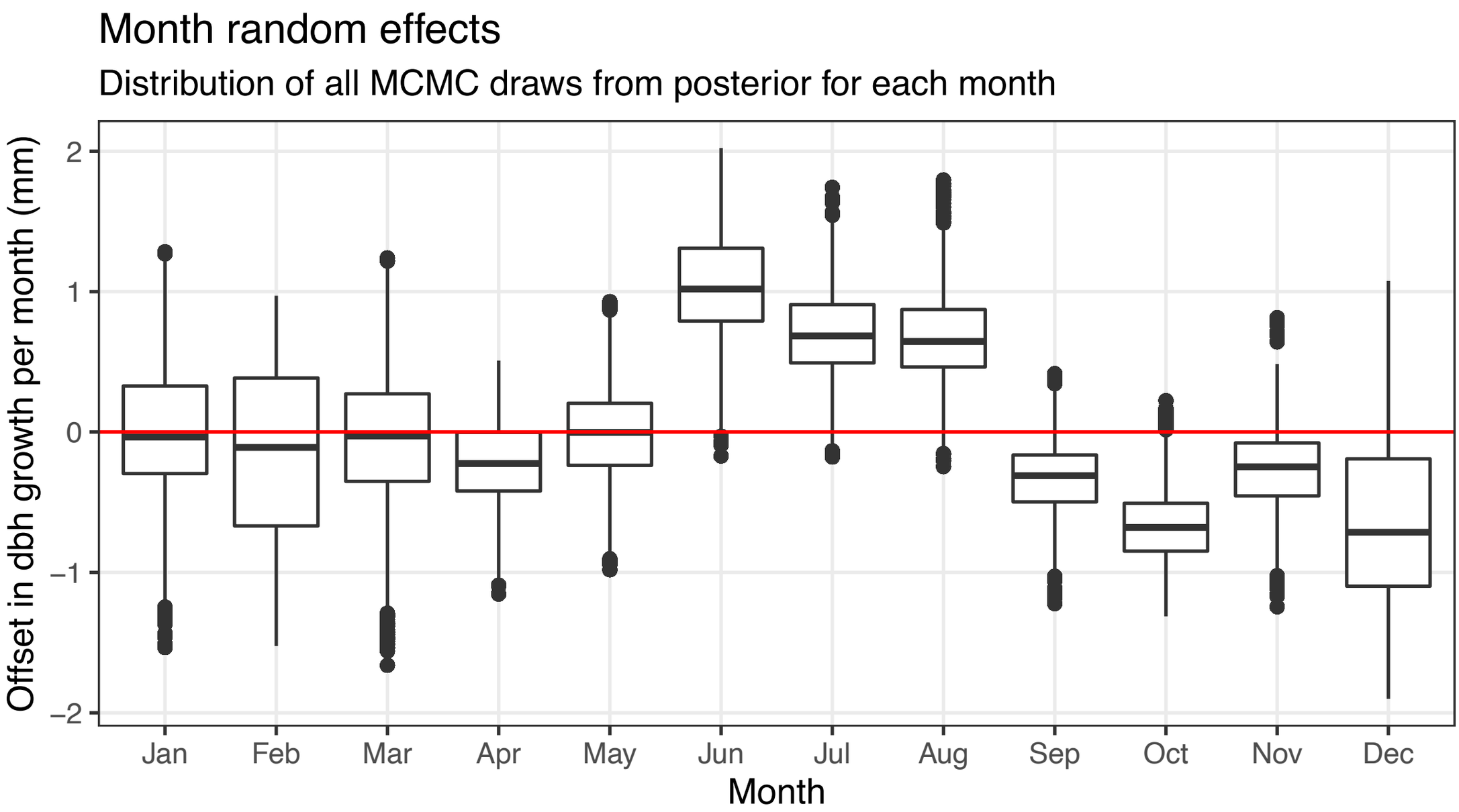
dbh trajectories



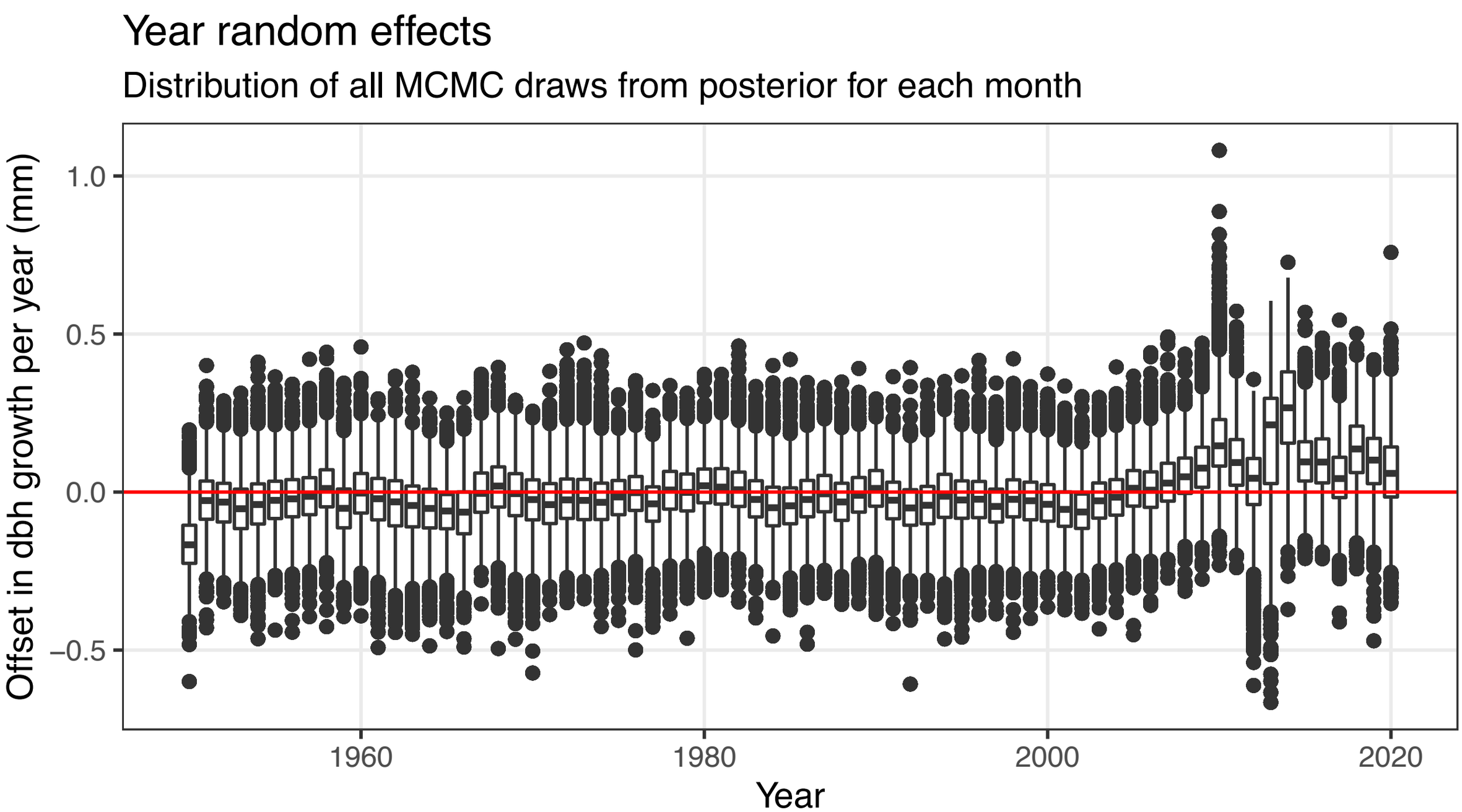
dbh trajectories



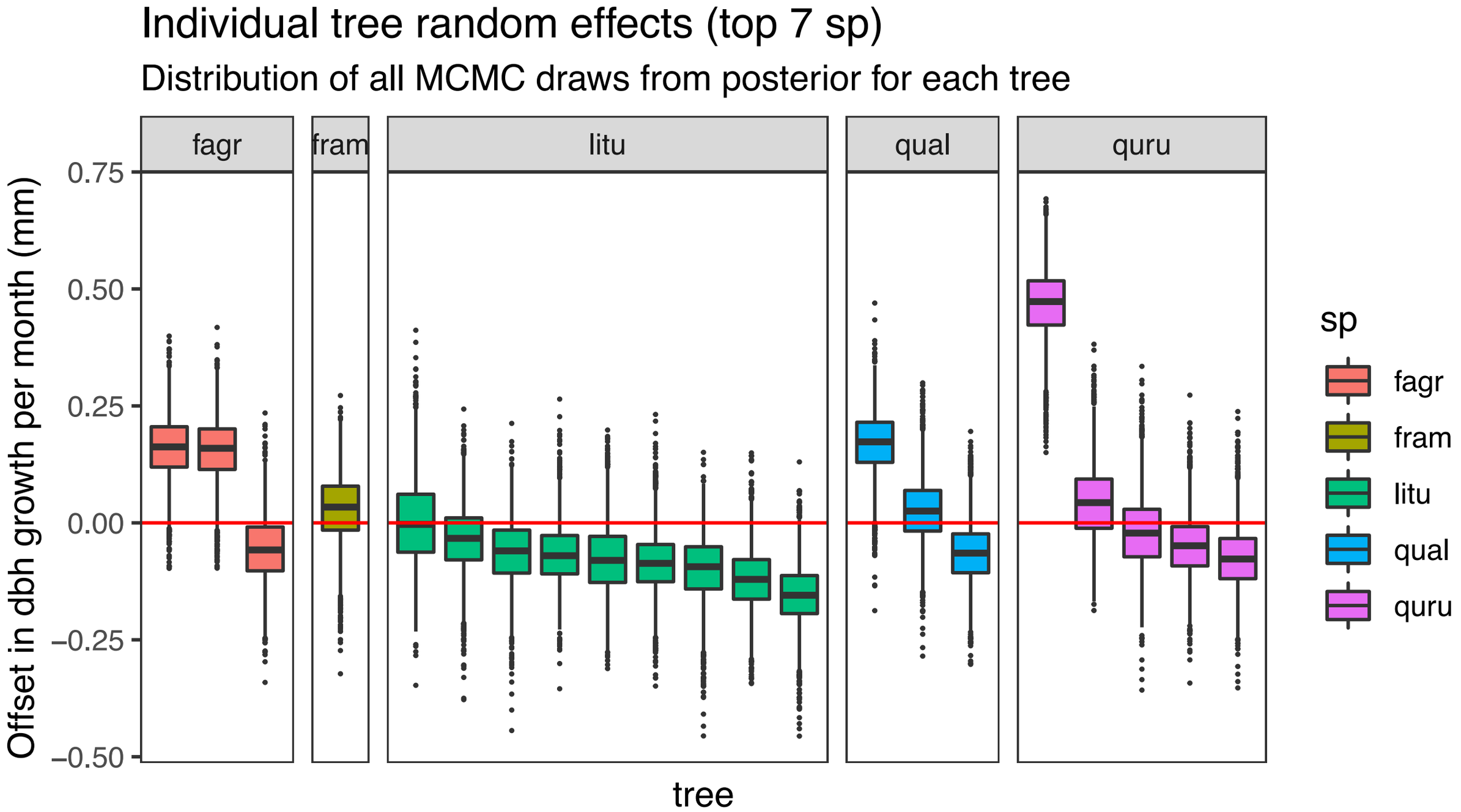
Posteriors: Month Random Effects



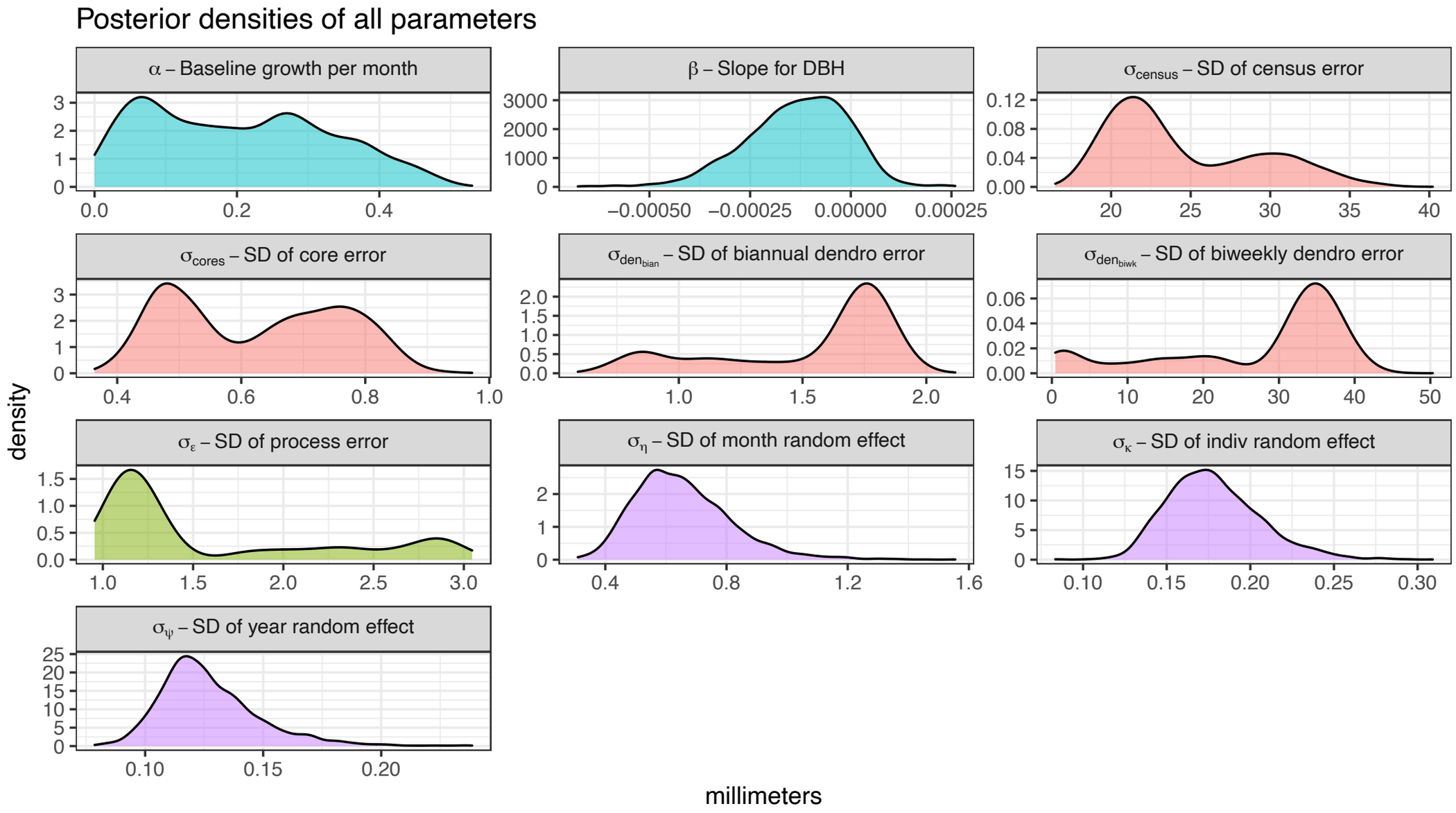
Posteriors: Year Random Effects



Posteriors: Individual Random Effects



Posteriors: Model parameters 🤔❌🧂



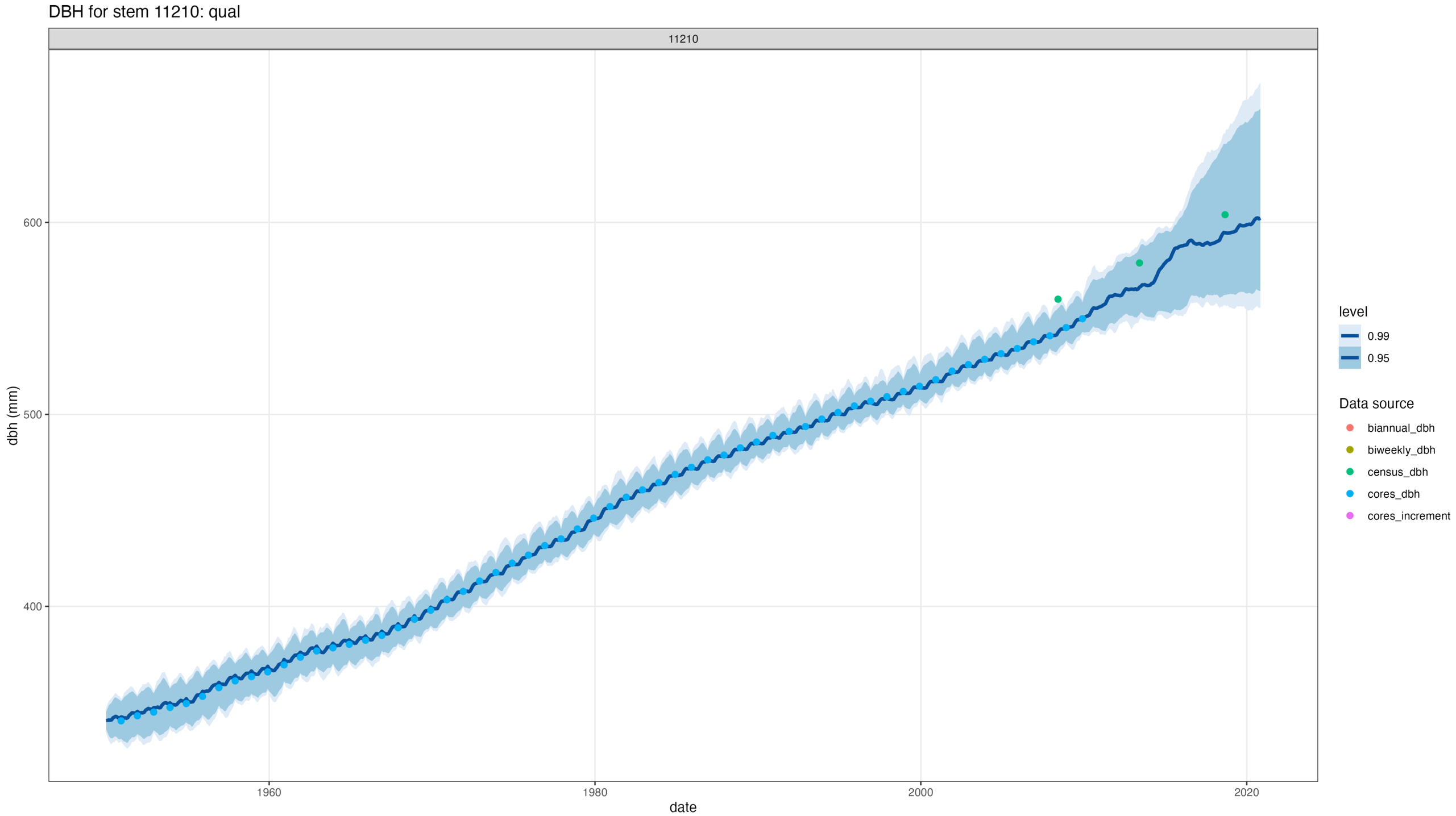
- cyan = process model parameters: baseline growth + dbh slope
- pink = measurement errors SD
- green = process error i.e. residual
- purple = random effects errors SD

Next steps

TODO

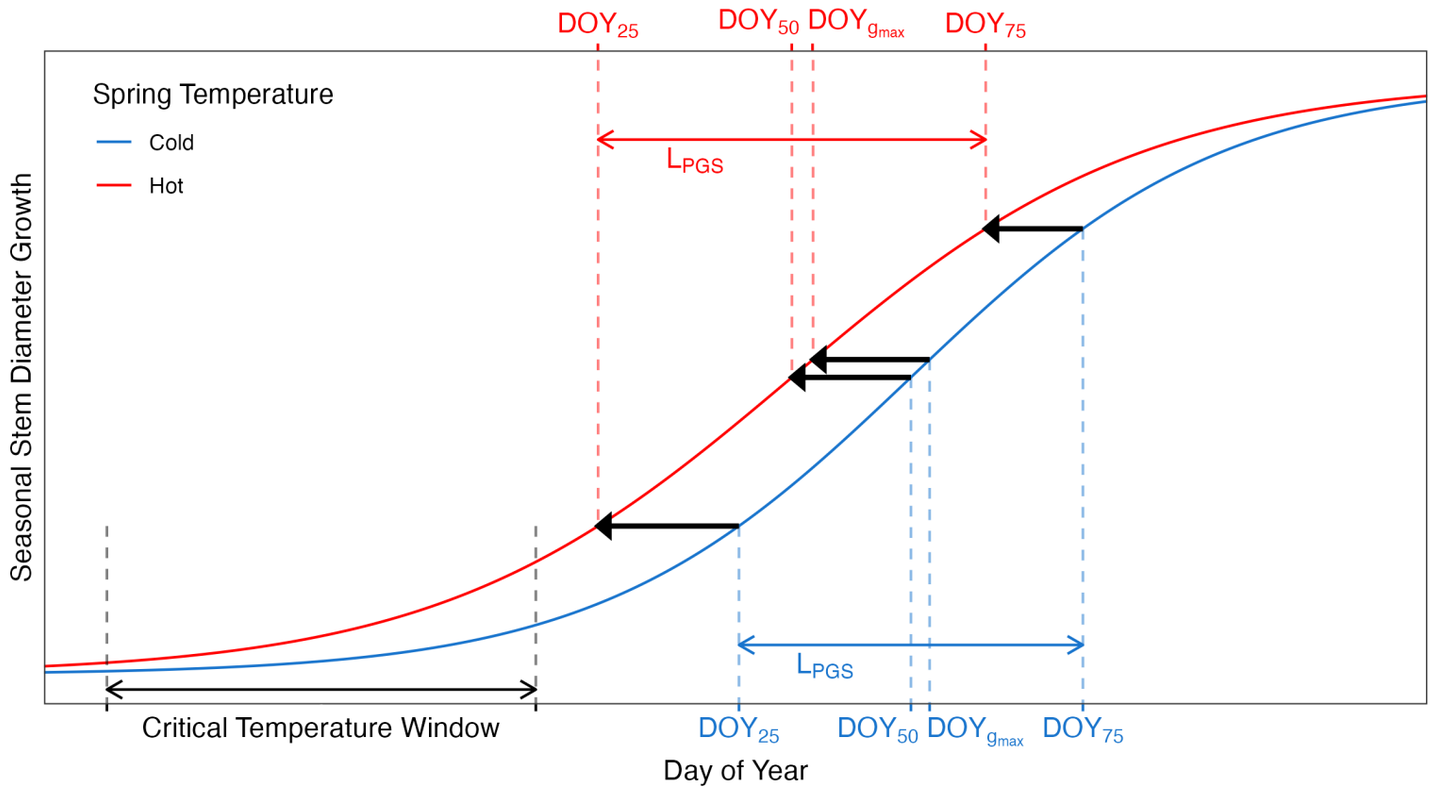
1. Restrict model to have non-negative true latent growth from t to t-1
2. Add climate covariate to process model e.g. β_{temp} based on Helcoski 2019
3. Deal with correlation of observations/errors

Correlation of observations/errors



Related: ForestGEO Seminar Series

"Warmer spring temperatures in temperate deciduous forests cause earlier tree growth but have little effect on annual woody productivity"



Wed January 19th 9-10am ET

Sign up

Thank you!

Slides available on Twitter
@rudeboybert