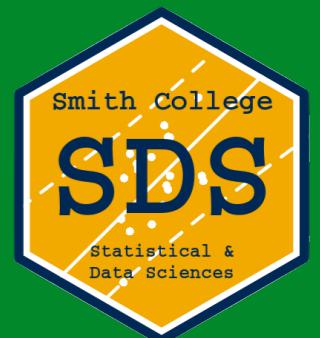


Fusing disparate measurement data for forecasting the growth of trees via Hidden Markov Models



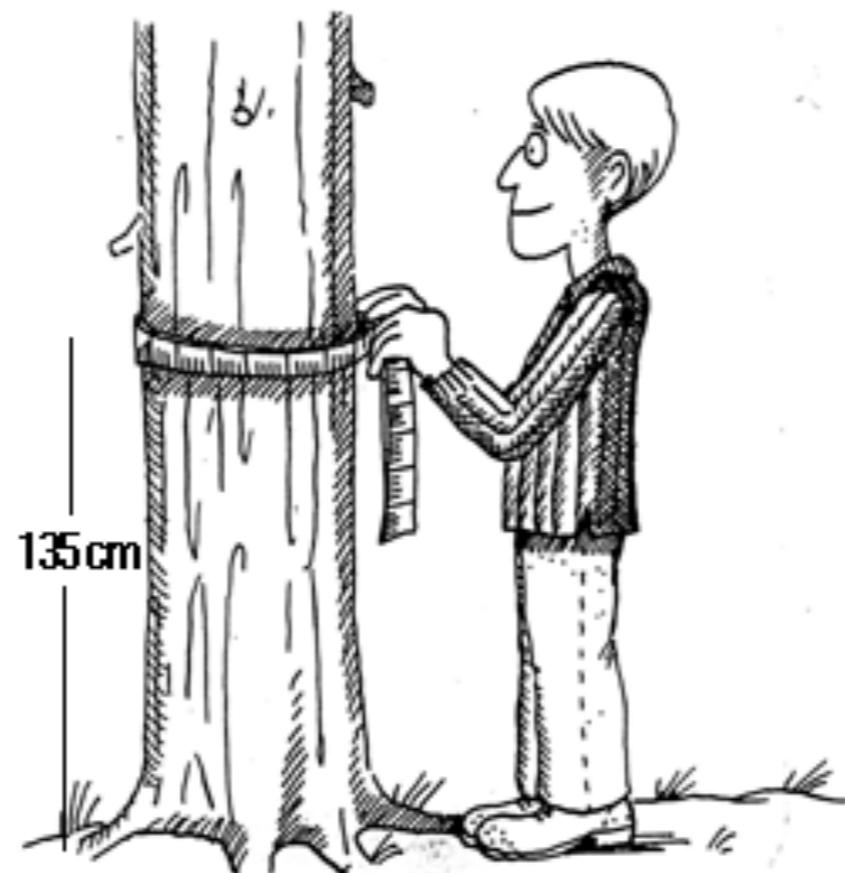
Prof. Albert Y. Kim
UMass Amherst Statistics Seminar Series
Friday, January 22, 2021



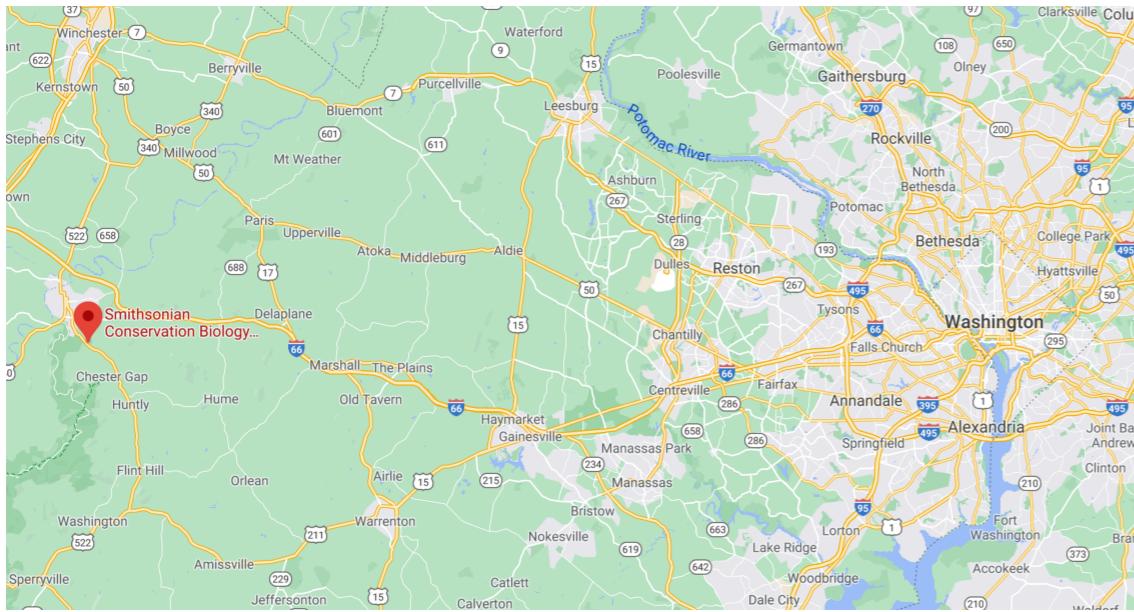
Context

Diameter at Breast Height (dbh)

After species & location, one of the most informative variables about a tree is dbh

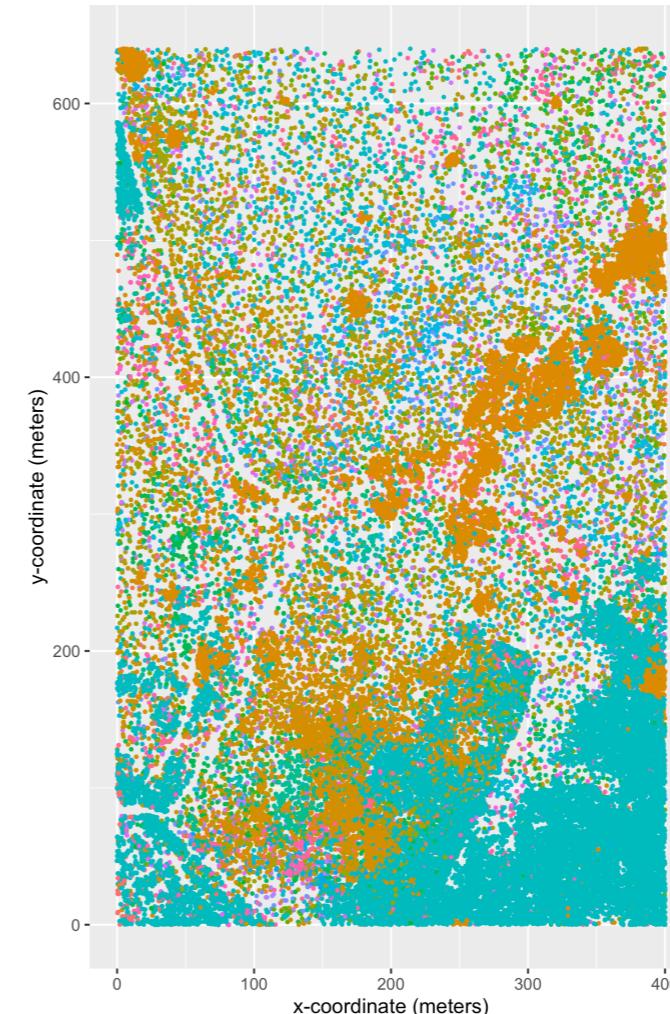


Smithsonian Conservation Biology Institute



25.6 ha = 35.85 soccer fields

Census 2018: 72,555 cataloged trees



dbh >10mm
are tagged

Data on GitHub

The screenshot shows a GitHub repository page for 'SCBI-ForestGEO / Dendrobands'. The repository has 5 stars, 0 forks, and 3 issues. The 'Code' tab is selected. A specific file, 'Dendrobands / data / scbi.dendroAll_2020.csv', is displayed. The file contains 1280 lines (1280 sloc) and is 190 KB. It was last updated on Jul 3 by 'rudeboybert'. Four contributors are listed. The CSV data is shown in a table:

1	tag	stemtag	survey.ID	year	month	day	biannual	intraannual	sp	quadrat	lx	ly	measure	codes	notes
2	10469	1	2020.01	2020	3	11	1	0	litu	109	9.7	1	NA	RE	window too large to measure
3	10587	1	2020.01	2020	3	11	1	0	litu	113	2.6	13	61.41	NA	NA
4	10609	1	2020.01	2020	3	11	1	0	cagl	111	19.5	2.9	81.03	NA	double-checked

Two columns, 'tag' and 'measure', are highlighted with red boxes.

Equipment to measure doh



Measuring tape. Call
this “census” data



Tree coring +
dendrochronology.
Call this “core” data

Equipment to measure doh



Dendrobands +
Calipers: Call this
“dendro” data



Comparison Chart

Equipment to measure doh



Goal

Model

Hidden Markov Models

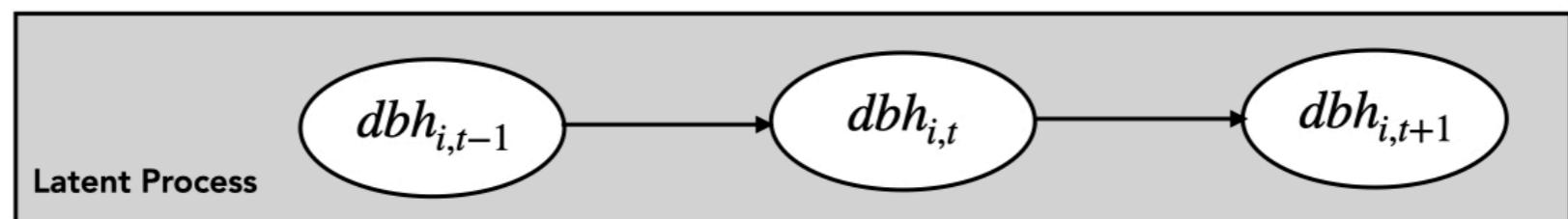
- Hidden: Data fusion via latent variables
- Markov: Observation
- Split out sources of error into those that are
 - Are not of direct interest
 - Are “one and done” i.e. measurement error
 - Propagate when forecasting

Minimum Viable Product

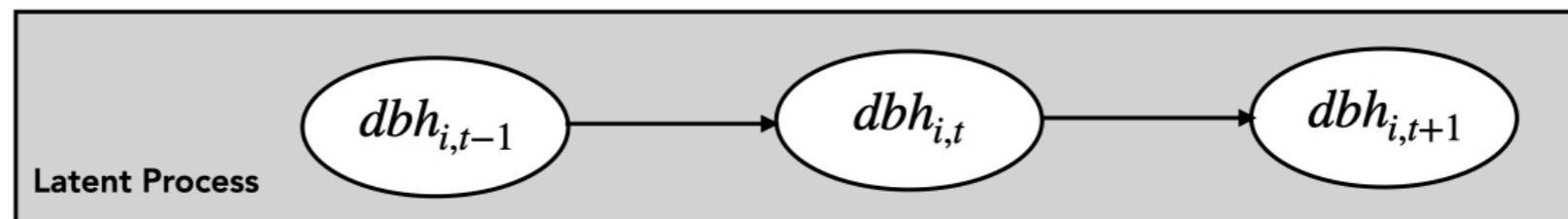
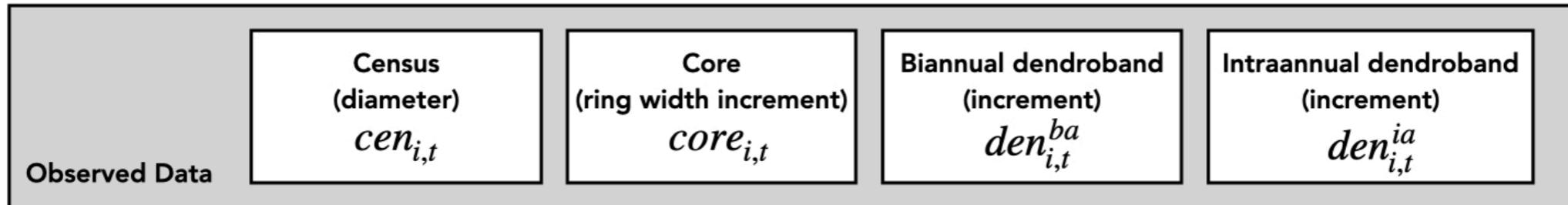
$$dbh_{i,t} = dbh_{i,t-1} + \beta_0 + \beta_i + \beta_t + \epsilon$$

- $dbh_{i,t}$: “True” latent dbh for individual i at time t
- β_0 : Baseline growth
- β_i : Individual tree i random effect
- β_t : Time point t random effect
- $\epsilon \sim \text{Normal}(0, \sigma_\epsilon^2)$

Model

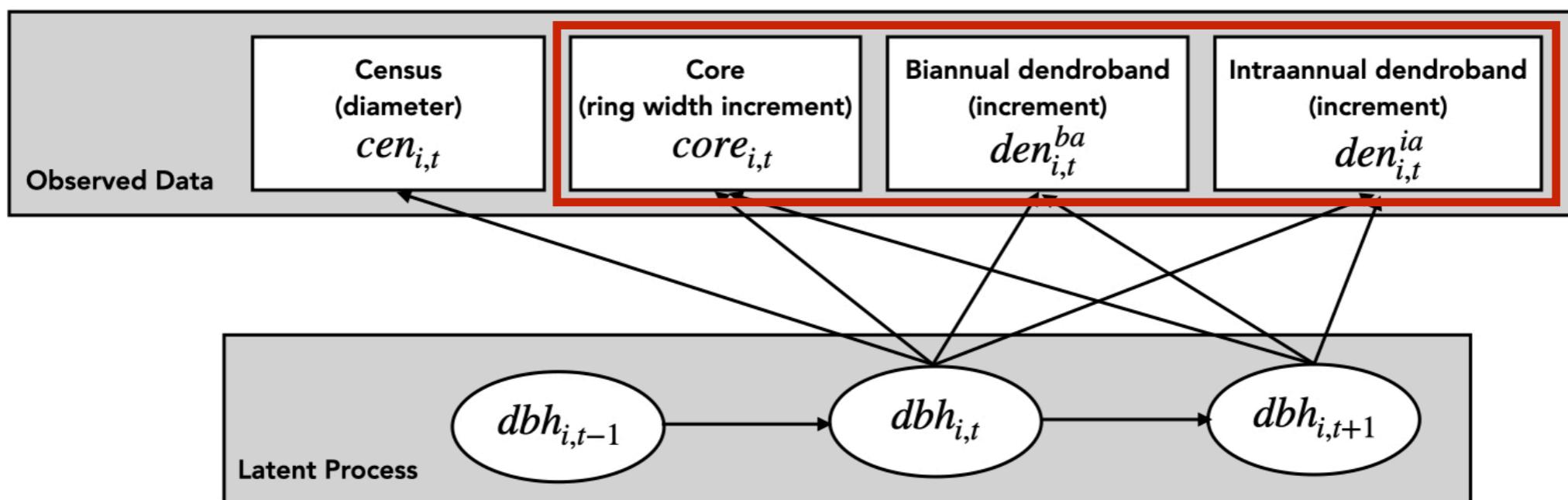


Model

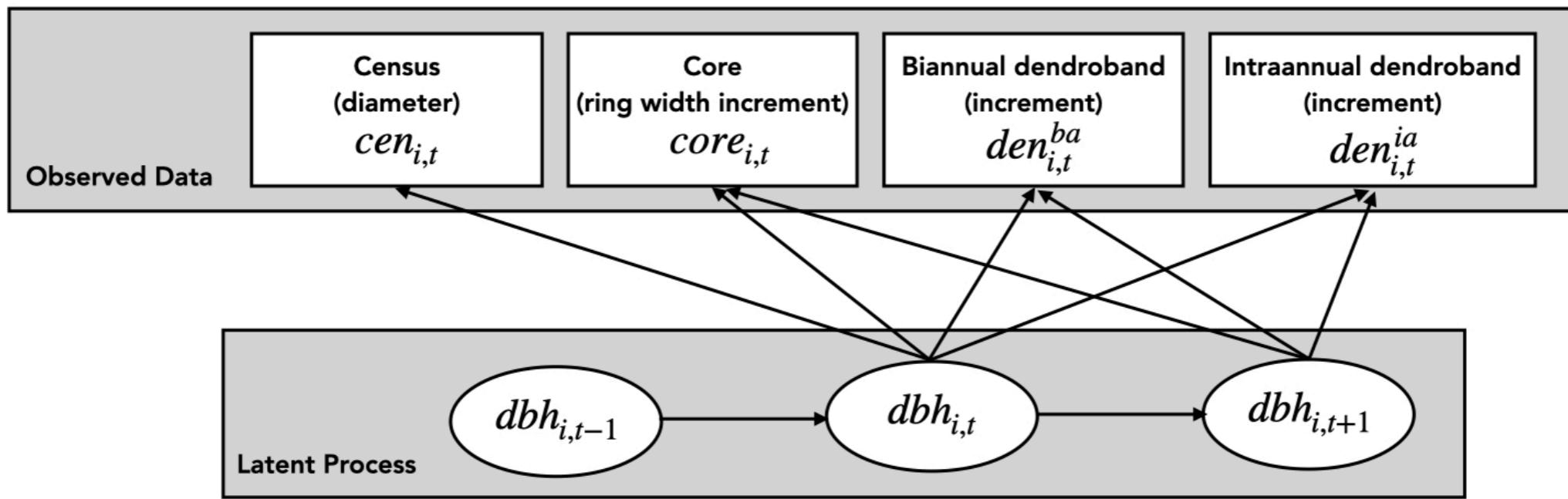


Model

$$\text{Increments} = dbh_{i,t-1} - dbh_{i,t}$$

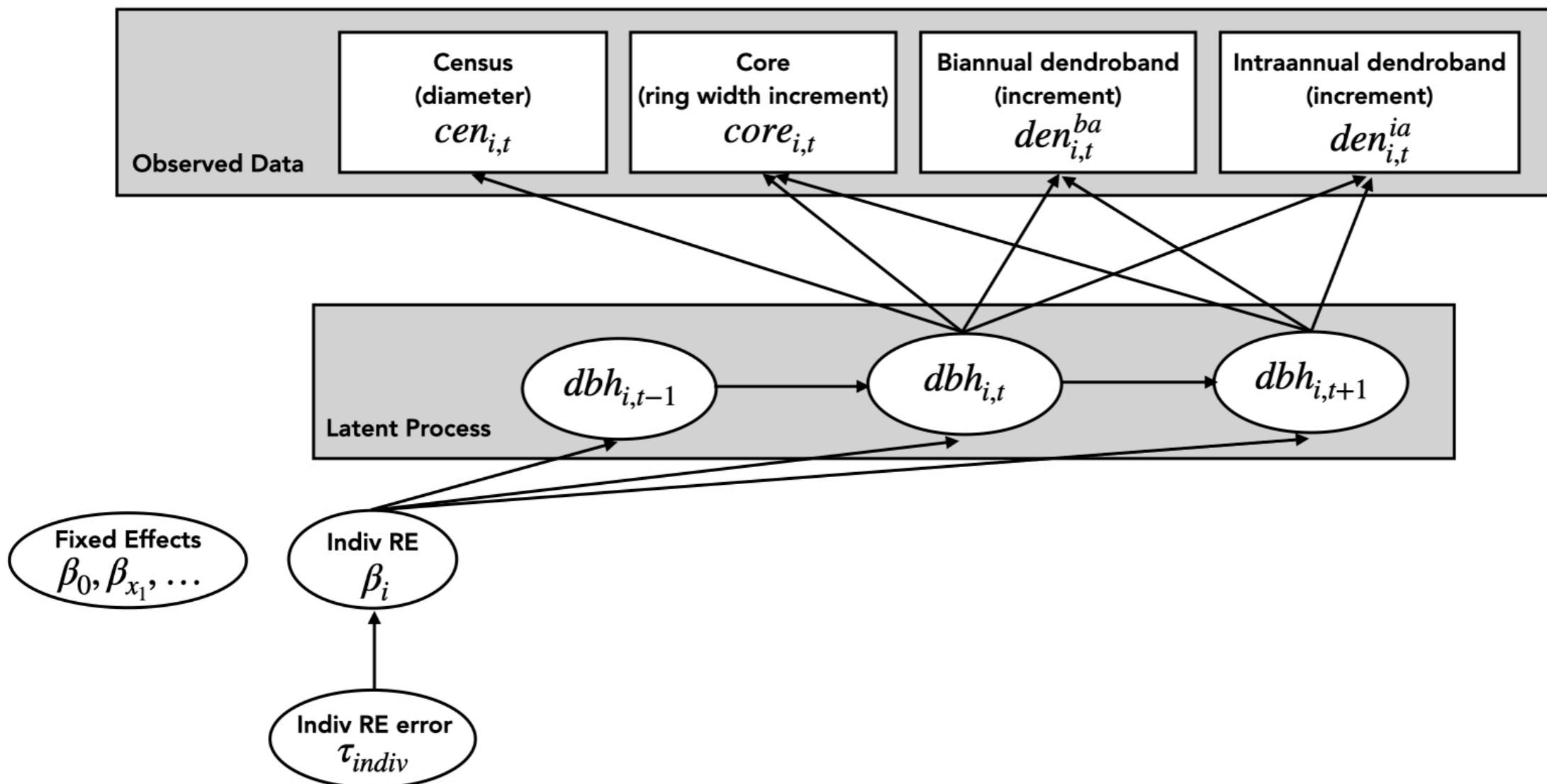


Model

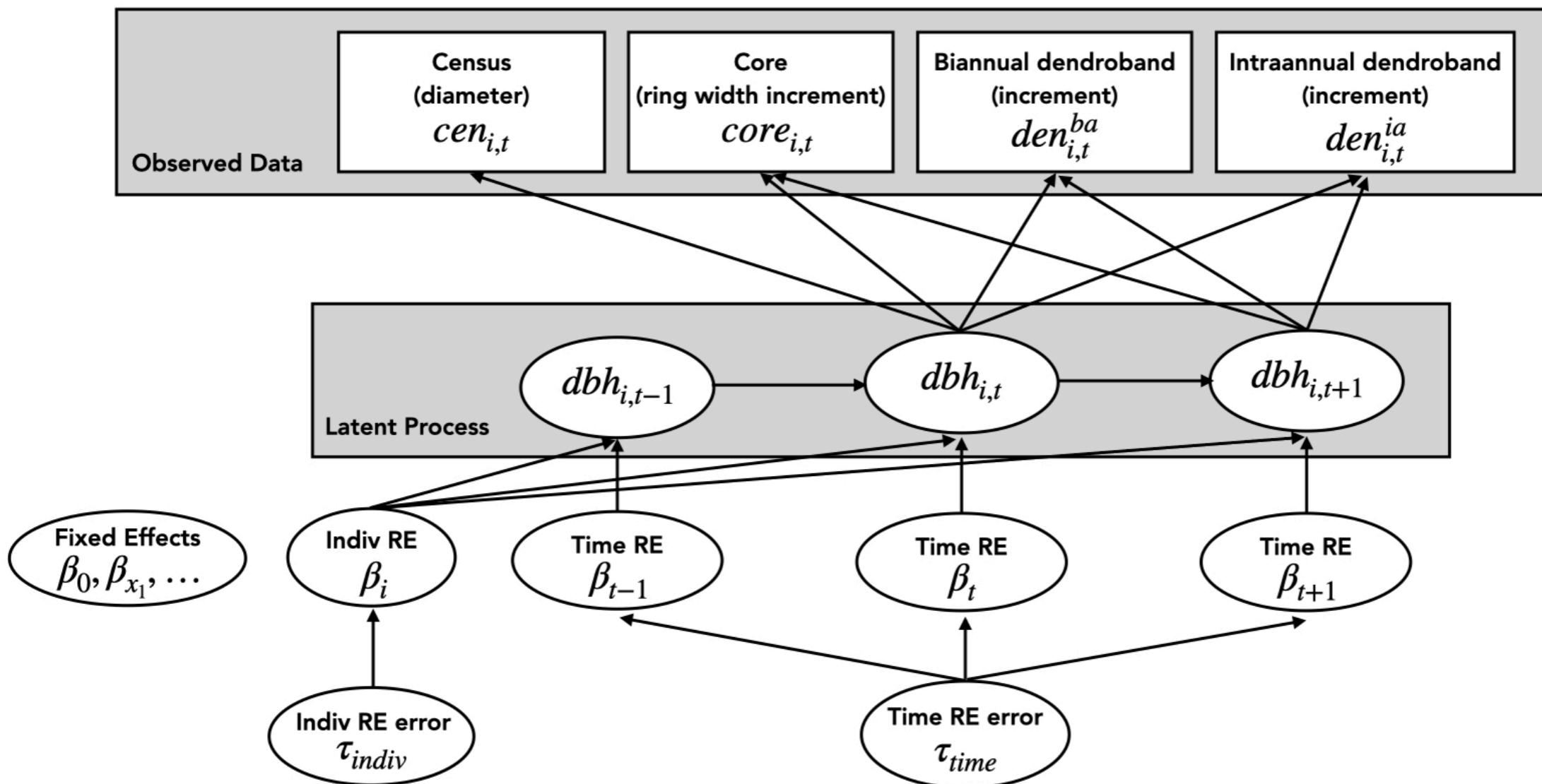


Fixed Effects
 $\beta_0, \beta_{x_1}, \dots$

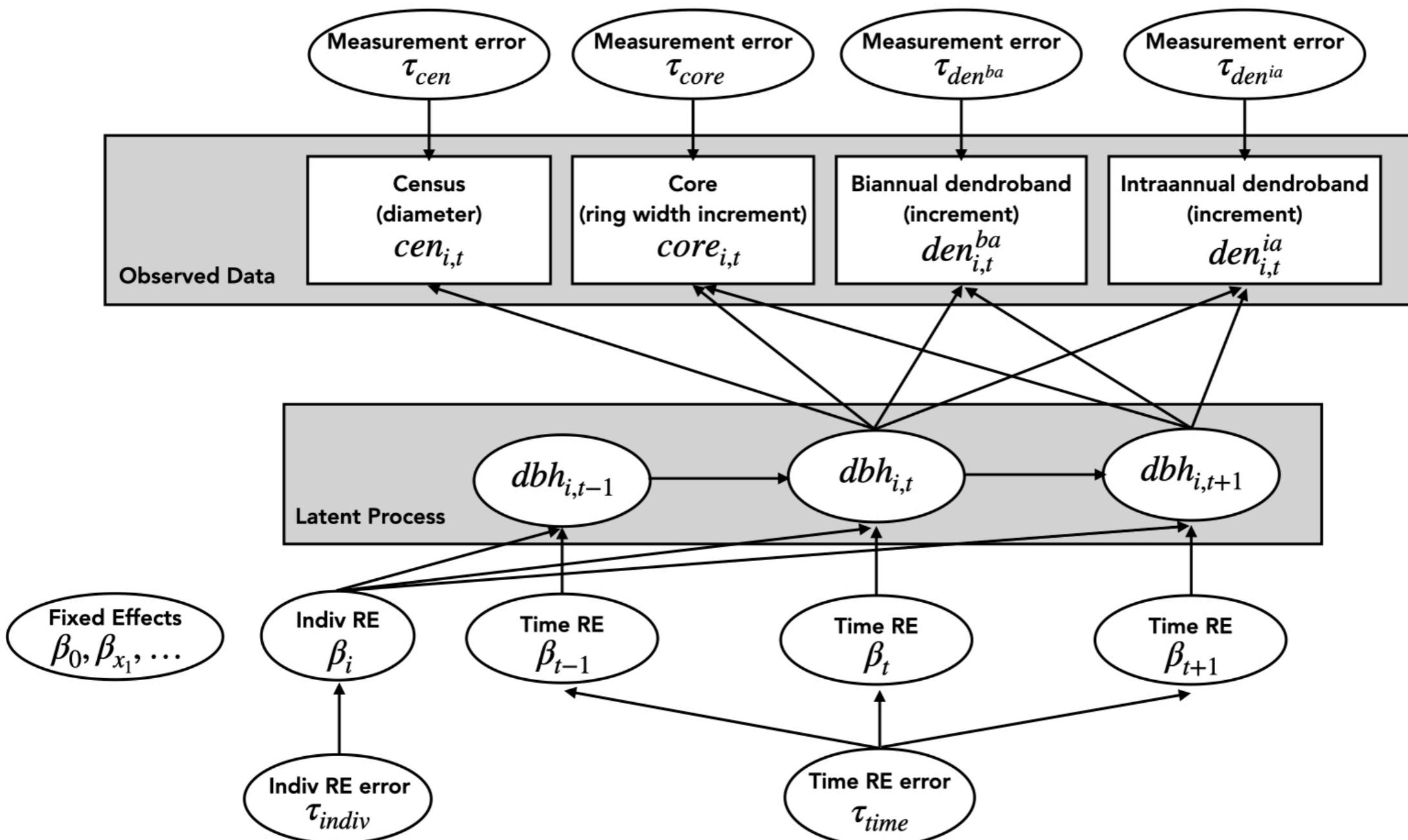
Model



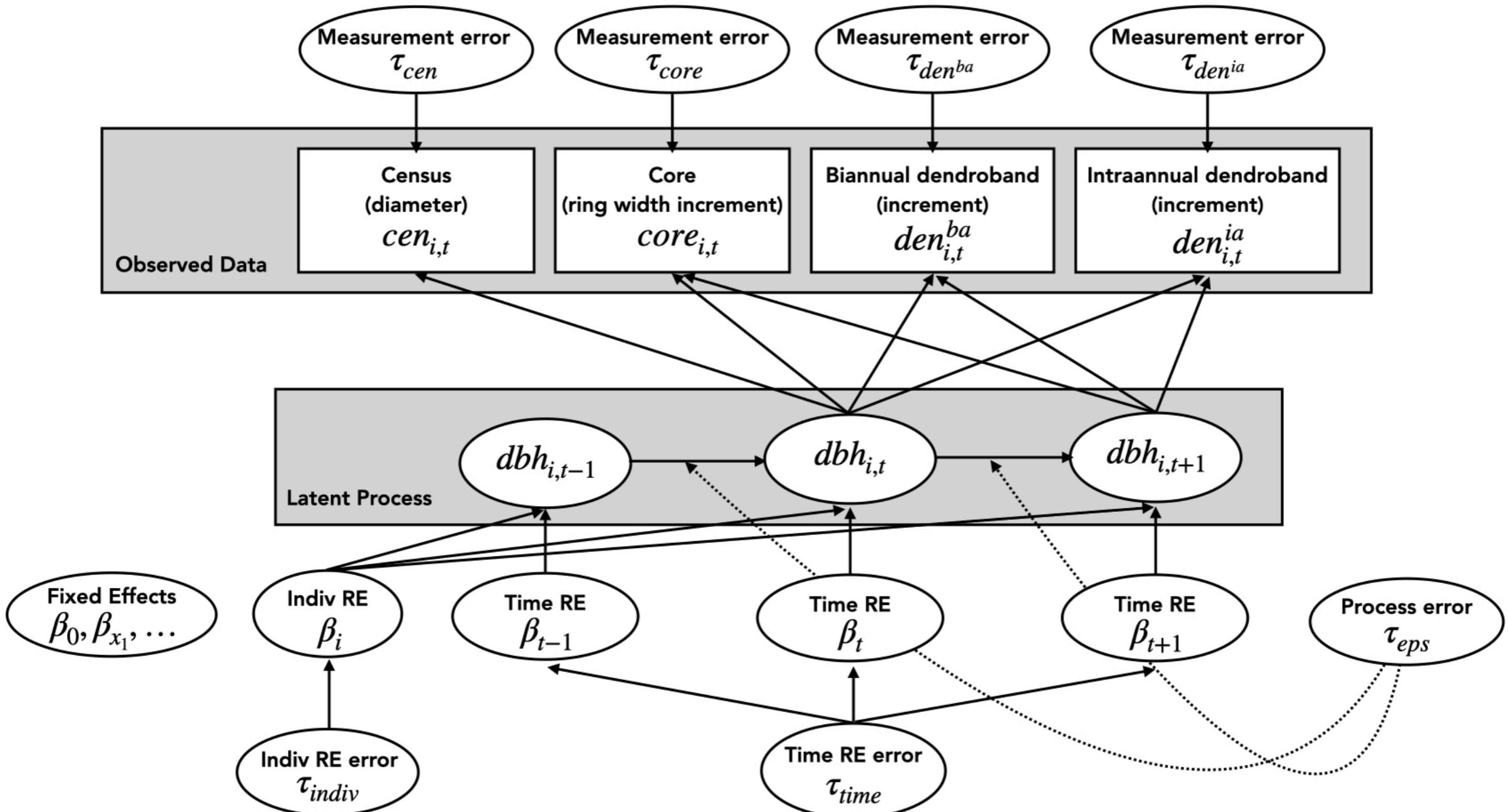
Model



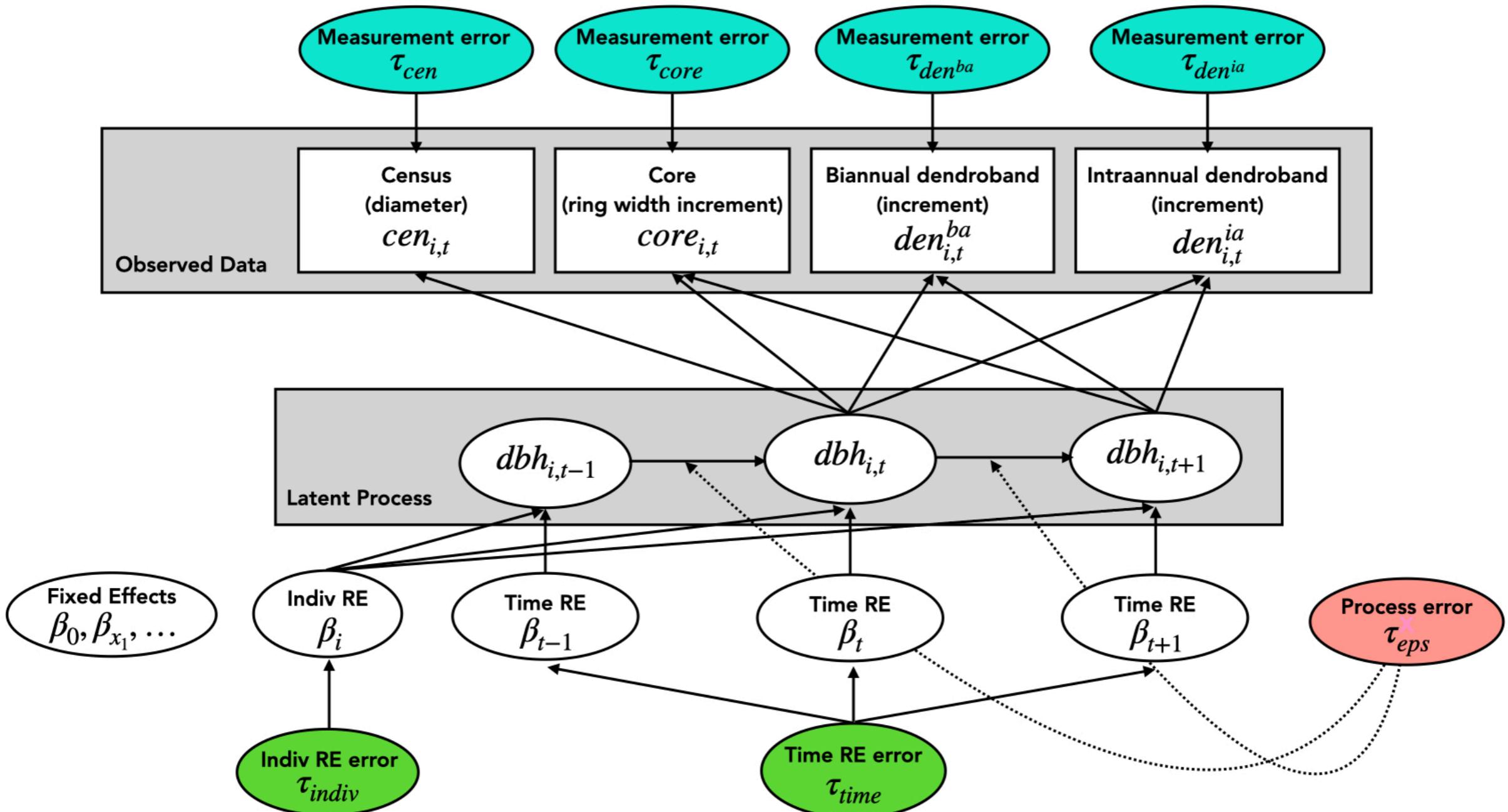
Model



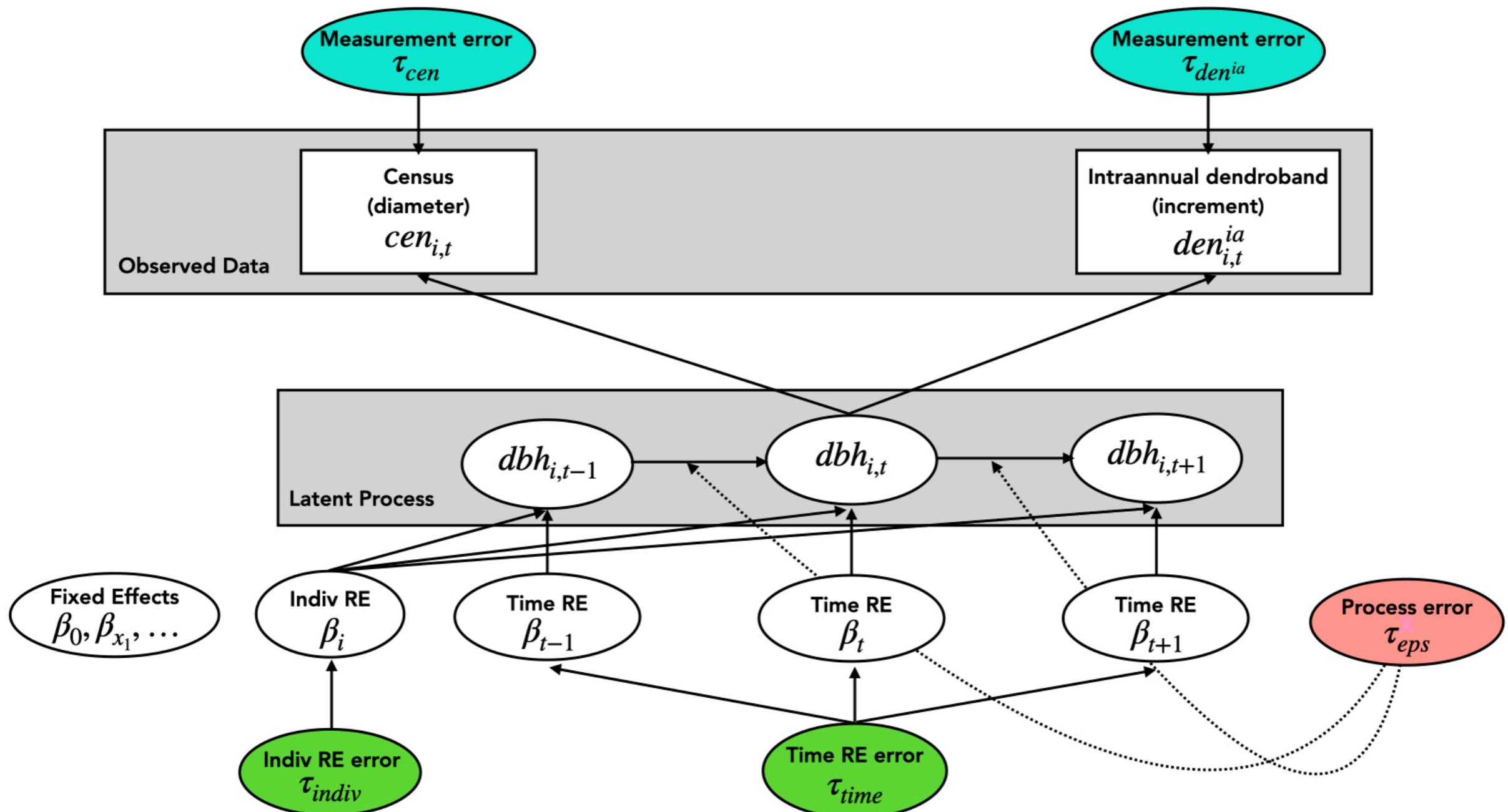
Model



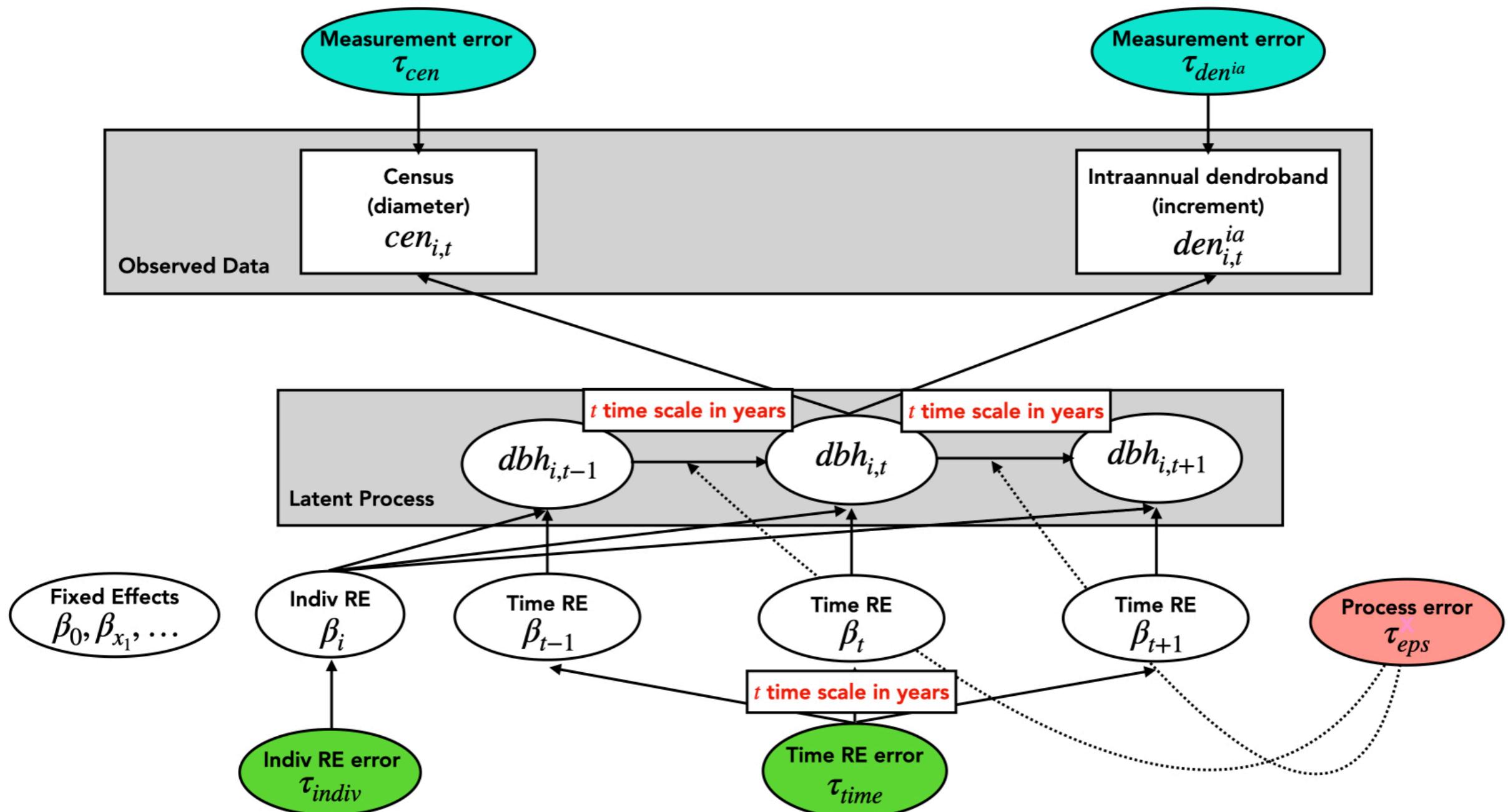
Model



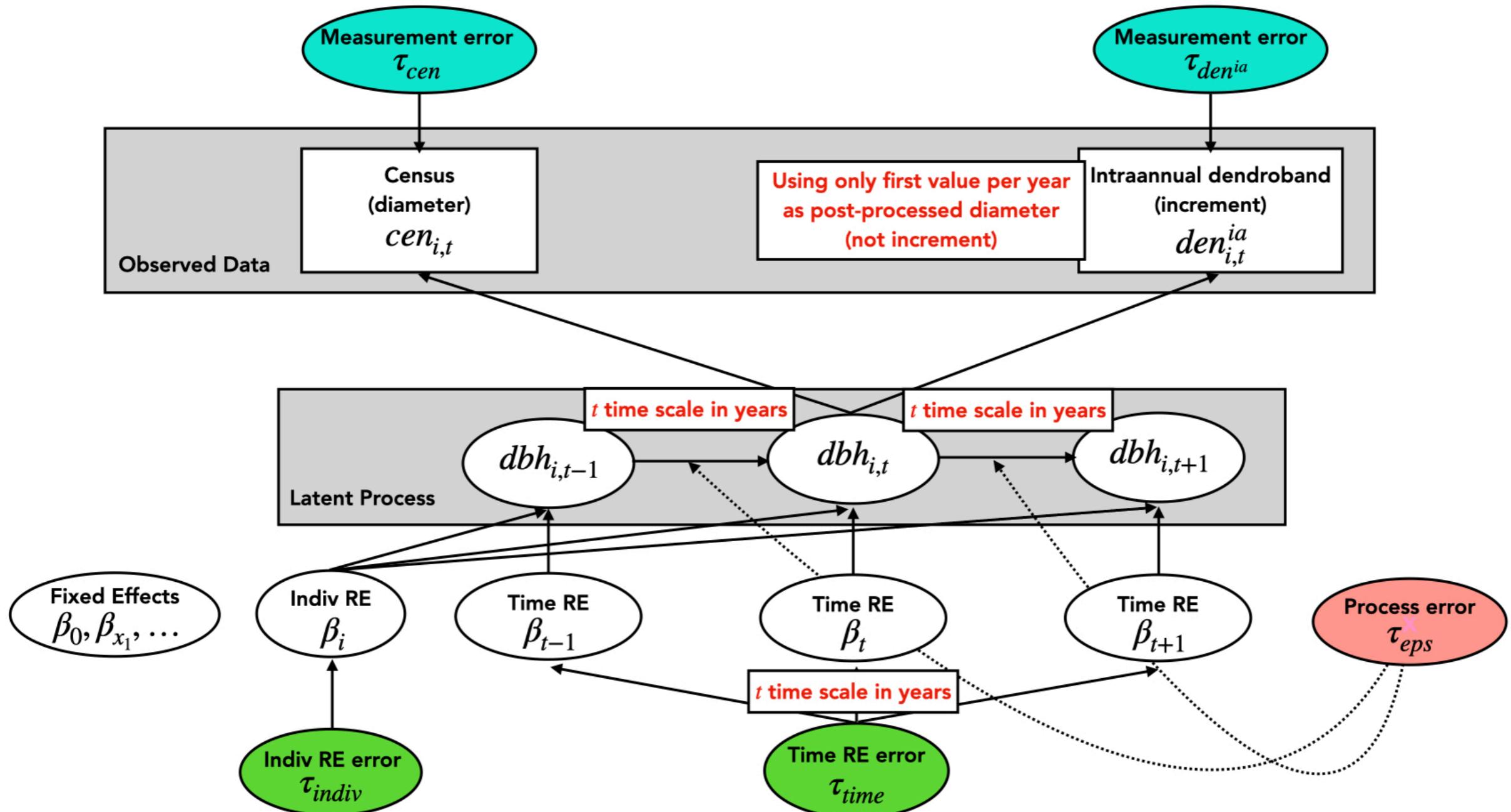
Model as of 2021/1/22



Model as of 2021/1/22



Model as of 2021/1/22

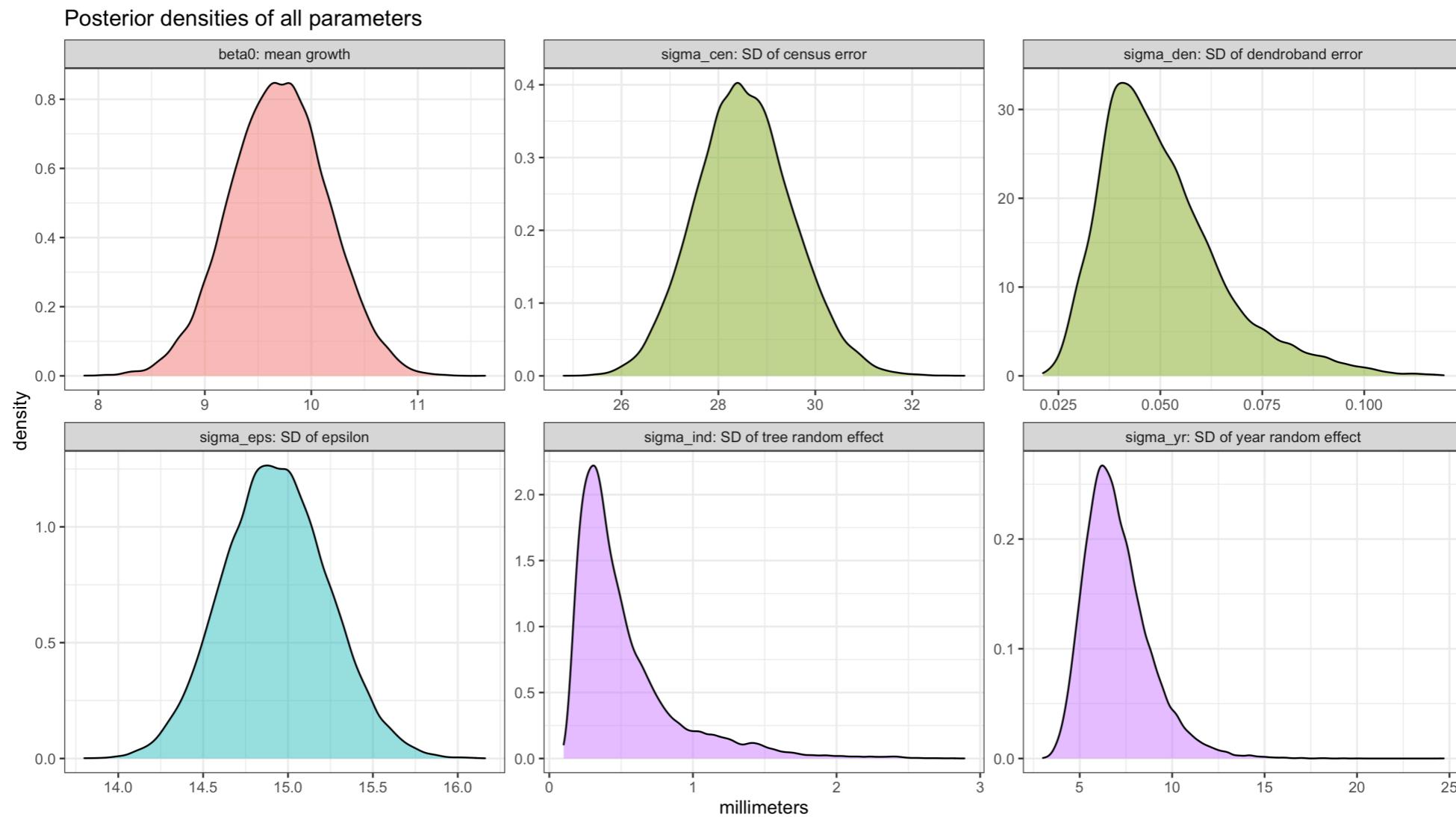


Results

MCMC specifications

- Implemented in JAGS
- 30k draws from posterior minus 10% burn-in
- Empirical Bayes (data informed) prior parameters
- Forecast into 2020 - 2022 by treating these years as missing values

Posterior Distributions



- mean $\beta_0 \sim 1\text{cm}$ growth per year
- $\sigma_{cen} > > \sigma_{den}$
- Year-to-year variation in growth > Between individuals variation
- σ_ϵ = remaining process error that propagates in forecasts across time

One particular tulip poplar

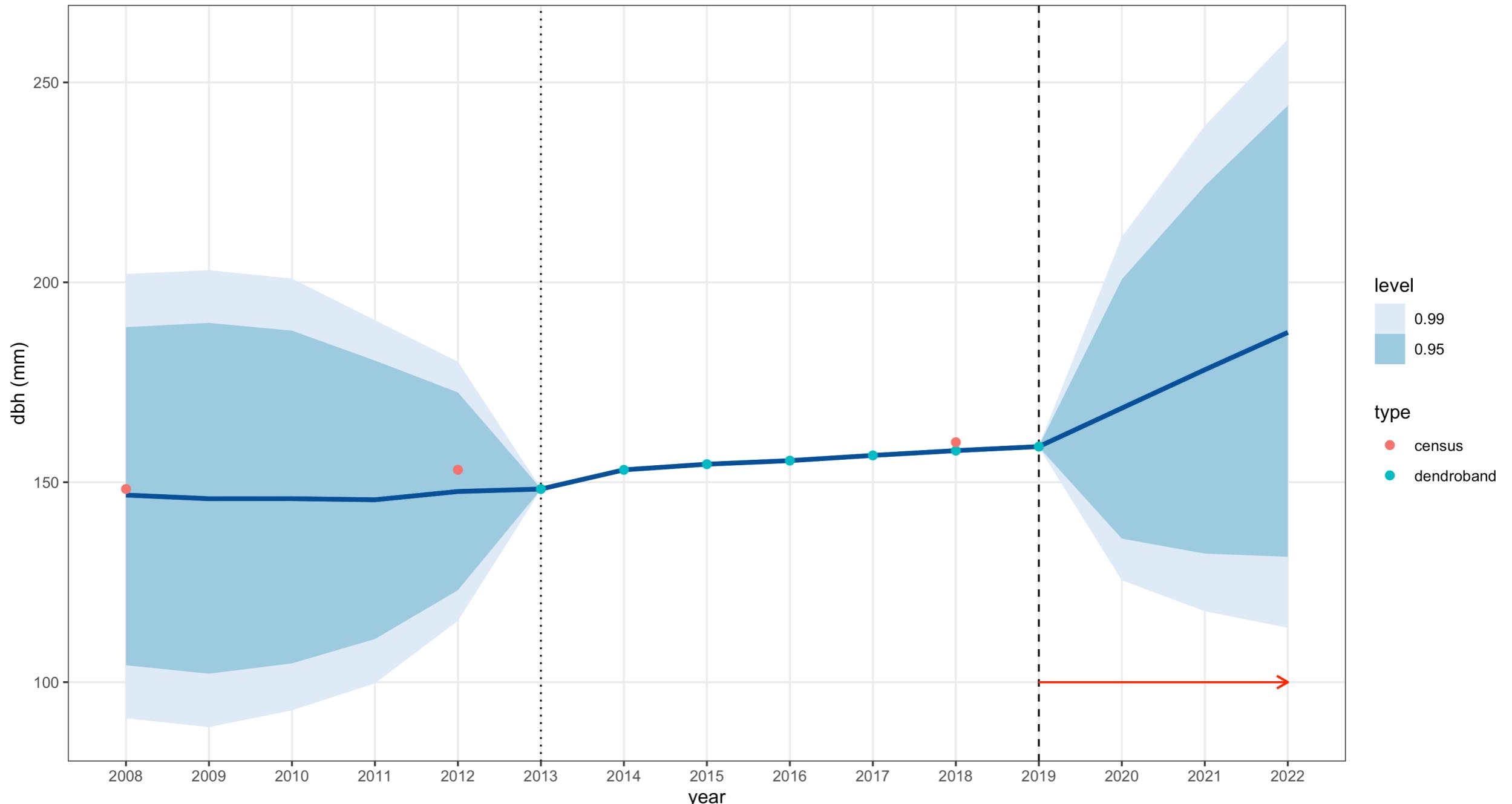
tag_stem	type	sp	`2007`	`2008`	`2009`	`2010`	`2011`	`2012`	`2013`	`2014`	`2015`	`2016`	`2017`	`2018`	`2019`
<chr>	<chr>	<chr>	<dbl>												
30339_3	census	litu	NA	148.	NA	NA	NA	153.	NA	NA	NA	NA	NA	160	NA
30339_3	dendroband	litu	NA	NA	NA	NA	NA	NA	149.	155.	156.	157.	157.	159.	160.



One particular tulip poplar diameter

$y = \text{modeled true latent } dbh_{i,t}$

Tag 30339: litu



Dendroband installed in 2013

Future Work

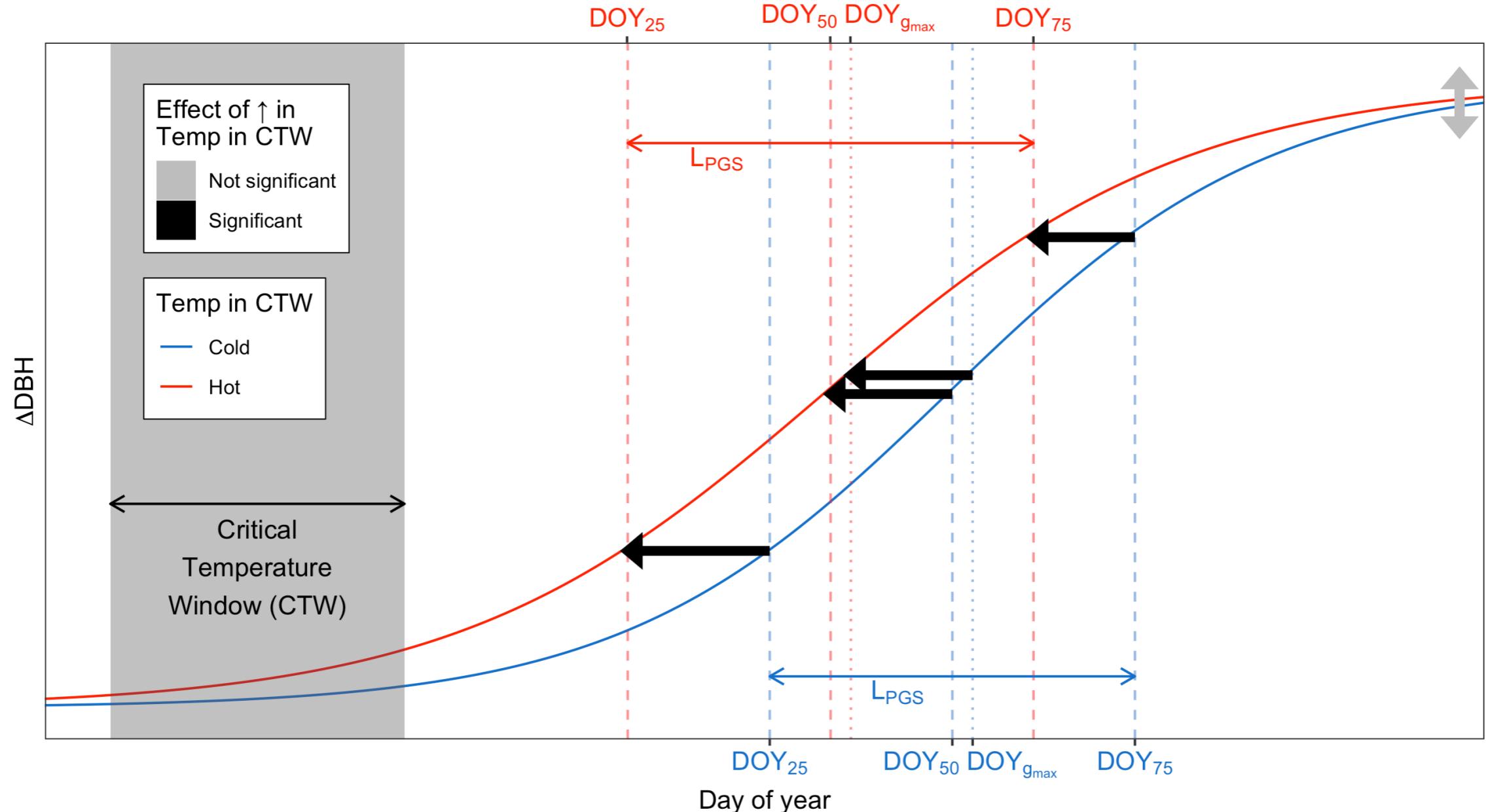
TODO

- Add remaining data:
 - All, not just first yearly observation, intra-annual dendroband
 - All biannual dendroband
 - Tree coring data
- Merge dendroband after comparing τ_{den}^{ba} versus τ_{den}^{ia} ?
- Improve model for latent variable
 - $dbh_{i,t} = dbh_{i,t-1} + \beta_0 + \dots + \epsilon$
 - Covariates: In particular species & starting diameter
- Choose appropriate time scale for t

Thanks!

Slides on Twitter
@rudeboybert

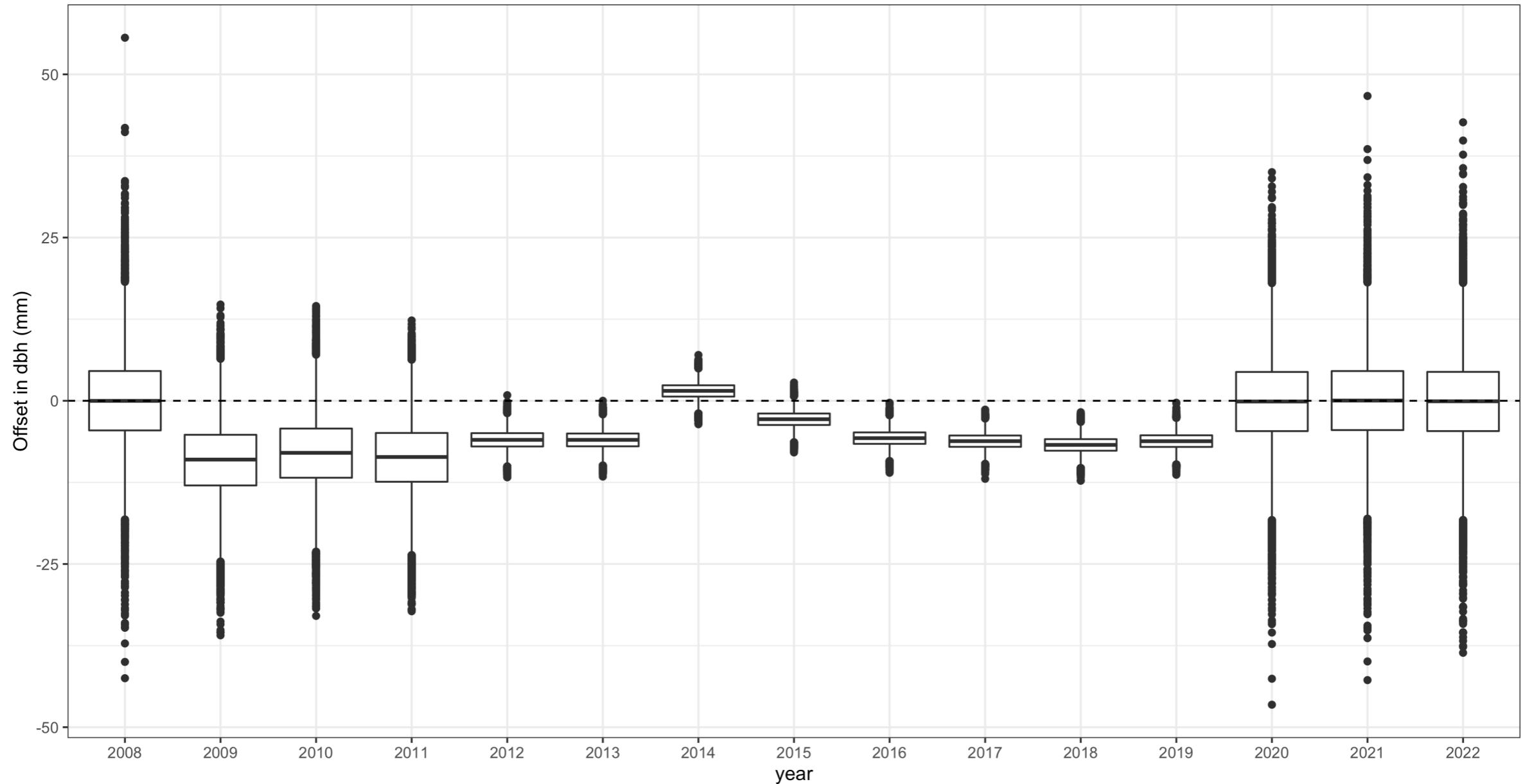
Intra-annual effect of climate



Year Random Effects

Year random effects

Distribution of all MCMC draws from posterior for each year



Individual Random Effects

Individual tree random effects

Distribution of all MCMC draws from posterior for each tree

