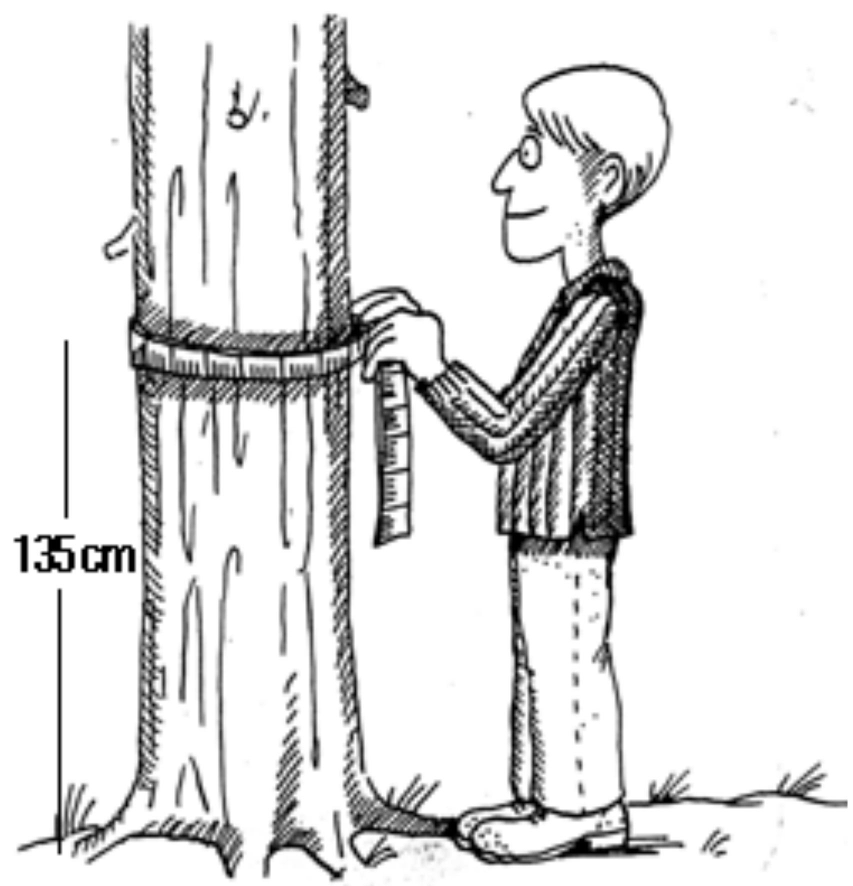
Fusing disparate measurement data for forecasting the growth of trees via Hidden Markov Models

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

Context

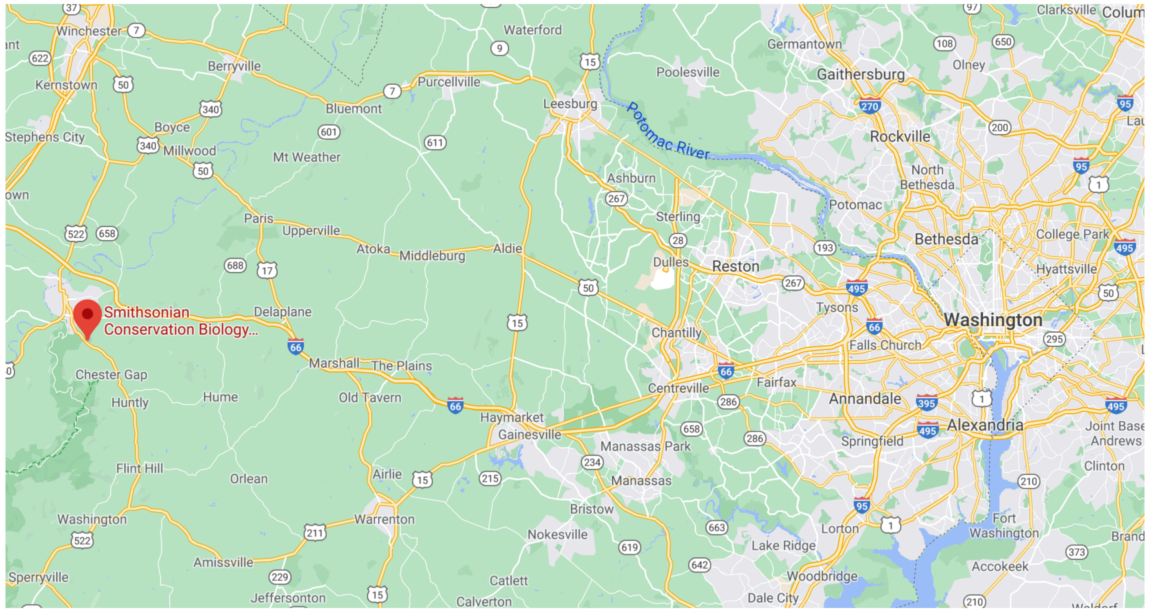
Diameter at Breast Height (dbh)

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

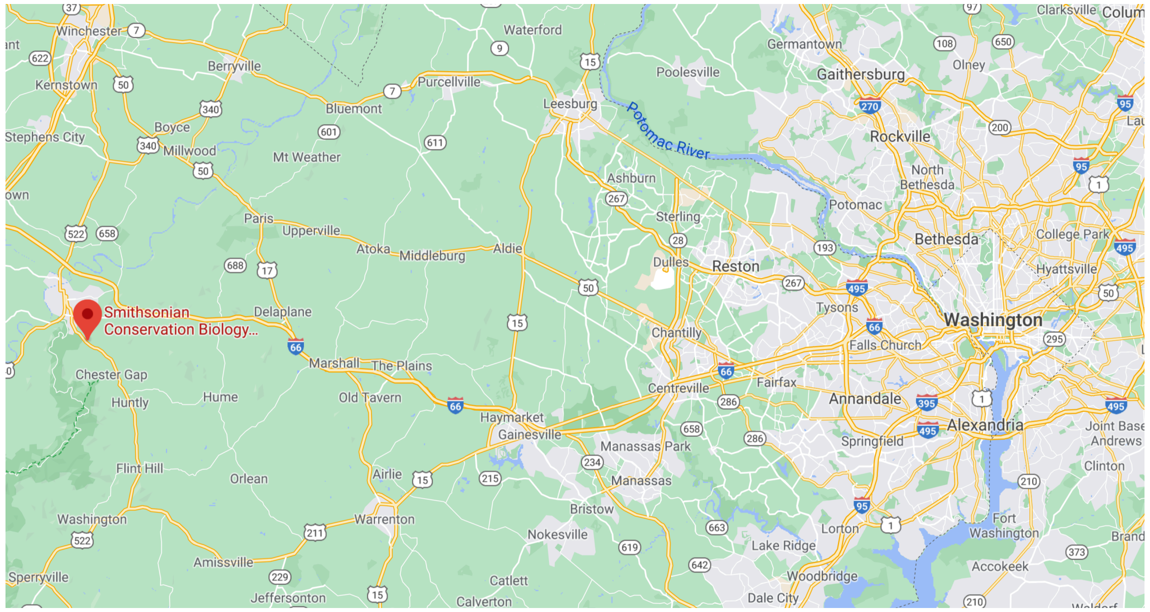


Smithsonian Conservation Biology Institute

Smithsonian Conservation Biology Institute

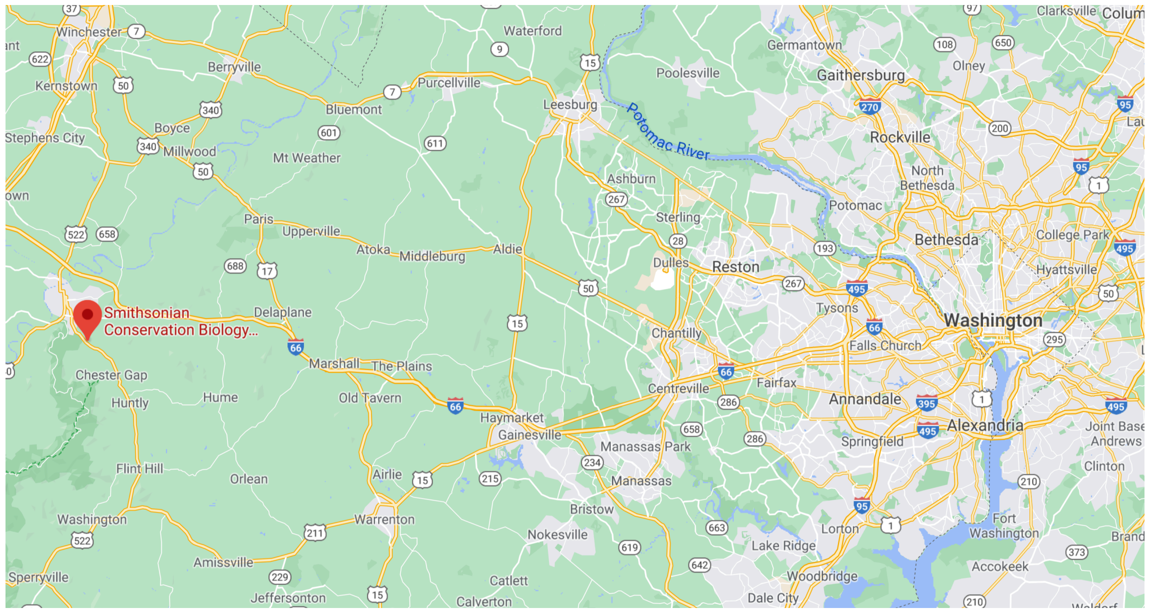


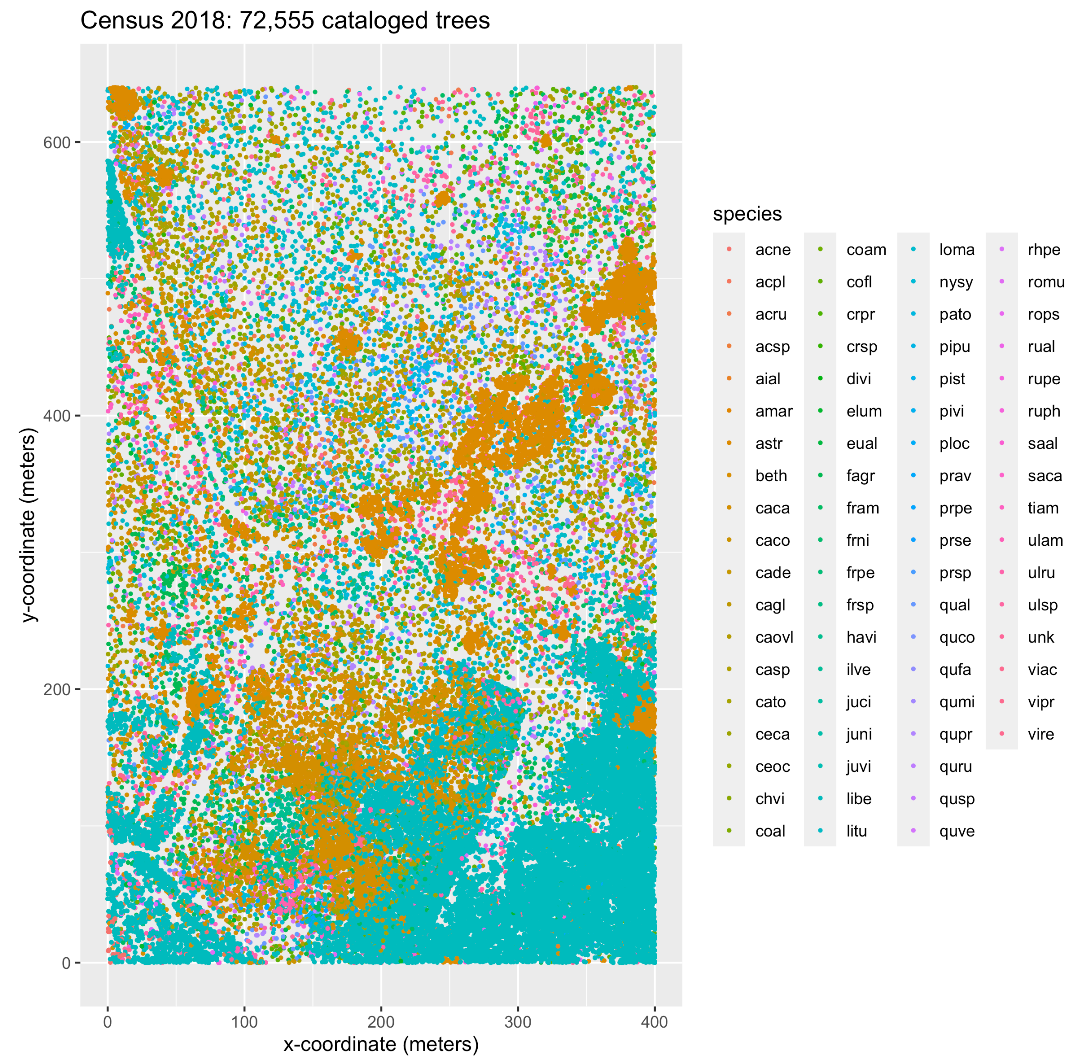
Smithsonian Conservation Biology Institute



25.6 ha = 35.85 soccer fields

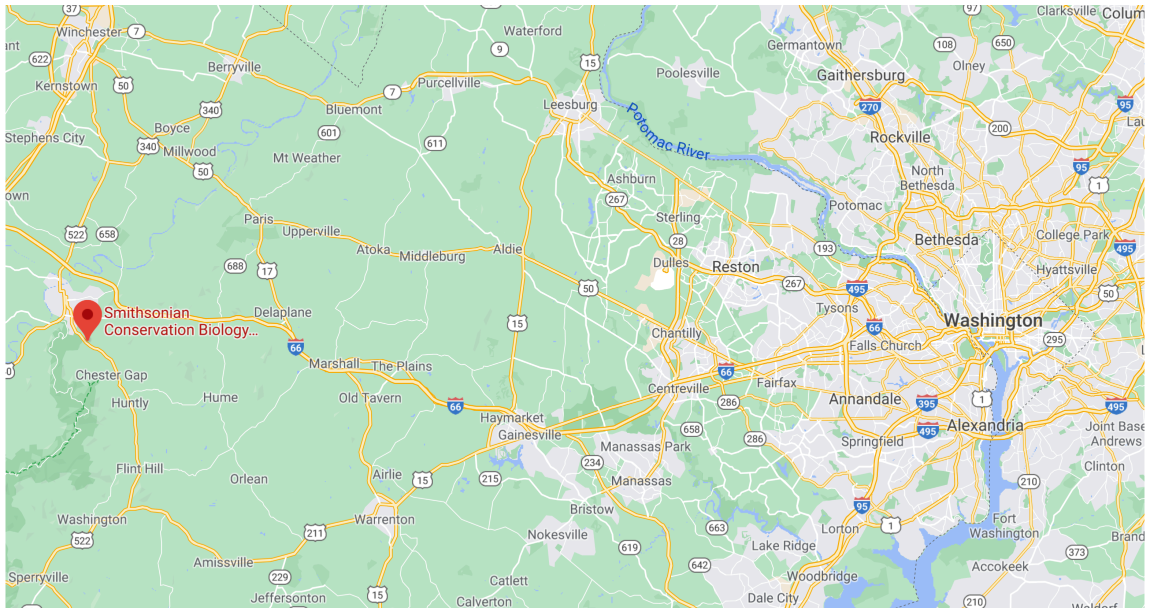
Smithsonian Conservation Biology Institute



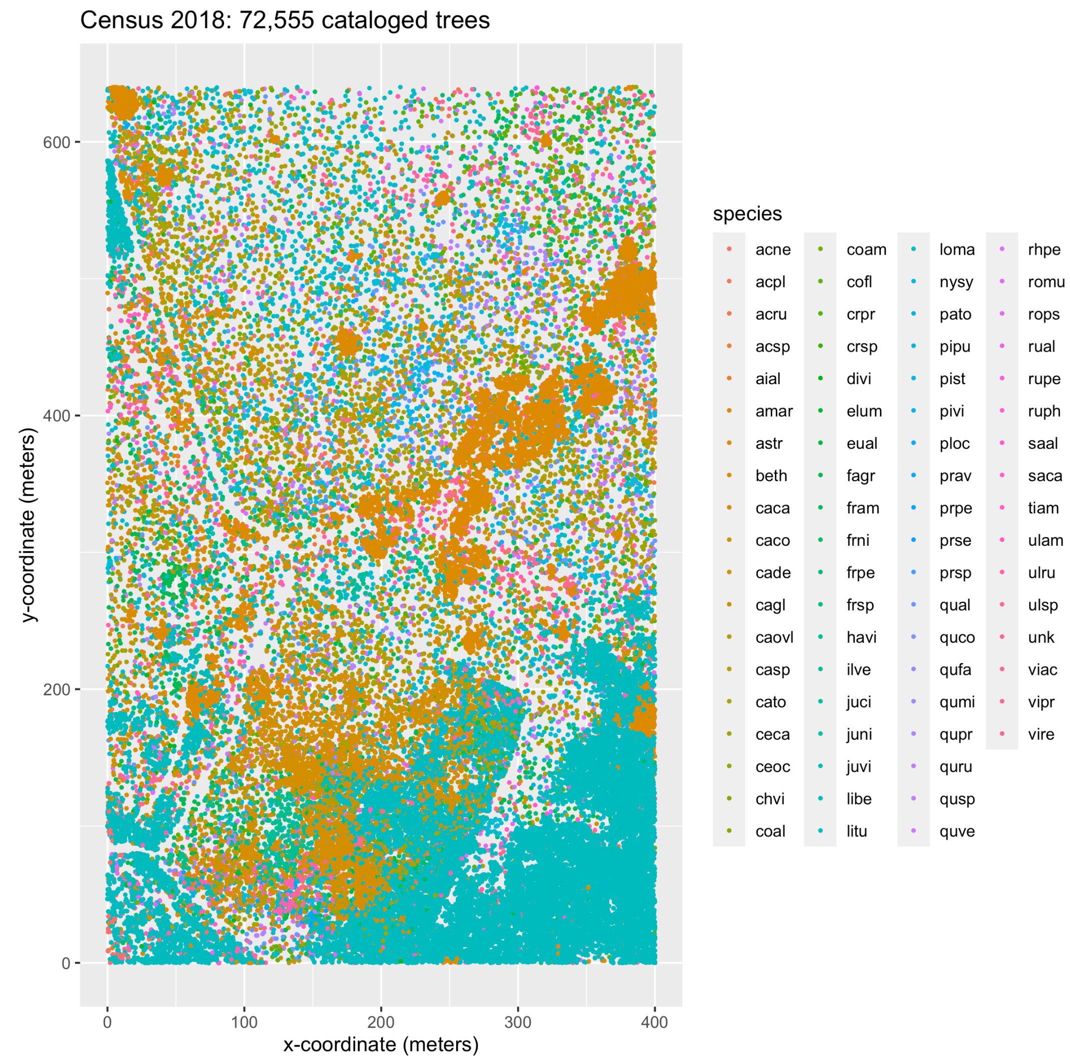


25.6 ha = 35.85 soccer fields

Smithsonian Conservation Biology Institute

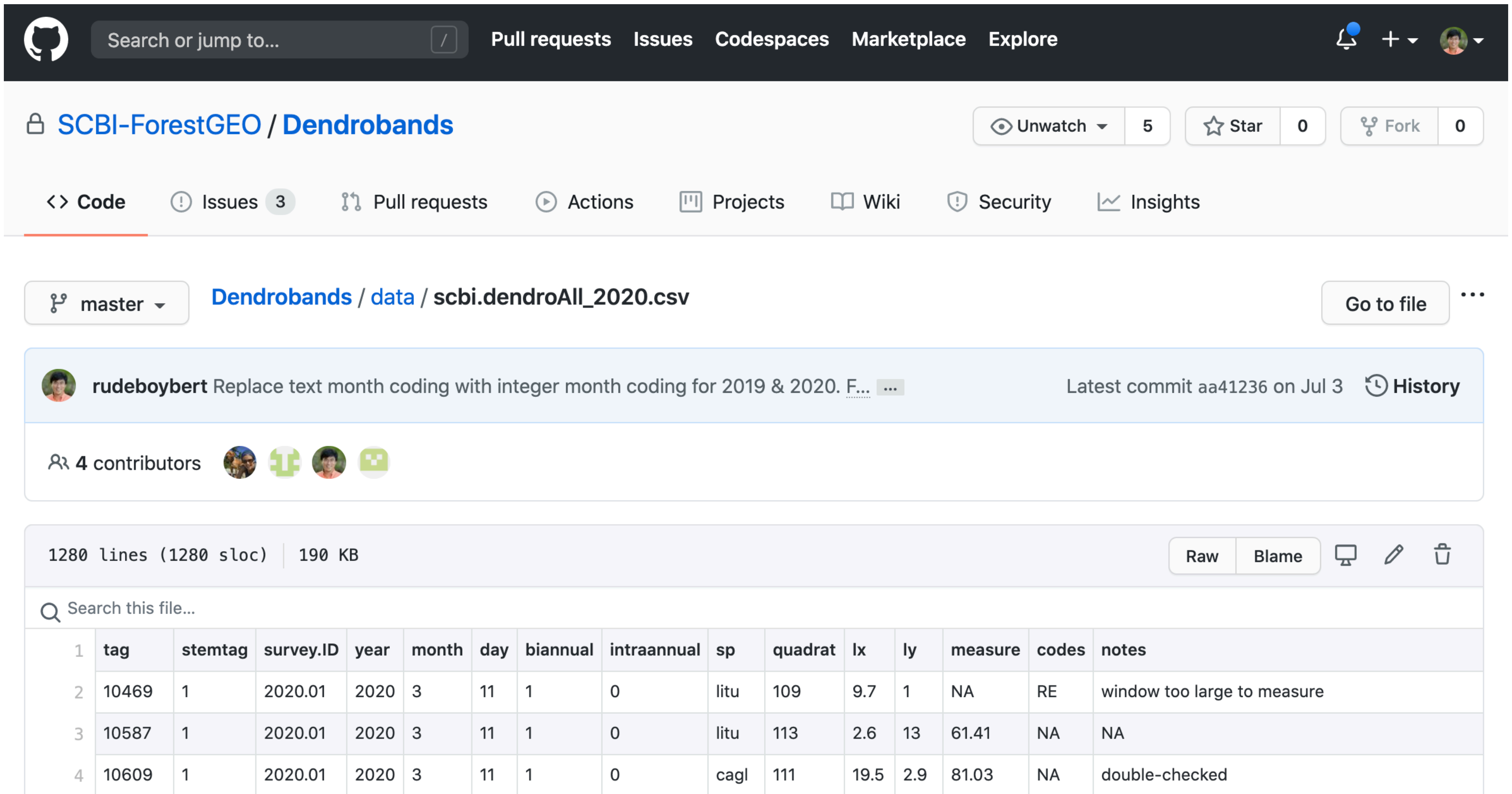


25.6 ha = 35.85 soccer fields



dbh >10mm are tagged

Data on GitHub



	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

Equipment to measure doh

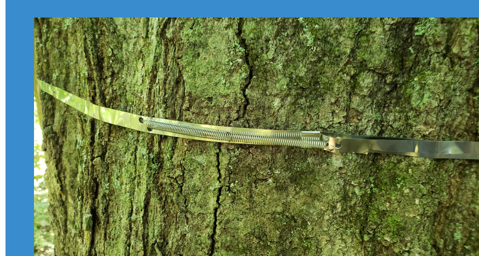
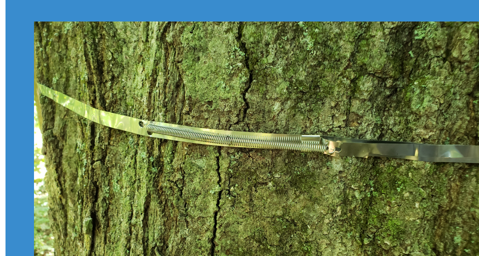
1. Measuring tape. Call this “census” data

2. Tree coring + dendrochronology. Call this “core” data

Equipment to measure doh

3. Dendrobands + Calipers:
Call this “dendro” data

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
	**Biannual dendroband (start & end of year)**	Increment (from baseline)	High setup, rapid follow-up	Climate induced variation in bark & device (-'ve growth)
	**Intraannual dendroband (every 2 weeks)**	"	"	"

Goal 🎯🎯🎯

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

Model

Hidden Markov Models

Hidden Markov Models

- Hidden: “Data fusion” via latent variables
  In our case: “true” dbh

Hidden Markov Models

- Hidden: “Data fusion” via latent variables
  In our case: “true” dbh
- Markov: y_t depends on y_{t-1}

Hidden Markov Models

- Hidden: “Data fusion” via latent variables
  In our case: “true” dbh
- Markov: y_t depends on y_{t-1}
- Partition sources of error into those that

Hidden Markov Models

- Hidden: “Data fusion” via latent variables
  In our case: “true” dbh
- Markov: y_t depends on y_{t-1}
- Partition sources of error into those that
  - Are not of direct interest

Hidden Markov Models

- Hidden: “Data fusion” via latent variables
  In our case: “true” dbh
- Markov: 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

Hidden Markov Models

- Hidden: “Data fusion” via latent variables
  In our case: “true” dbh
- Markov: 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

Minimum Viable Product

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

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

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

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

Minimum Viable Product

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

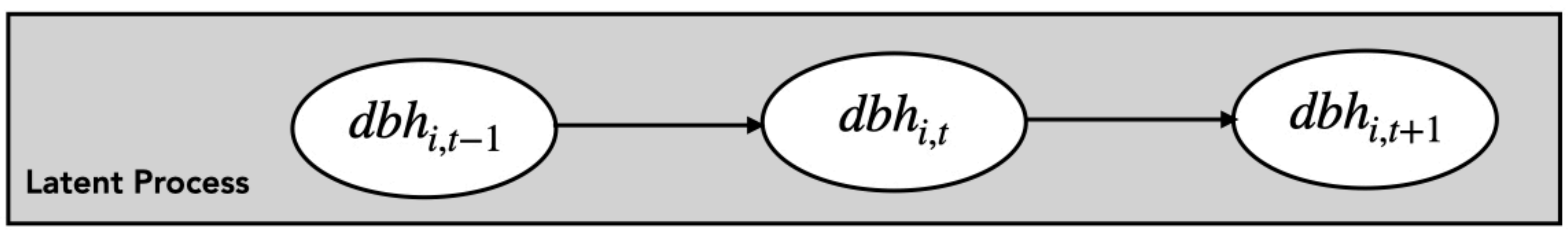
- $dbh_{i,t}$: "True" latent dbh for individual i at time t
- β_0 : Baseline growth
- β_i : Individual tree i random effect
- β_t : Time point t random effect

Minimum Viable Product

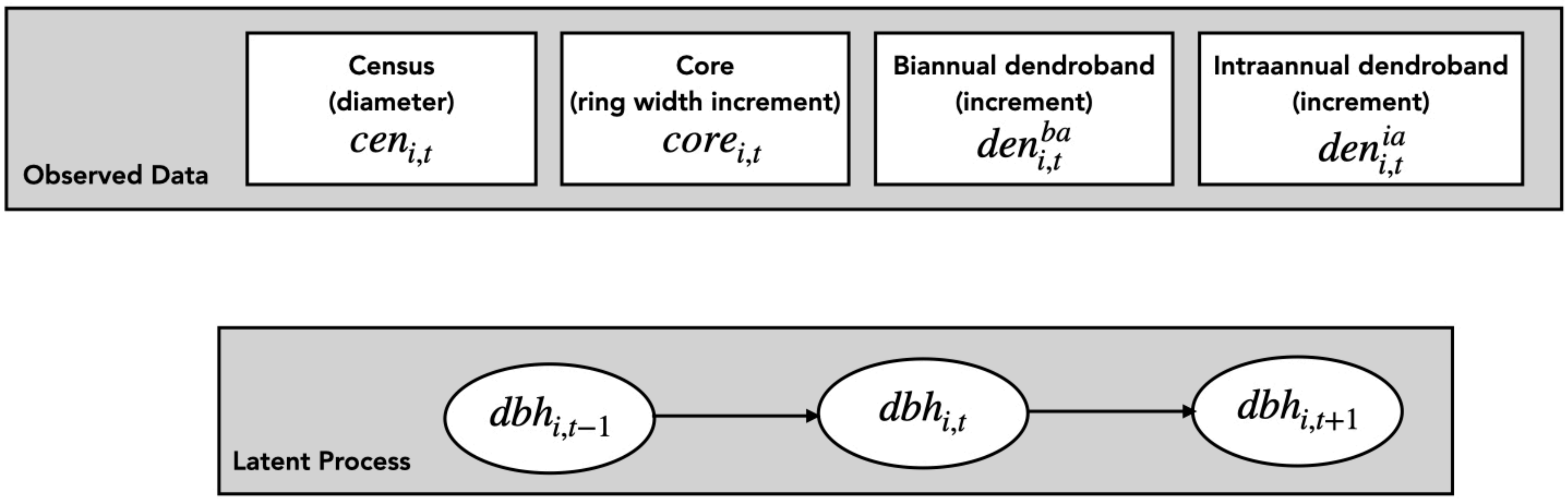
$$dbh_{i,t} = dbh_{i,t-1} + \beta_0 + \beta_i + \beta_t + \epsilon$$

- $dbh_{i,t}$: "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
- $\epsilon \sim$ Normal $\left(0, \sigma_\epsilon^2\right)$

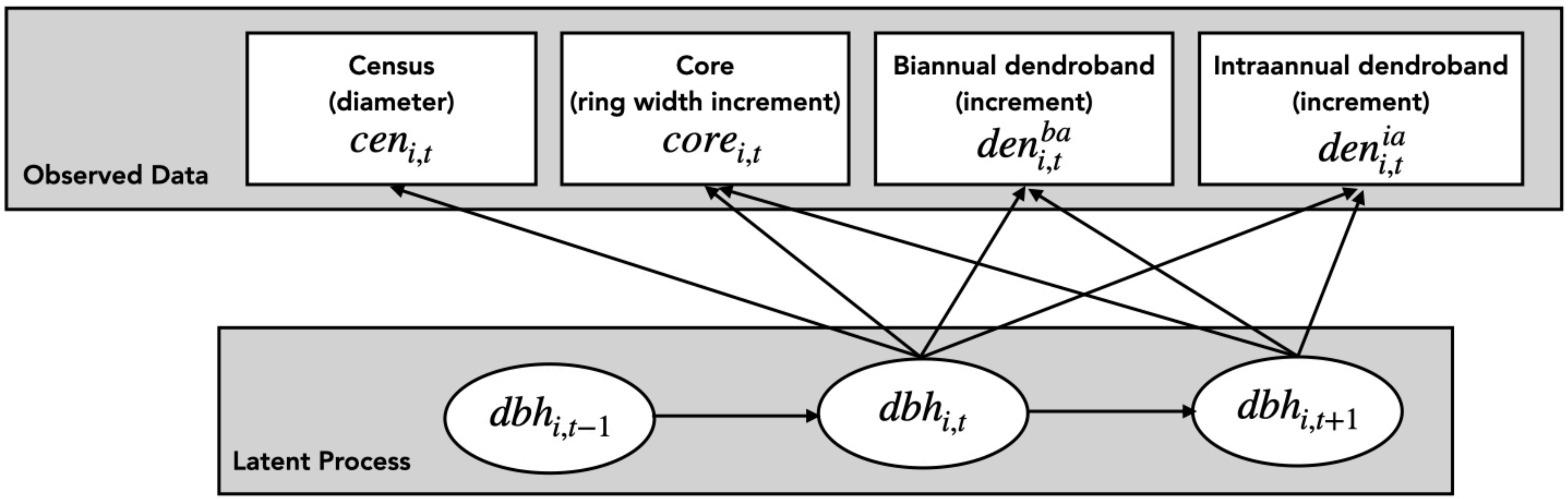
Model



Model

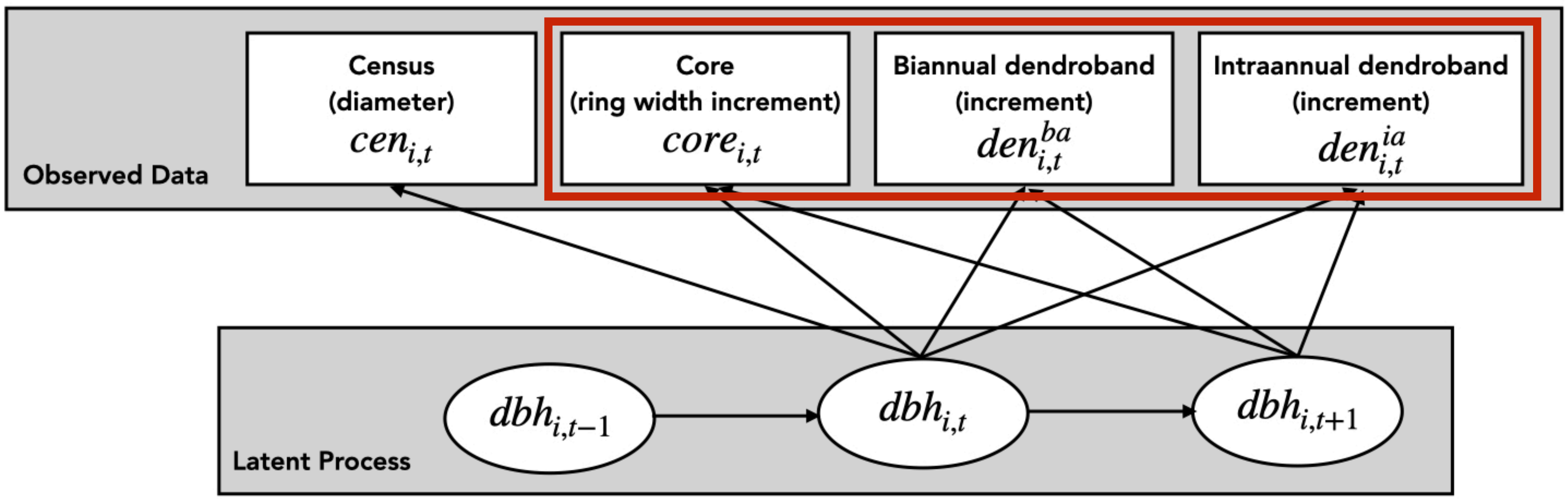

Census
(diameter)
$cen_{i,t}$
Core
(ring width increment)
$core_{i,t}$
Biannual dendroband
(increment)
$den^{ba}_{i,t}$
Intraannual dendroband
(increment)
$den^{ia}_{i,t}$
Observed Data
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Latent Process

Model



Census
(diameter)
$cen_{i,t}$
Core
(ring width increment)
$core_{i,t}$
Biannual dendroband
(increment)
$den^{ba}_{i,t}$
Intraannual dendroband
(increment)
$den^{ia}_{i,t}$
Observed Data
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Latent Process

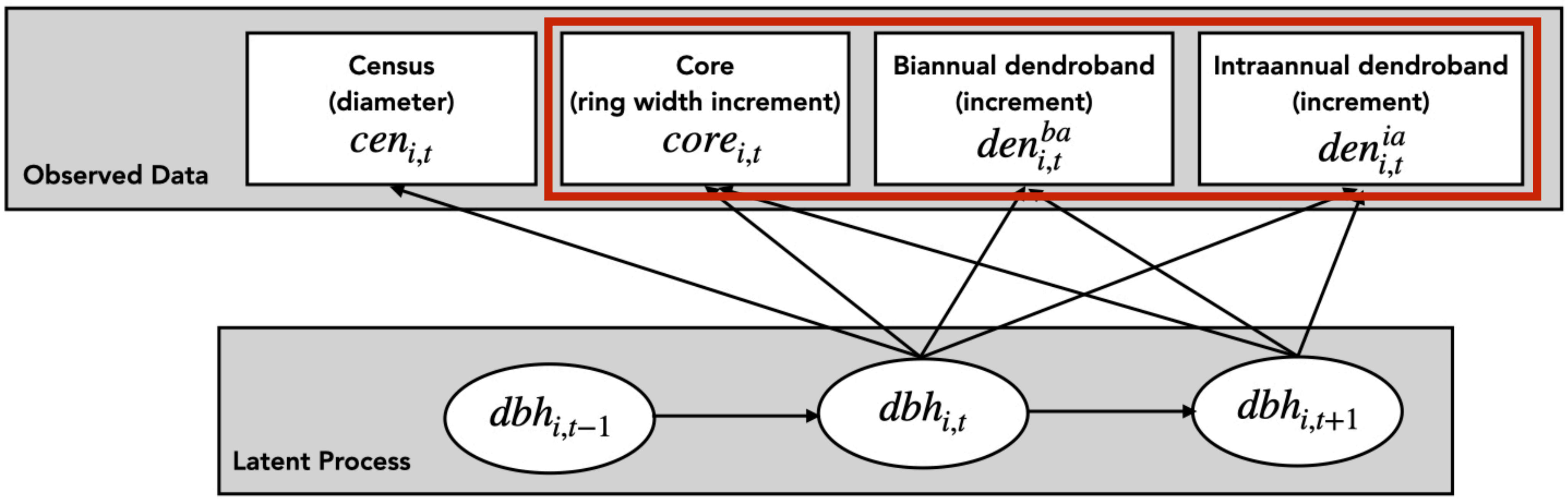
Model



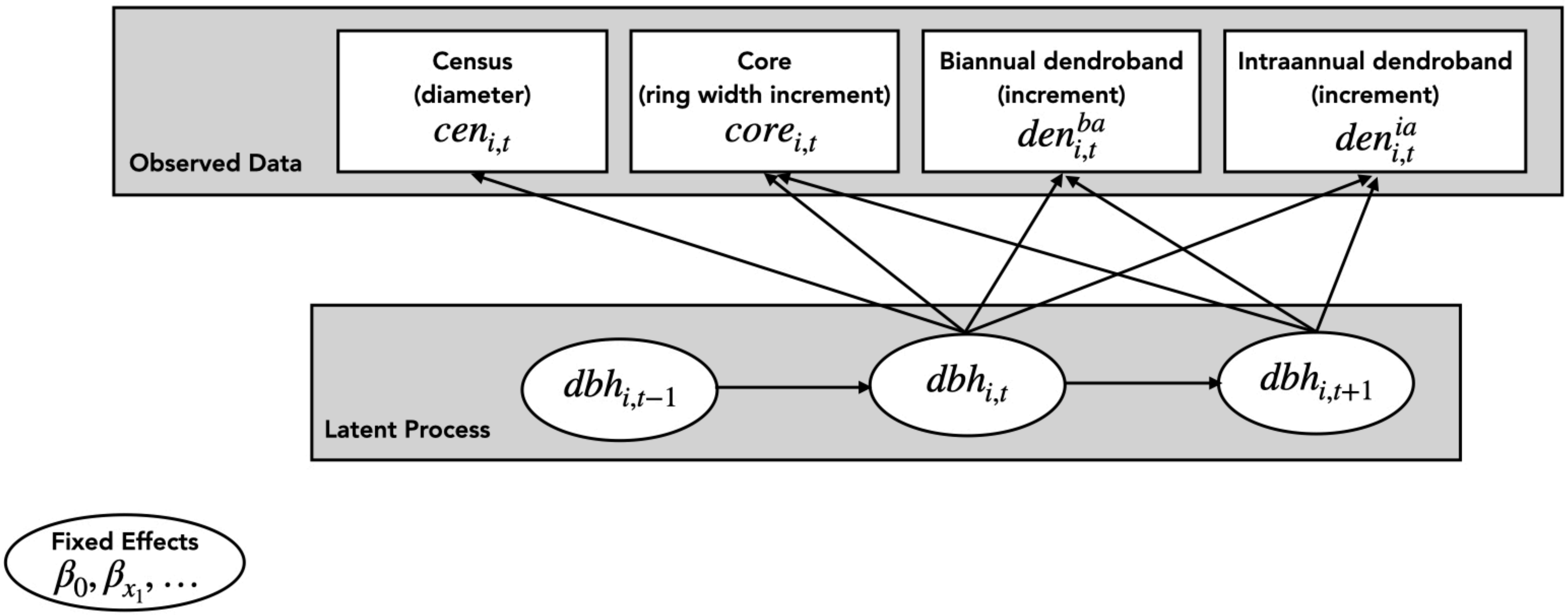
Census
(diameter)
$cen_{i,t}$
Core
(ring width increment)
$core_{i,t}$
Biannual dendroband
(increment)
$den^{ba}_{i,t}$
Intraannual dendroband
(increment)
$den^{ia}_{i,t}$
Observed Data
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Latent Process

Model

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

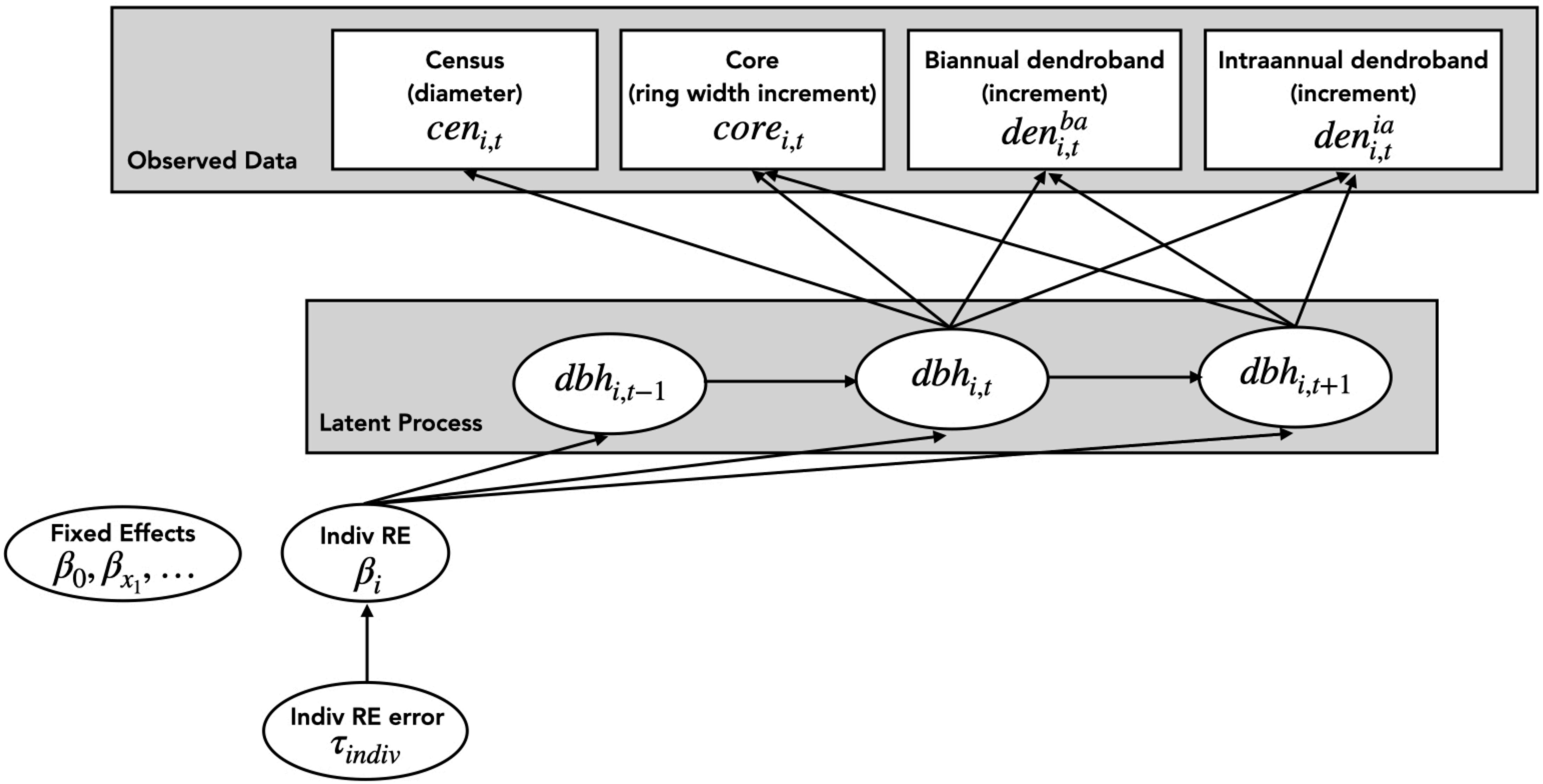


Model

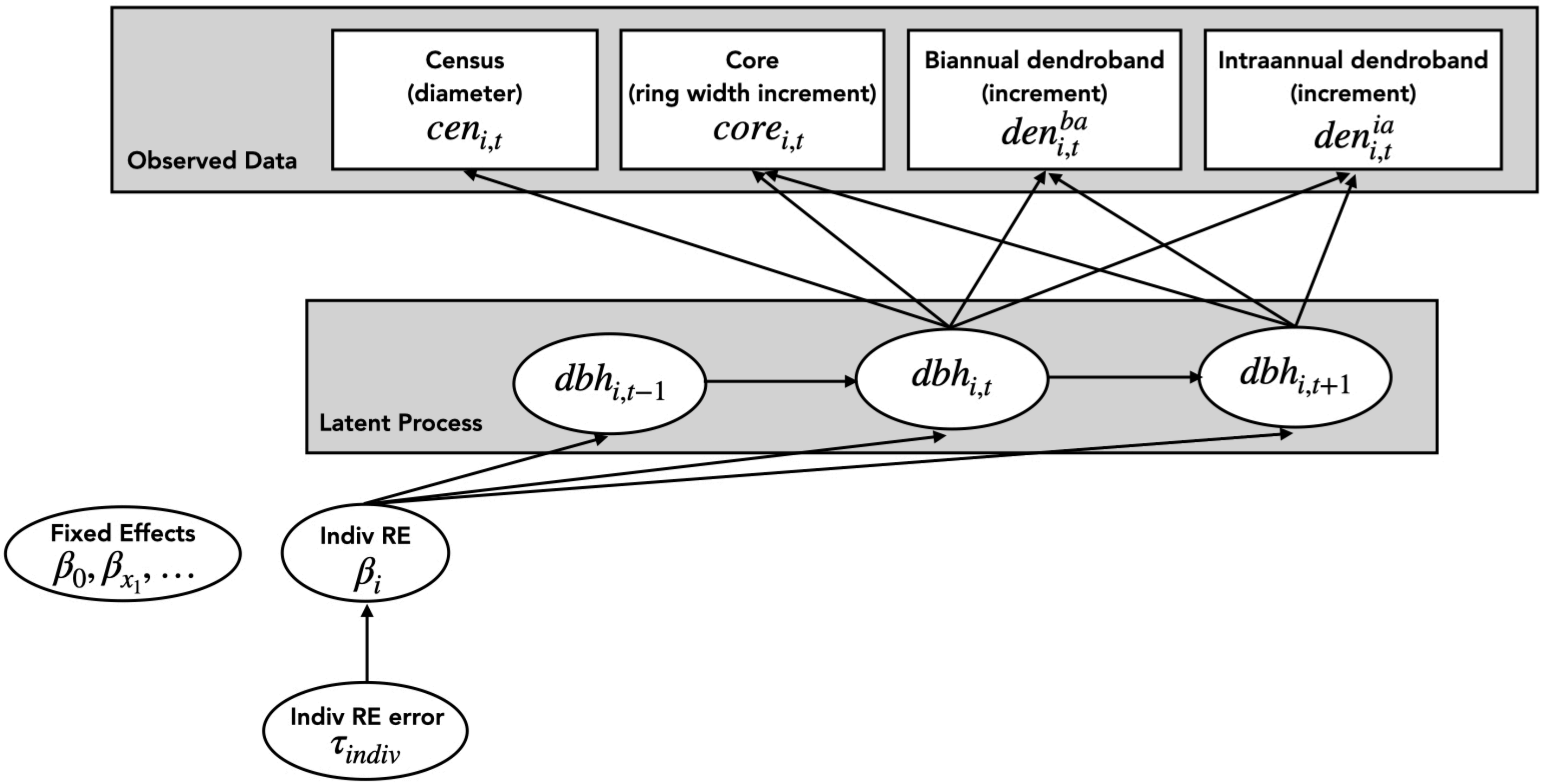


Census
(diameter)
$cen_{i,t}$
Core
(ring width increment)
$core_{i,t}$
Biannual dendroband
(increment)
$den_{i,t}^{ba}$
Intraannual dendroband
(increment)
$den_{i,t}^{ia}$
Observed Data
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Latent Process
Fixed Effects
$\beta_0, \beta_{x_1}, \ldots$

Model

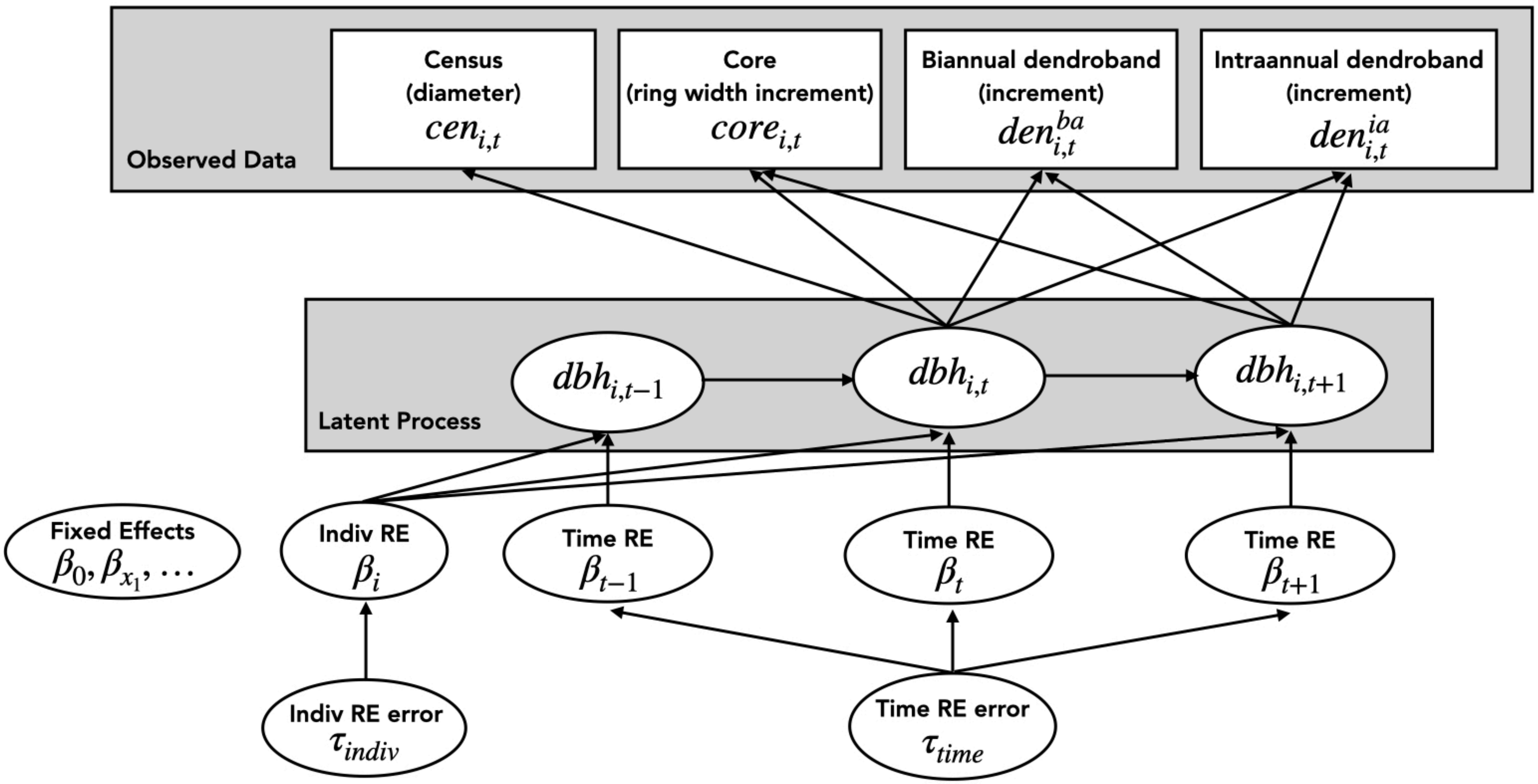

Observed Data
Census
(diameter)
$cen_{i,t}$
Core
(ring width increment)
$core_{i,t}$
Biannual dendroband
(increment)
$den_{i,t}^{ba}$
Intraannual dendroband
(increment)
$den_{i,t}^{ia}$
Latent Process
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Fixed Effects
$\beta_0, \beta_{x_1}, \ldots$
Indiv RE
β_i
Indiv RE error
τ_{indiv}

Model



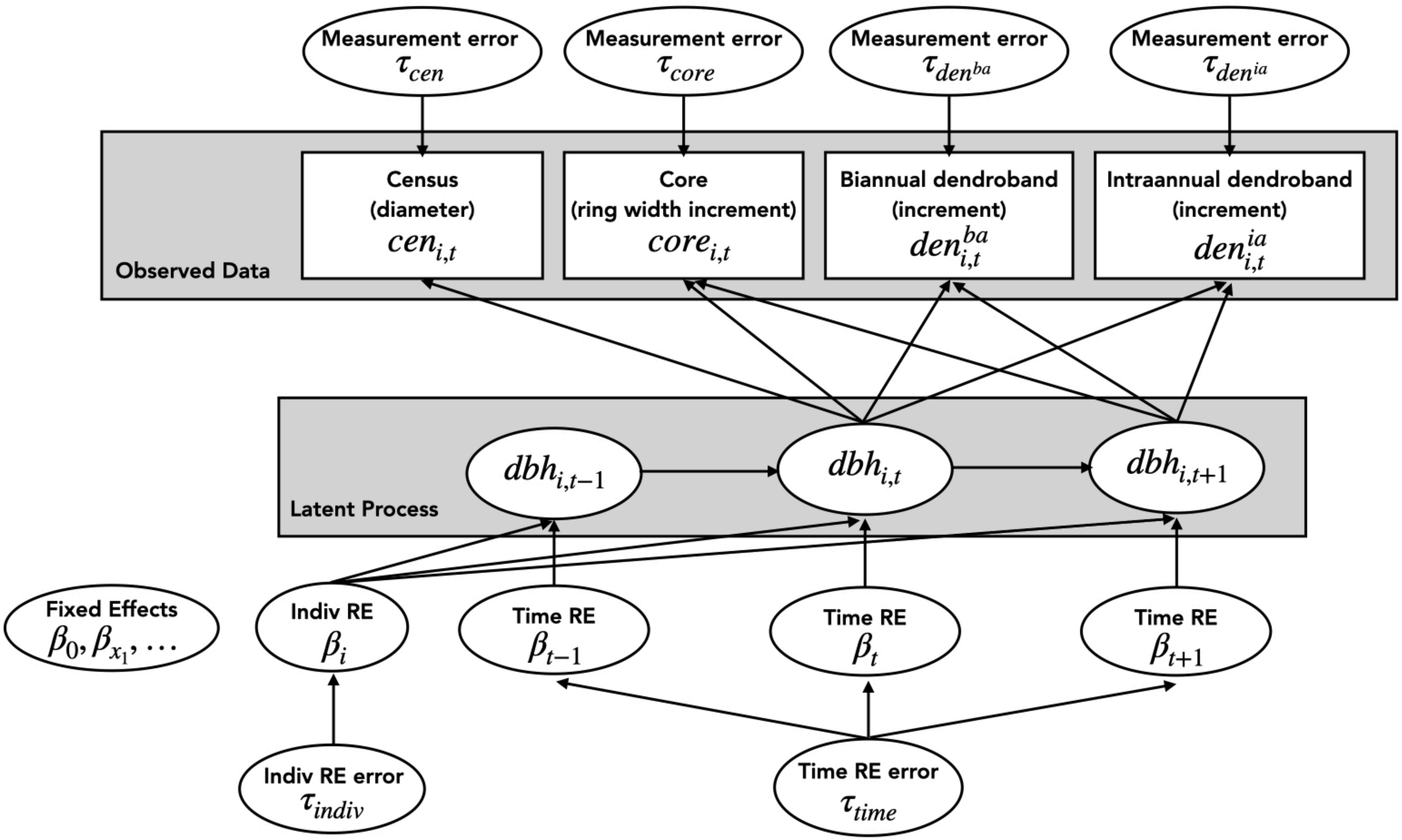
FYI: Express variances via precision $\tau = \dfrac{1}{\sigma^2}$

Model

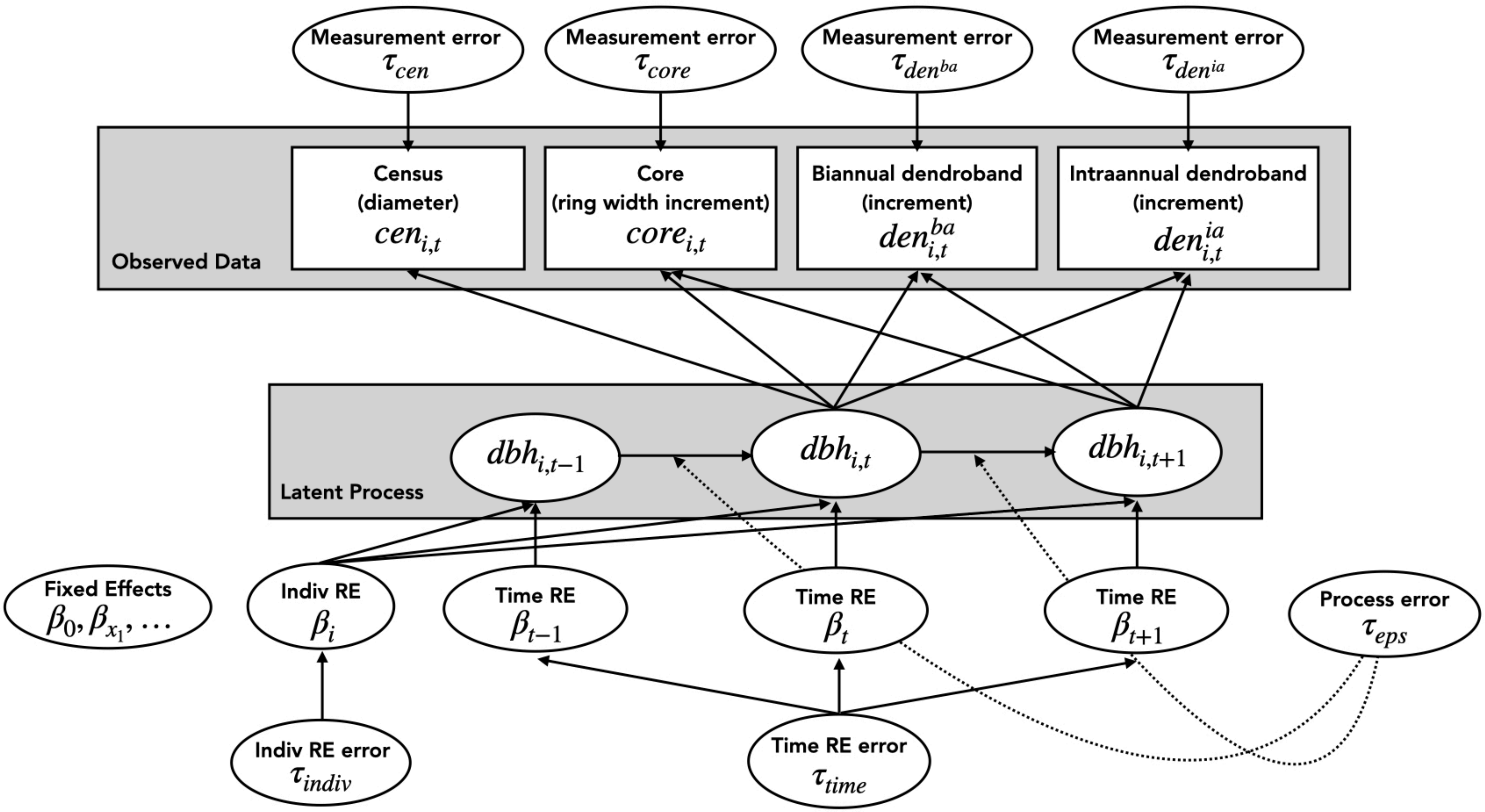


Observed Data
Census (diameter) $cen_{i,t}$
Core (ring width increment) $core_{i,t}$
Biannual dendroband (increment) $den^{ba}_{i,t}$
Intraannual dendroband (increment) $den^{ia}_{i,t}$
Latent Process
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Fixed Effects $\beta_0, \beta_{x_1}, \ldots$
Indiv RE β_i
Time RE β_{t-1}
Time RE β_t
Time RE β_{t+1}
Indiv RE error τ_{indiv}
Time RE error τ_{time}

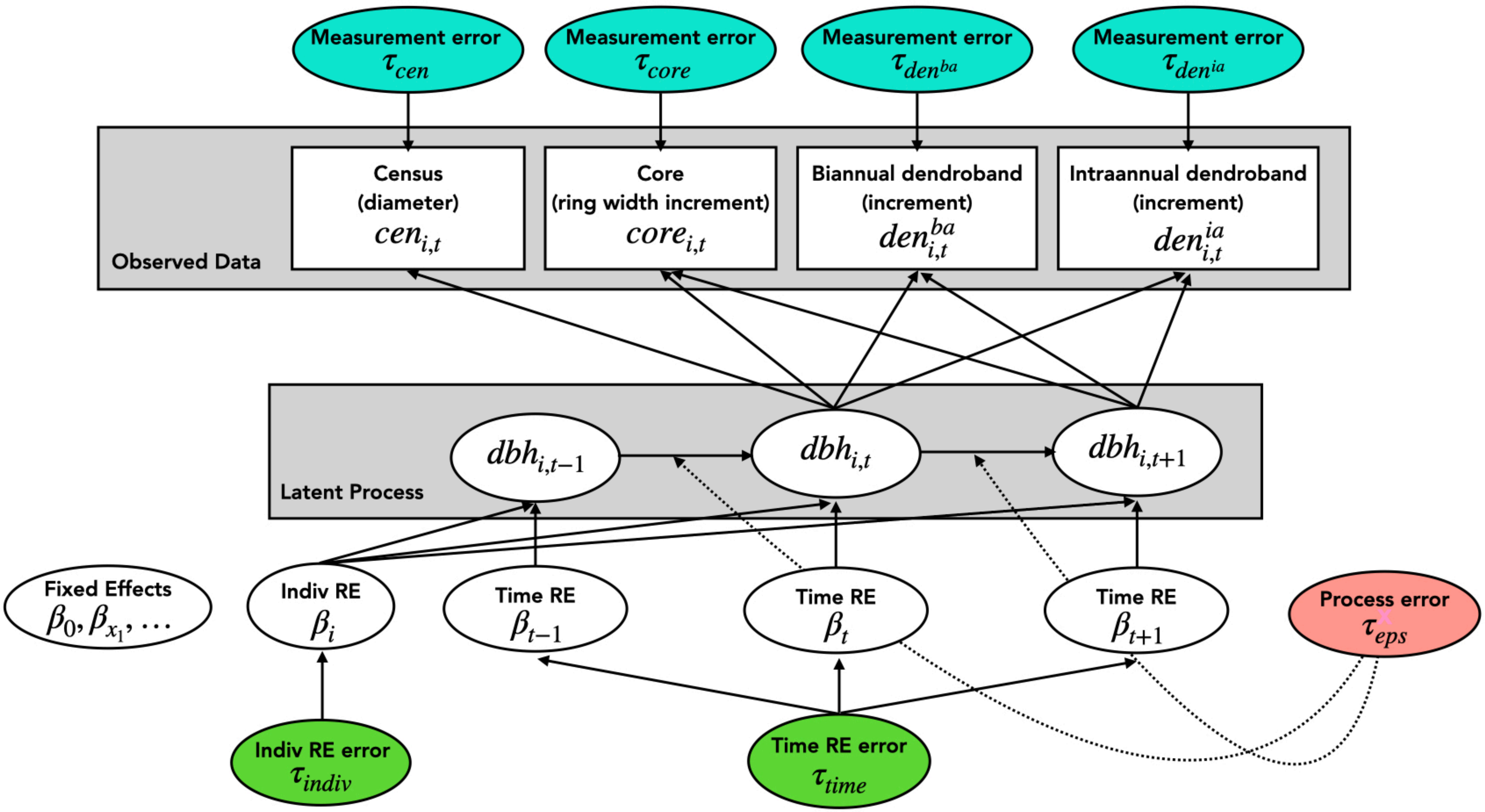
Model


Measurement error τ_{cen}
Measurement error τ_{core}
Measurement error $\tau_{den^{ba}}$
Measurement error $\tau_{den^{ia}}$
Census (diameter) $cen_{i,t}$
Core (ring width increment) $core_{i,t}$
Biannual dendroband (increment) $den^{ba}_{i,t}$
Intraannual dendroband (increment) $den^{ia}_{i,t}$
Observed Data
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Latent Process
Fixed Effects $\beta_0, \beta_{x_1}, \ldots$
Indiv RE β_i
Time RE β_{t-1}
Time RE β_t
Time RE β_{t+1}
Indiv RE error τ_{indiv}
Time RE error τ_{time}

Model

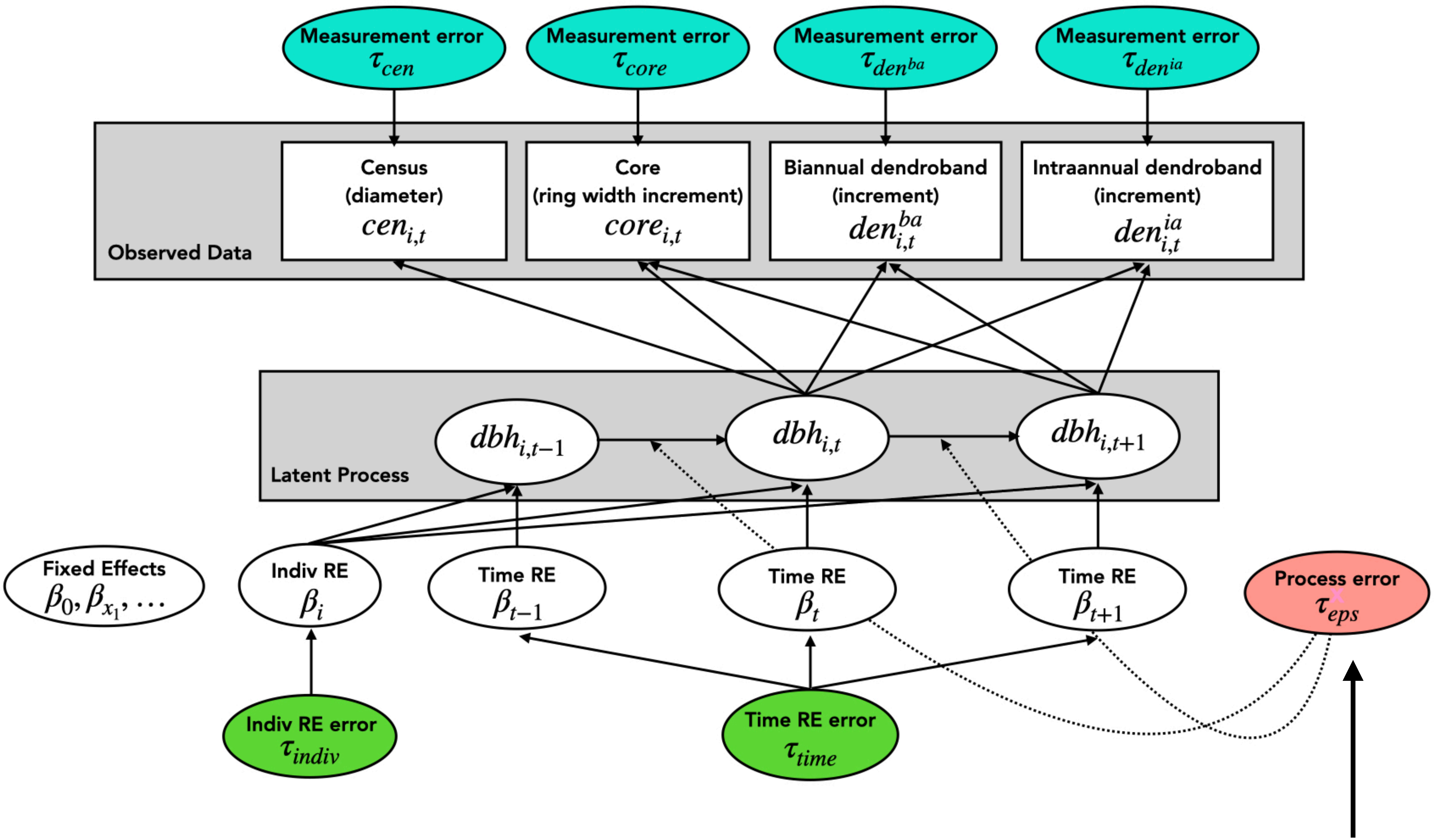

Measurement error τ_{cen}
Measurement error τ_{core}
Measurement error $\tau_{den^{ba}}$
Measurement error $\tau_{den^{ia}}$
Census (diameter) $cen_{i,t}$
Core (ring width increment) $core_{i,t}$
Biannual dendroband (increment) $den^{ba}_{i,t}$
Intraannual dendroband (increment) $den^{ia}_{i,t}$
Observed Data
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Latent Process
Fixed Effects $\beta_0, \beta_{x_1}, \ldots$
Indiv RE β_i
Time RE β_{t-1}
Time RE β_t
Time RE β_{t+1}
Process error τ_{eps}
Indiv RE error τ_{indiv}
Time RE error τ_{time}

Model



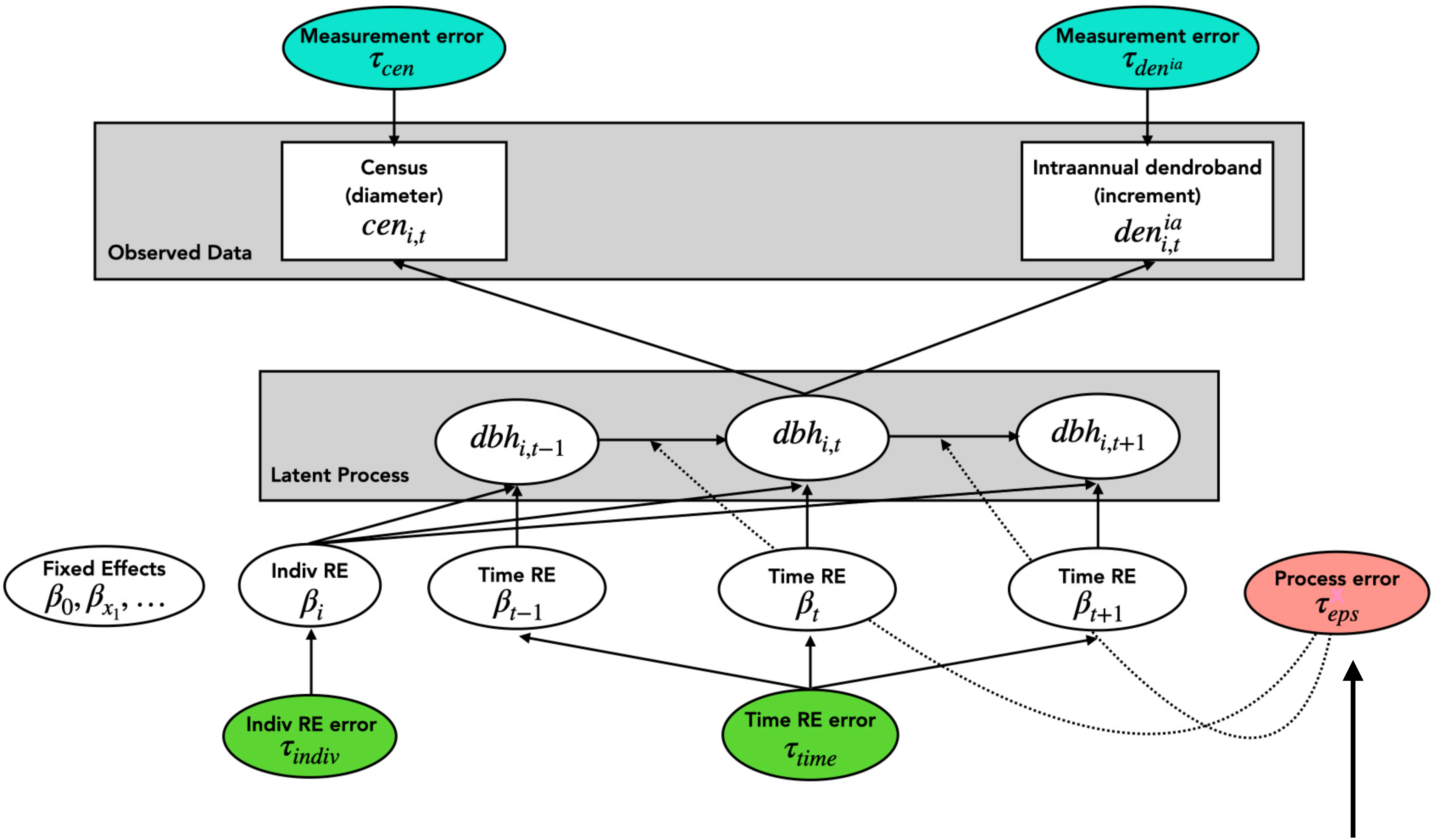
Measurement error
τ_{cen}
Measurement error
τ_{core}
Measurement error
$\tau_{den^{ba}}$
Measurement error
$\tau_{den^{ia}}$
Census
(diameter)
$cen_{i,t}$
Core
(ring width increment)
$core_{i,t}$
Biannual dendroband
(increment)
$den^{ba}_{i,t}$
Intraannual dendroband
(increment)
$den^{ia}_{i,t}$
Observed Data
$dbh_{i,t-1}$
$dbh_{i,t}$
$dbh_{i,t+1}$
Latent Process
Fixed Effects
$\beta_0, \beta_{x_1}, \ldots$
Indiv RE
β_i
Time RE
β_{t-1}
Time RE
β_t
Time RE
β_{t+1}
Process error
τ_{eps}
Indiv RE error
τ_{indiv}
Time RE error
τ_{time}

Model



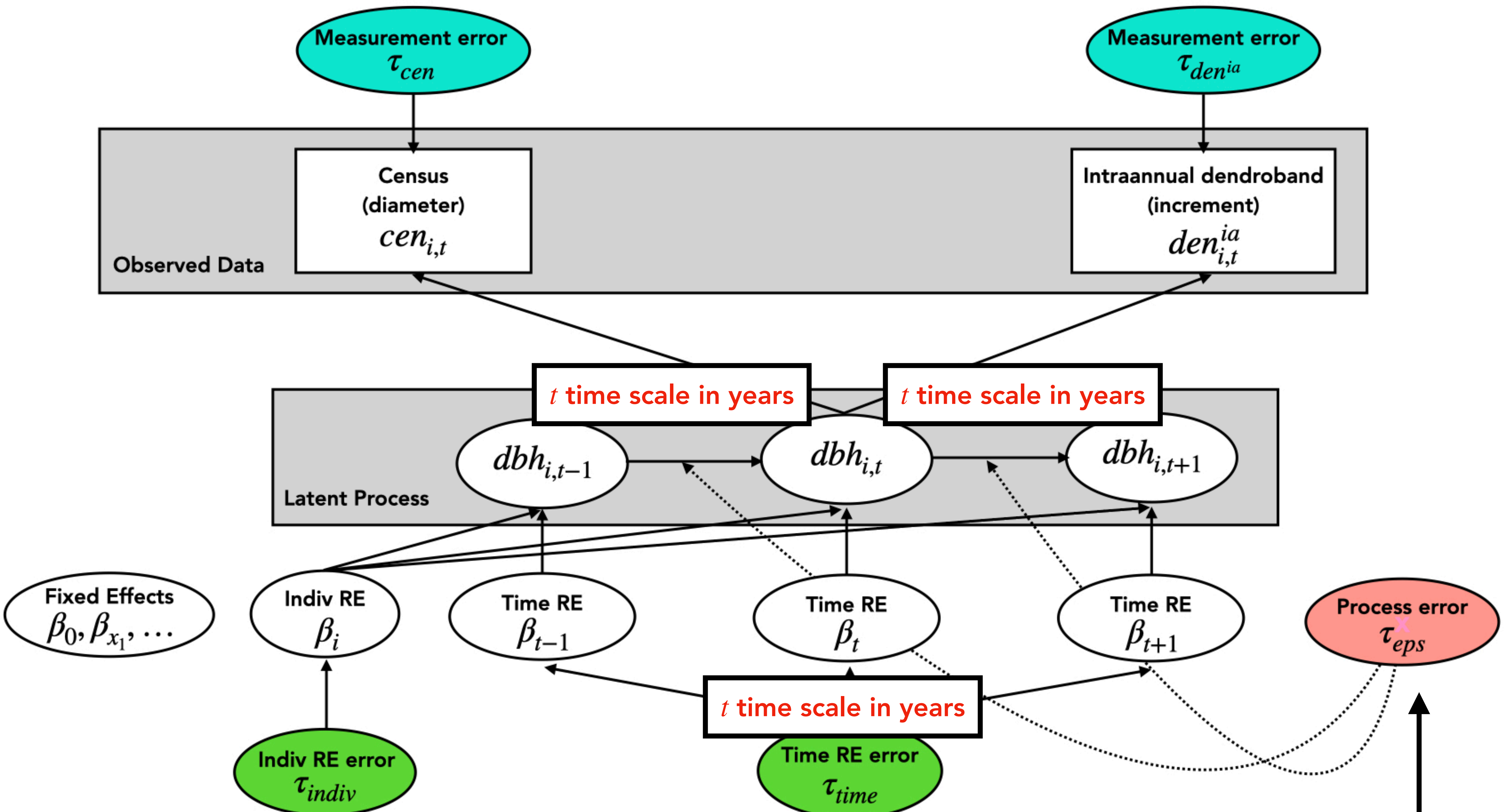
Moral: only error this propagates across time in forecasts

Model as of 2021/1/22



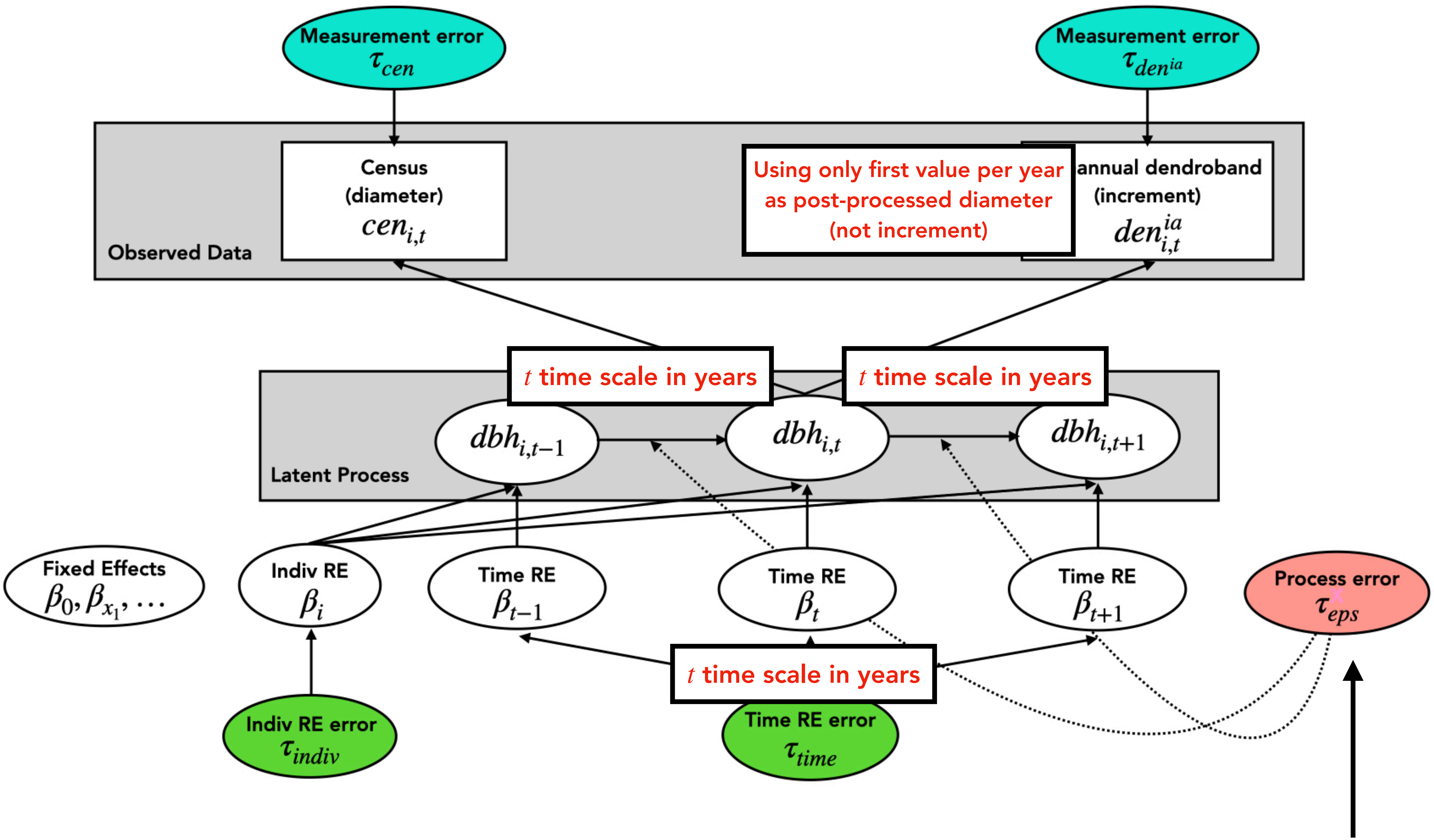
Moral: only error this propagates across time in forecasts

Model as of 2021/1/22



Moral: only error this propagates across time in forecasts

Model as of 2021/1/22



Moral: only error this propagates across time in forecasts

Results

MCMC specifications

MCMC specifications

- n = 155 trees with intraannual dendrobands (tend to be larger canopy trees)

MCMC specifications

- n = 155 trees with intraannual dendrobands (tend to be larger canopy trees)
- Implemented in JAGS

MCMC specifications

- n = 155 trees with intraannual dendrobands (tend to be larger canopy trees)
- Implemented in JAGS
- 30k draws from posterior minus 10% burn-in

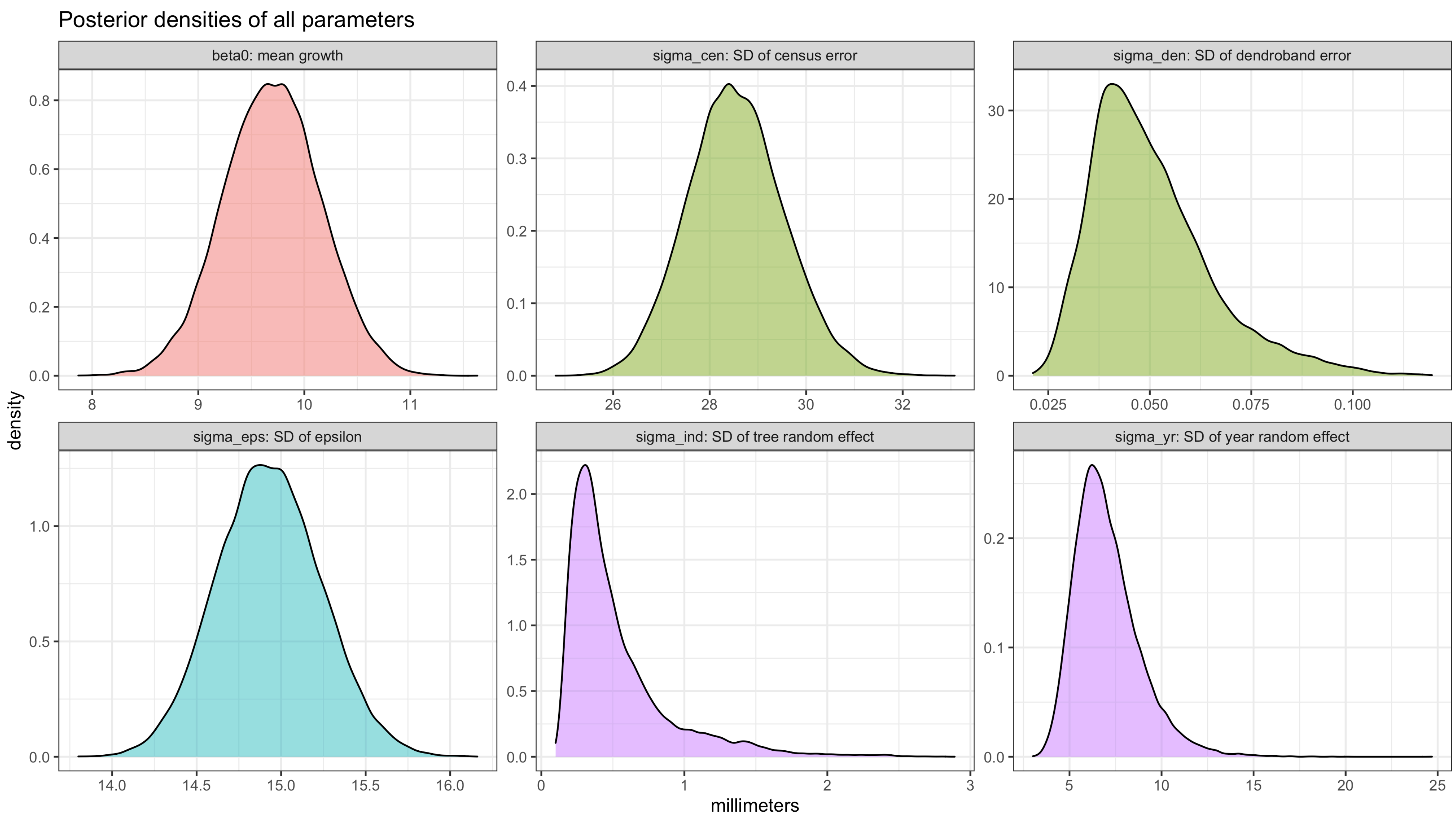
MCMC specifications

- n = 155 trees with intraannual dendrobands (tend to be larger canopy trees)
- Implemented in JAGS
- 30k draws from posterior minus 10% burn-in
- Empirical Bayes (data informed) prior parameters

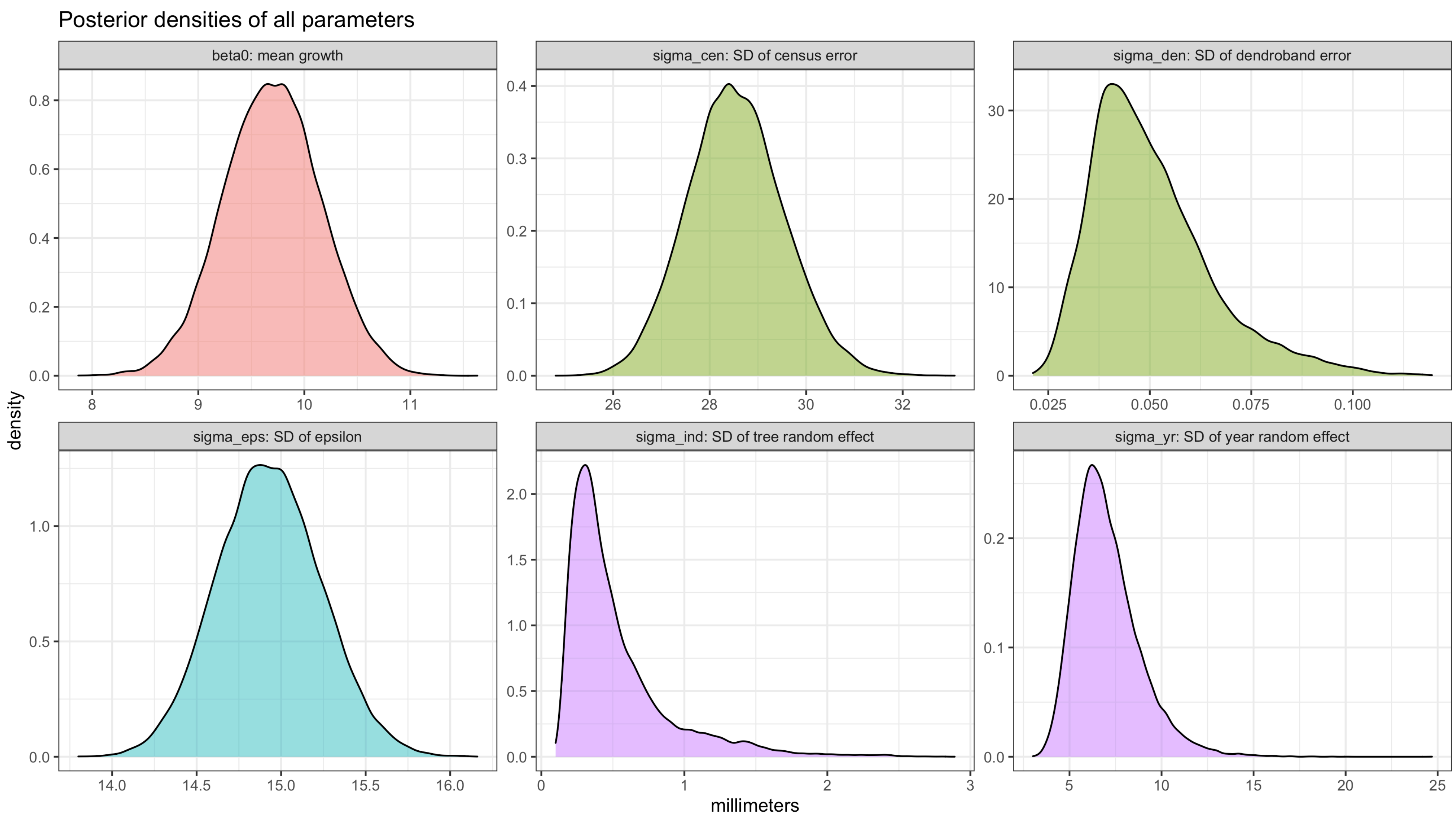
MCMC specifications

- n = 155 trees with intraannual dendrobands (tend to be larger canopy trees)
- 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

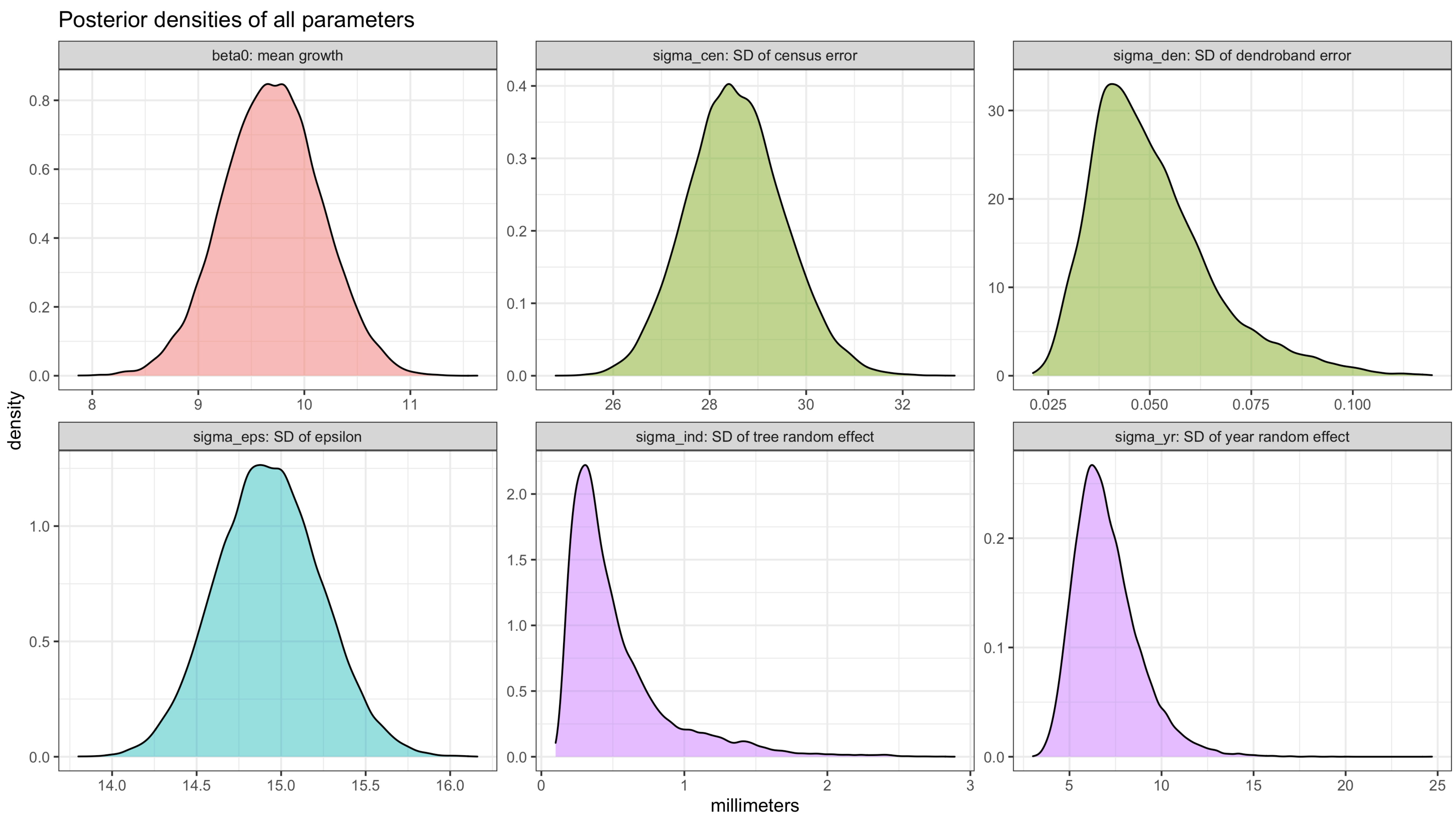


Posterior Distributions



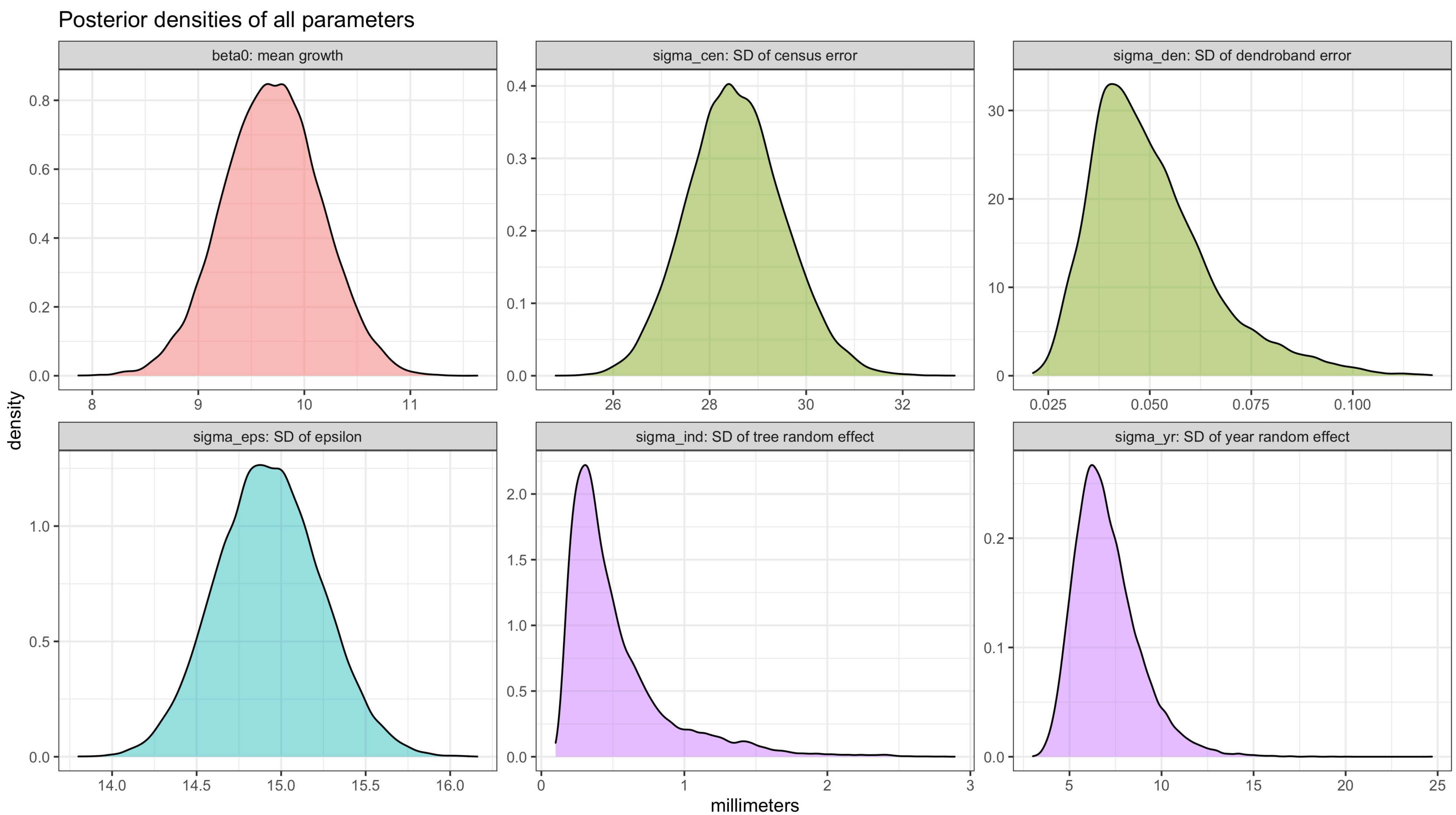
- mean β_0 ~ 1cm growth per year (needs sanity checking)

Posterior Distributions



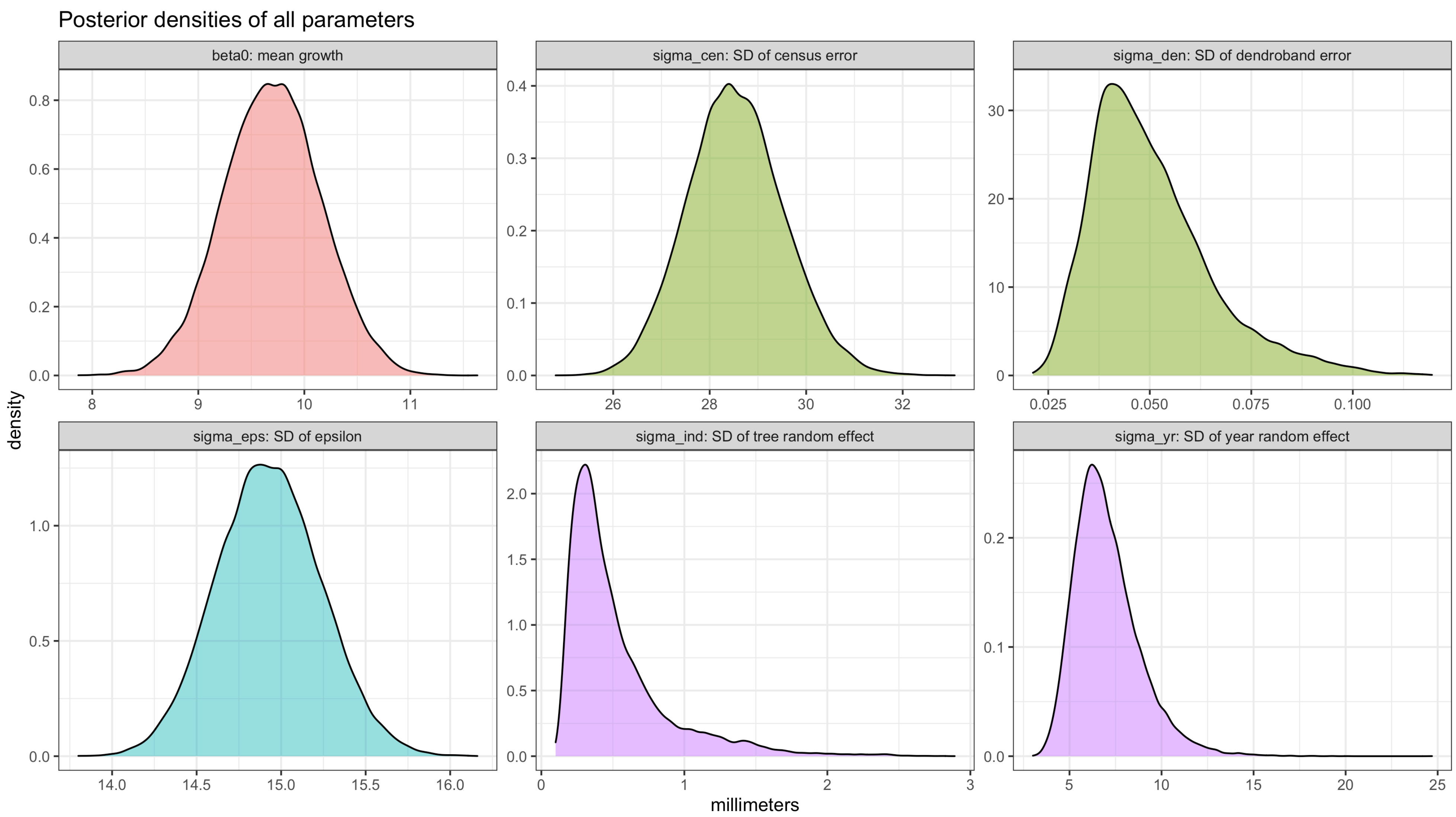
- mean β_0 ~ 1cm growth per year (needs sanity checking)
- $\sigma_{cen} >> \sigma_{den}$

Posterior Distributions



- mean β_0 ~ 1cm growth per year (needs sanity checking)
- $\sigma_{cen} >> \sigma_{den}$
- Year-to-year variation in growth > Between individuals variation

Posterior Distributions



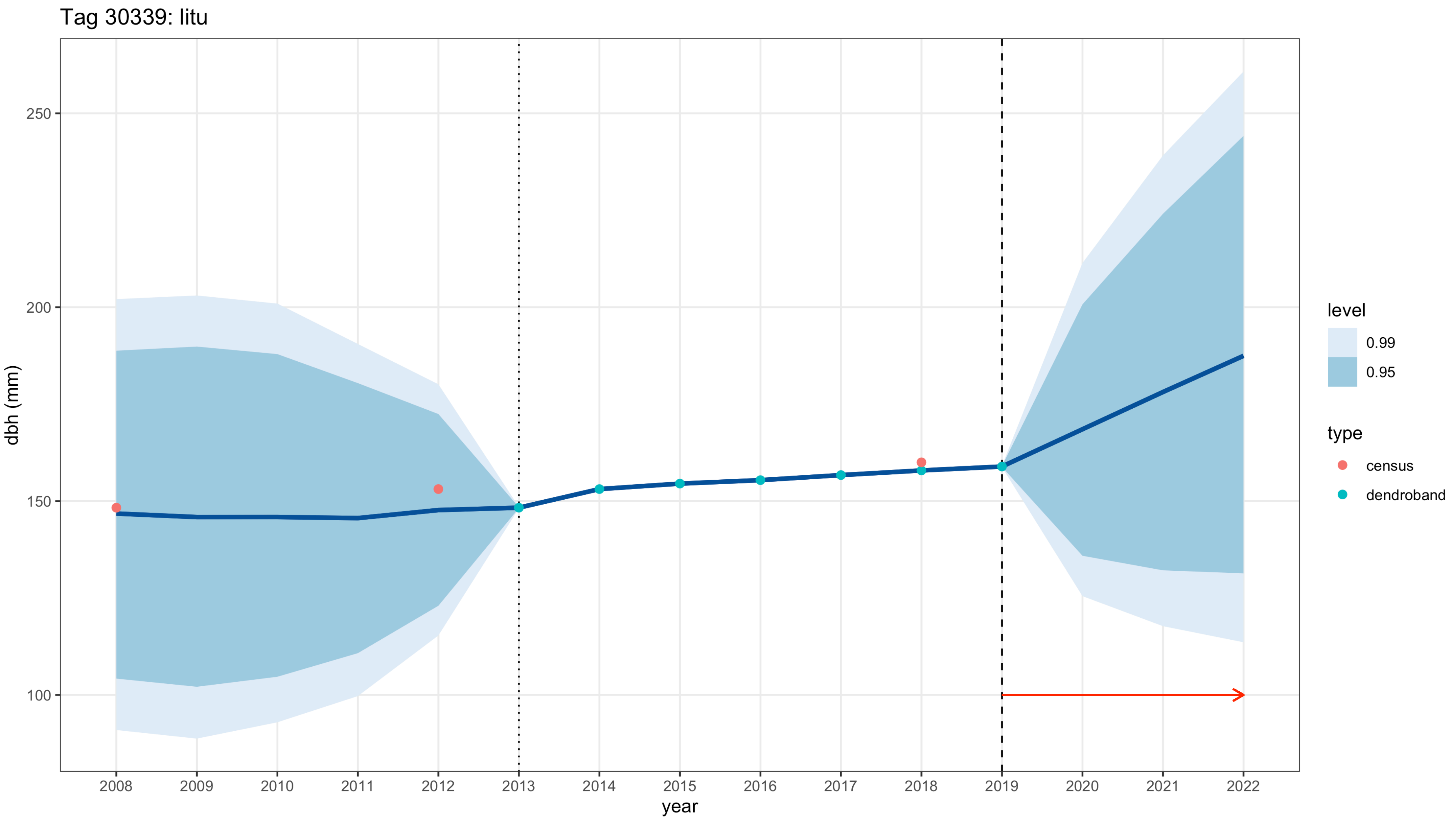
- mean β_0 ~ 1cm growth per year (needs sanity checking)
- $\sigma_{cen} >> \sigma_{den}$
- Year-to-year variation in growth > Between individuals variation
- σ_ϵ = remaining process error that propagates in forecasts across time

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>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<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 = modeled true latent $dbh_{i,t}$



Dendroband installed in 2013

Future Work

TODO

TODO

- Add remaining data:

TODO

- Add remaining data:
  - All, not just first yearly observation, intra-annual dendroband

TODO

- Add remaining data:
  - All, not just first yearly observation, intra-annual dendroband
  - All biannual dendroband

TODO

- Add remaining data:
  - All, not just first yearly observation, intra-annual dendroband
  - All biannual dendroband
  - Tree coring data

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}?

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

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

TODO

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

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
  - $dbh_{i,t} = dbh_{i,t-1} + \beta_0 + \ldots + \epsilon$
  - Covariates: In particular species & starting diameter
- Choose appropriate time scale for t

Thanks!

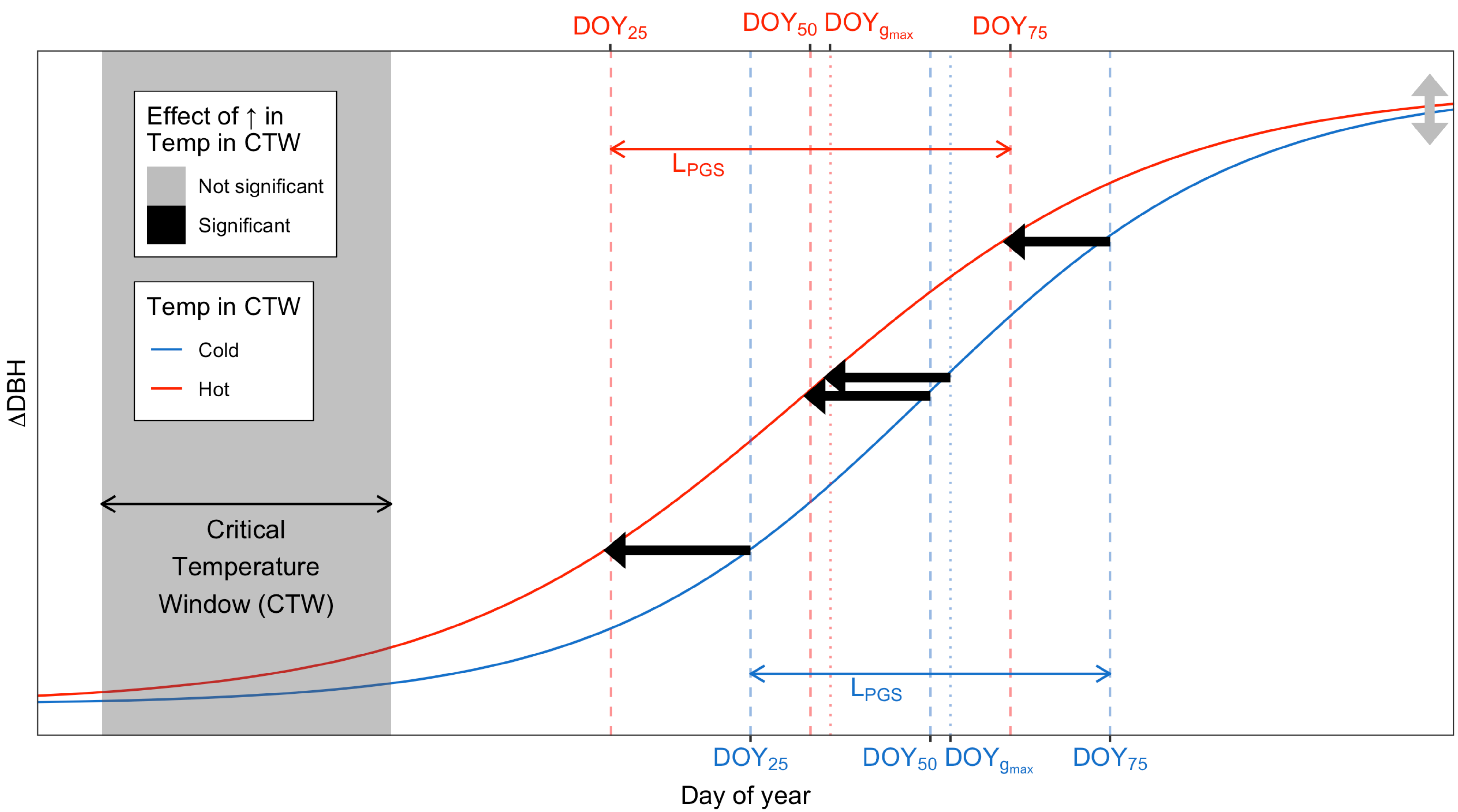
Slides on Twitter @rudeboybert

Nature doesn't always play nice

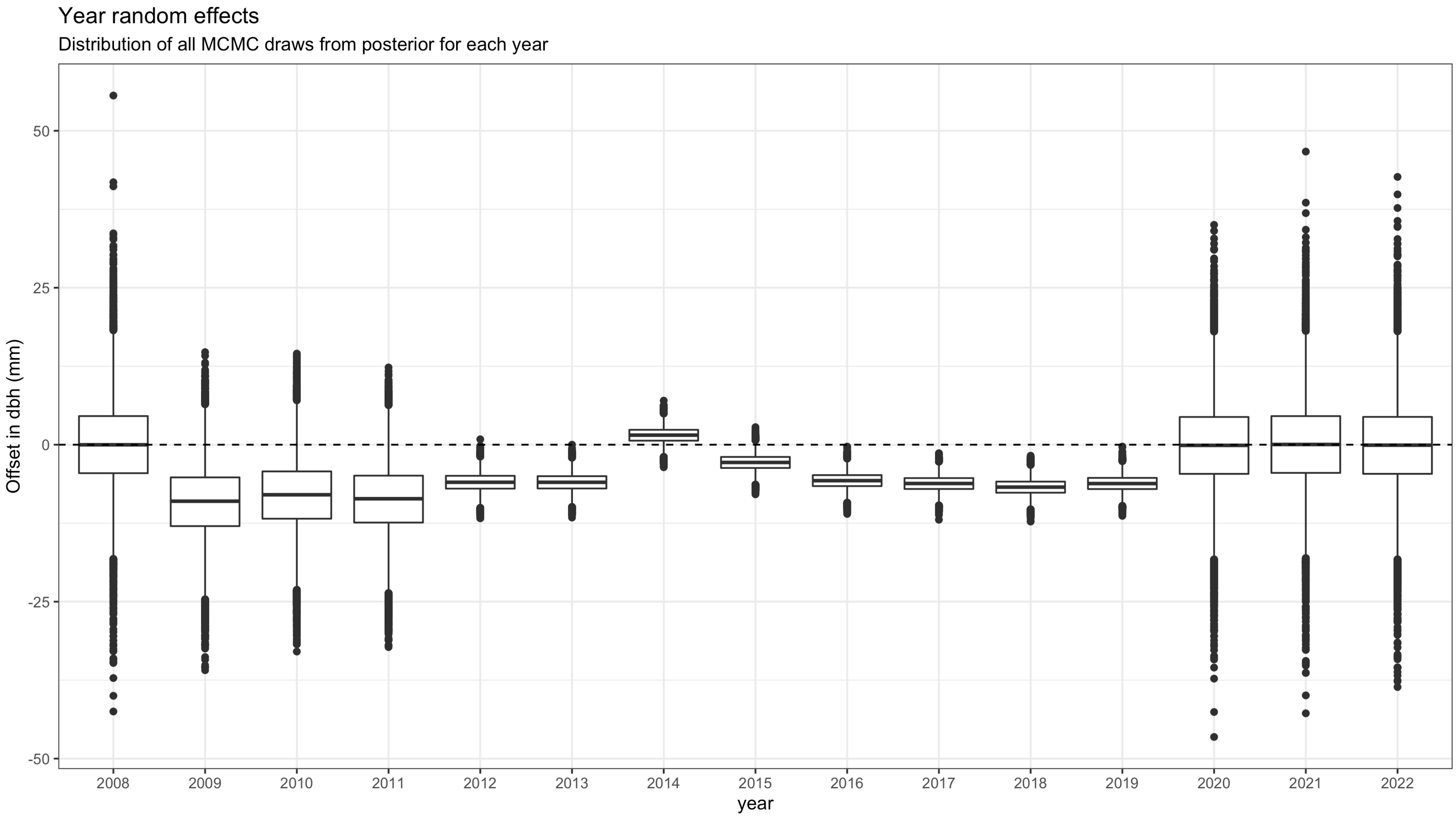
Nature doesn't always play nice

Thus could biannual dendroband data be more error prone than intraannual?

Intra-annual effect of climate



Year Random Effects



Individual Random Effects