Multilevel Urban Tree Allometric equations

erker

November 9, 2018

Introduction

Multilevel models have been used for decades in tree growth equations (Lappi and Bailey 1988). Multilevel modeling is an attractive approach because it provides a coherent framework to account for the many levels of observation or of groupings in data and to pool information across groups. This paper has two main contributions. First, we demonstrate the use of Stan via the "brms" package in R to fit bayesian nonlinear multilevel models to predict tree diameter growth from age Carpenter et al. (2017); Bürkner (2017). Second, we apply the method to the Urban Tree Database McPherson et al. (2016a,b). This dataset is the result an over a decade long effort to collect age and size data on thousands of trees in 17 cities across the US. Multilevel modeling has the potential to extract more information from the data and improve predictions compared to the existing modeling approach. Improving predictions of tree size from tree age will improve our ability to predict the important ecosystem services these trees provide urban dwellers.

Stan is a probabilistic programming language for bayesian inference Carpenter et al. (2017). It uses No-U-Turn sampler, an adaptive form of Hamiltonian Monte Carlo sampling, to effectively draw samples from the specified log posterior density. Here, we access Stan

via the R package brms Bürkner (2017). brms allows the user to specify the likelihood and priors in syntax similar to the R package lme4 commonly used for frequentist? multilevel (mixed effects) models. This makes harnessing the power of Stan much simpler and concise because it doesn't require the user to know how to write efficient Stan code and can convert a few lines of R code into many lines of Stan. brms is not as flexible as stan, but still can be used to fit many types of models including nonlinear multilevel regression models, such as ours here. Some of the key advantages of fitting a model in Stan via brms include relatively simple syntax and efficient posterior sampling for multilevel non-linear models. The bayesian approach gives better estimates of parameter uncertainty and provides a formal way to include prior information.

The existing approach to modeling the diameter growth of trees in the urban tree database (UTD) was to fit a separate model for each tree species in each city and test several model forms with different weights and then select the model with the lowest Akaike Information Criterion (AIC). This approach has several limitations, many highlighted in the report. First, while the model form selected provided the lowest AIC, many of the estimates are not biologically realistic (for example they begin to increase sharply at old ages, cubic fits, or decrease at old ages, quadratic fits). Therefore, the researchers discouraged applying the models beyond the range of the data, or sometimes even within the range of data if the estimates were unrealistic. These unrealistic estimates and the inability to extrapolate severely limits managers' ability to predict growth over meaningful time scales (a century rather than a few decades). A second limitations is that some models predict negative diameters, an impossibility. Third, models are only provided for the cities and the species sampled. If a manager wants to predict the diameter growth of a tree species in an unsampled city, the researchers recommend using the model from the reference city in the same climate region. However many of the reference cities are on the border of climate regions and there is known large variability in growth within regions (see figure XX in utd report which is from McPherson.... comparing Cheyenne to Ft Collins). Furthermore, if a manager wants to

predict growth for an unsampled species or a species that was sampled in a different city it is not obvious which equation/model to use and the additional uncertainty that this introduces is not quantified.

Our approach addresses the above limitations. First, we use a weibull curve, commonly used in foresty growth equations and biologically realistic, which makes extrapolation to ages outside the data range less fraught. Second, using this sigmoidal curve and modeling diameter with a gamma distribution ensures our estimates of diameter are positive. Third, by modeling the weibull curve parameters as functions of species, city, and climate, we are able to borrow information across cities and across species to provide predictions an associated uncertainty of diameter growth even in even in cities or species with very little or no data.

Sigmoidal curves very similar to the weibull we use here have been used before in modeling urban tree diameter growth as a function of age. Frelich (1992) use the Chapman-Richards growth curve of form $y = B_0(1 - exp(B_1 x))^{B_2}$ to predict DBH from age for healthy trees (12 species, 221 trees total) in Minneapolis and St. Paul, Minnesota. This equation form worked very well (8 out of 12 species had an R^2 over .9), but the trees used were only healthly open grown trees, which is not representative of urban trees generally. Following Frelich (1992), in an early version of the urban tree database, McPherson and Simpson (1999) fit the same curve to a small number of observations and adjusted parameters for different locations based on the number of frost free days based on expert opinion. Peper et al. (2001a) and Peper et al. (2001b) compared this modified weibull curve to logarithm regression and selected the logarithm regression based on a higher R^2. Subsequent urban tree growth equations did not use sigmoidal curves McPherson et al. (2016b).

Multilevel modeling was first introduced to forestry by Lappi and Bailey (1988) and has since been widely used to account for multiple levels of variability (observations correlated within groups) in allometric and growth equations. Hall and Bailey (2001) give a general overview of the approach and an example of loblolly pine height growth in the Southeastern US. Levels in their model are plot and tree nested within plot, and they use tree density

as a plot level covariate. Nothdurft et al. (2006) provide another example of modeling tree height growth using norway spruce in Germany. They use the Sloboda function and levels in the model are plot and trees nested in plot with elevation as a plot level covariate. Li et al. (2011) provide a bayesian example modeling balsam fir height growth in Maine using the Chapman-Richards equation, also with tree nested within plot. They found the bayesian approach has similar parameter point estimates to the frequentist approach. In an urban tree context, Peper et al. (2014) model DBH growth with repeat measures data on individual trees and test a varying coefficients model because they have repeat measures on individual trees.

Indeed the test dataset, "orange", in the statistical programming language, R, is used to demonstrate the fitting of nonlinear multilevel (mixed effects) models R Core Team (2016) orange.

Compared to past work, our approach is new from a modeling standpoint in the use of a bayesian approach with both nested and non-nested groupings and group level predictors. The Bayesian multilevel/hierarchical modeling framework has many strengths as discussed in Li et al. (2011) (and others) and includes the ability to sample parameter values from the entire posterior, rather than maximum likelihood estimates. Nonlinear multilevel models in the frequentist approach depend on linear approximations such as first-order taylor expansion, which is not required in the bayesian framework. Summary statistics of parameters (mean, median, quantiles) can be easily calculated from posterior samples and the common assumption in the frequentist paradigm that parameters are normally distributed can be relaxed. The ability to incorporate prior information from experts or past studies on parameters is another strength of the bayesian approach.

A paragraph on the impact of these equations for managing forests to predict/forecast ecosystem services.

Methods

Data

The urban tree database (UTD) consists of measurements on 14487 trees of 170 species in 17 cities. However, largely because of the difficulty is measuring tree age, there are only 12687 trees with complete age and diameter data (161 species, 17 cities, 309 species by city combinations 1).

Some species were measured in multiple cities, but not most. The number of trees of each city by species combination sampled ranged from 1 (both Liquidambar styraciflu and Prunus serrulata in Queens, NY) to 79 (Quercus laurifolia in Charleston, SC). The median number of trees in each species-city combination was 37.

Age is defined in this dataset as time since planting, since this is the record kept by cities. Actual age of the trees may be several years more. Diameter (cm) of the trees is measured at breast height (1.37m above ground).

In the UTD, trees are classified taxonomically down to cultivar for some individuals, but here we aggregate cultivars up to the species level. Species are then nested within Genera.

The 17 cities in the UTD cover much of the US geographically, 2, and much of the variation in climate, 3. However, New York City only has a few observations and the data for Indianapolis is missing too.

Rather than using the aggregated sunset zones as done in UTD, we used growing degree days (GDD) and precipitation data from climate NOAA's climate normals to continuously vary equation parameters across climate. Figure 3 shows each census tract centroid in the conterminous US plotted in GDD and precipitation space. We appoximated the GDD and precipitation for each tract by assigning the values of the weather station closest to the centroid. This allows us to vary our model continouously across geographic space in a way that better captures the natural gradients of climate.

Figure 1: Number of trees sampled of each species and city combination in the urban tree database.

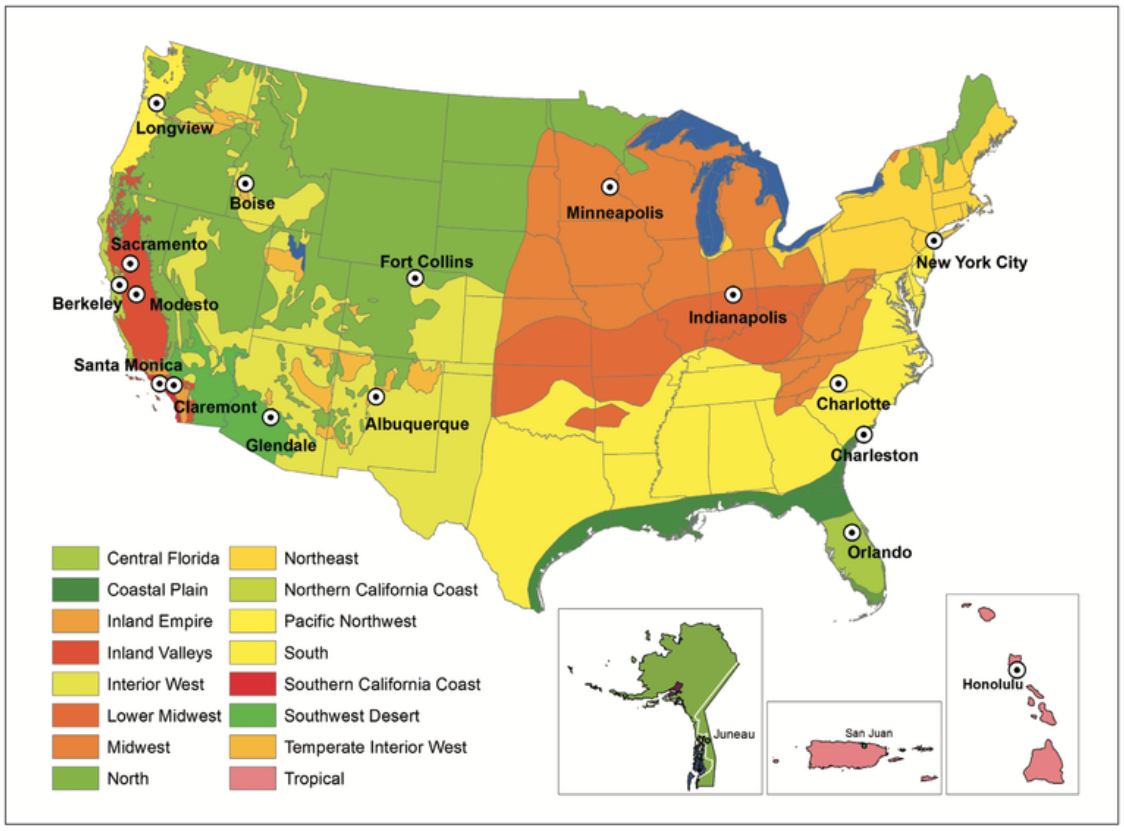


Figure 9—Climate zones were aggregated from 45 Sunset climate zones into 16 zones. Each zone has a reference city where tree growth data were collected. Sacramento, California, was added as a second reference city (with Modesto) to the Inland Valleys zone.

Figure 2: 16 climate regions and 17 representative cities in the UTD (McPherson et al., 2016b).

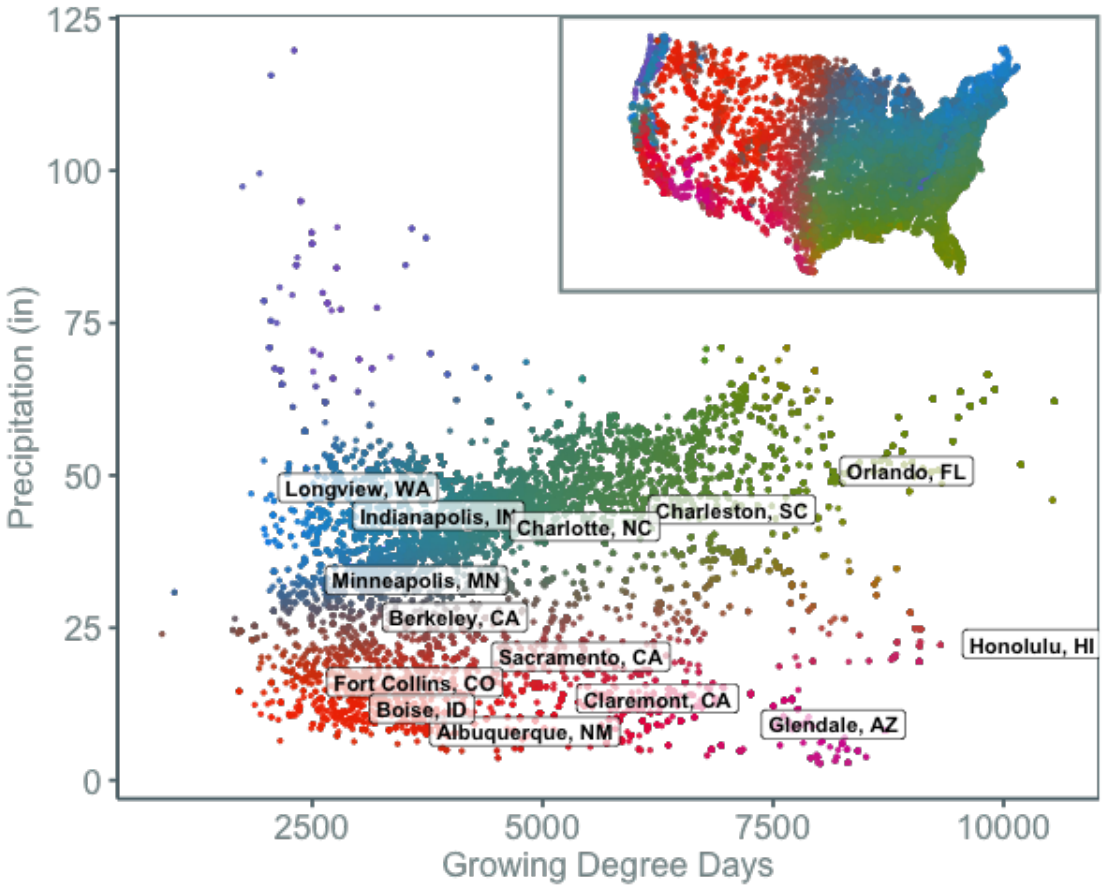


Figure 3: US census tract centroids with UTD reference cities overlaid in growing degree day (GDD) and precipitation climate space and matching color gradient in geographic space. The reference cities cover climate space well, and variation in precipitation and growing degree days is continuous.

Modelling

Model requirements

We sought a model of tree growth that would adequately represent the known biological dynamics of tree growth, namely that diameter growth rate starts slow, reaches a maximum at a young age, and then gradually declines to nearly zero. Diameter for trees much always increase however slightly because the growth of new wood is essential for proper function. This is different than tree height which often reaches its asympototte. Instead the asympote in our curve could be considered a pragmatic way to constrain diameter growth, or a practical asympote. While trees could theoretically continue to increase in diameter indefinitely, they don't in reality. The asympote represents this practical maximum diameter. An additional feature to the data is that age is time since transplanting. This means trees can have substantial diameter at age 0.

The type of curve that meets these criteria would be an asymmetric sigmoidal curve with an intercept. A modified weibull is such a curve that has worked well in forestry and is the one we use here Weiskittel et al. (2011). However, there are many other curves such as the Chapman-Richards that meet these criteria and could be used.

Another characteristic of tree growth curves is heterosckedasticity, namely that as the age of a tree increases, so does the variability around the mean. Often past modellers controlled this using log - log transformations Troxel et al. (2013), but we wanted to keep units in their original scale. We tested fitting models where the variance was a linear function or a smoothed spline function of age. However, this still could yield negative predictions at low ages. Instead we adopted the approach of modeling DBH from a gamma distribution, which yielded more realistic posterior predictions.

We fit models of generally increasing complexity starting a single weibull curve for all trees and then varying the curve parameters by city, by genus and species, and by climate. Using approximate leave-one-out cross validation we selected the model with the highest estimated

log posterior density Vehtari et al. (2017). The following section details this model, and other models tested are in supplementary materials.

Model

- Bayesian Mulitlevel Modeling a bit of background?
- likelihood

$$y_{igsc} \sim Gamma(\mu_{igsc}, \alpha_y)$$

$$\mu_{igsc} = \beta^0_{igsc} + \beta^1_{igsc}(1 - \exp(-\beta^2_{igsc} x^{\beta^3_{igsc}}_{igsc}))$$

where:

y_{igsc} is the diameter at breast height of tree i of genus, g, species, s, and city, c. y_{igsc} has a gamma distribution with mean, μ_{igsc}, and shape, α_y.

$i = 1, \ldots, n_{sc}$; where n_{sc} is the number of trees sampled for species, s, and city, c.

$g = 1, \ldots, G$; where G is the number of genera (G)

$s = 1, \ldots, S_g$; where S_g is the number of species in genus g.

x_{igsc} is the transplant age in years of tree $igsc$ (i.e. years since transplanting).

β^0_{igsc} is the intercept, or the diameter of a tree at time of transplanting.

β^1_{igsc} (plus β^0_{igsc}) is the asymptote of the sigmoidal weibull curve.

β^2_{igsc} and β^3_{igsc} affect the rate of growth. β^2_{igsc} provides flexibility to have slow or fast growth at young ages.

For each β^j_{igsc}, j = 0,1,2:

$$\beta^j_{igsc} = \beta^j_0 + \gamma^j_{gs} + \delta^j_c$$

for β^3_{igsc}:

$$\beta^3_{igsc} = \beta^3_0 + \tau_1 * \text{precip}_c + \tau_2 * \text{gdd}_c + \tau_3 * (\text{precip}_c * \text{gdd}_c) + \gamma^3_{gs} + \delta^3_c$$

where β^j_0 is the mean for β^j_{igsc}. γ^j_{gs} is genetic (genus and species) effect on β^j. δ^j_c is the city effect on β^j

The species effect is nested within the genus effect. Both are normally distributed, such that:

$$\gamma^j_{gs} \sim N(\gamma^j_g, \sigma^j_{genus:species})$$

$$\gamma^j_g \sim N(0, \sigma^j_{genus})$$

The effect of city is normally distributed:

$$\delta^j_c \sim N(0, \sigma^j_{city})$$

- Priors

  The priors were selected to make biologically unrealistic parameters highly improbable, but they have a small effect on the posterior estimates. The β's and α_y are gamma distributed, while the variance parameters are half-normal. More details on the selection of priors are available in supplementary materials and code. Many priors that might appear to be narrow (e.g Half-Normal(0,.1)) are actually fairly wide given the scale of the predictors and the reasonable range of some parameters.

genus: many; species: many; cities: many (not β_1); climate: b3linint; hetero: no; family: Gamma FULL DATA no Palm

- model R code

```
library(dplyr)
library(brms)



genus <- "many"
species <- "many"
cities <- "many_notB1"
climate <- "b3linint"
hetero <- "no"
family <- "Gamma"

data_form <- formula(DBH ~ b0 + 100 * b1 * (1 - exp(-(b2/100) * AGE^(b3))))
b0_form <- formula(b0 ~ (1 | City) + (1 | Genus/Species))
b1_form <- formula(b1 ~ (1 | Genus/Species))
b2_form <- formula(b2 ~ (1 | City) + (1 | Genus/Species))
b3_form <- formula(b3 ~ precip * gdd +  (1 | City) + (1 | Genus / Species))



form <- bf(data_form, b0_form, b1_form, b2_form, b3_form, nl = T)



nlprior <- c(prior(gamma(9, 3), nlpar = "b0",lb = 0),
             prior(gamma(34, 19.4), nlpar = "b1",lb = 0),
             prior(gamma(69.4, 55.5), nlpar = "b2", lb = 0),
             prior(gamma(16, 26), nlpar = "b3", lb = 0),
             prior(normal(0.01, 0.015), nlpar = "b3", coef = "gdd"),
             prior(normal(0.01, 0.01), nlpar = "b3", coef = "precip"),
```

```
              prior(normal(0.005, 0.005), nlpar = "b3", coef = "precip:gdd"),
              prior(gamma(20, 1), class = "shape"),
              prior(normal(0, .3), class = "sd", nlpar = "b0", group = "City"),
              prior(normal(0, .1), class = "sd", nlpar = "b2", group = "City"),
              prior(normal(0, .1), class = "sd", nlpar = "b3", group = "City"),
              prior(normal(0, .4), class = "sd", nlpar = "b0", group = "Genus"),
              prior(normal(0, .1), class = "sd", nlpar = "b0", group = "Genus:Specie
              prior(normal(.1, .4), class = "sd", nlpar = "b1", group = "Genus"),
              prior(normal(0, .1), class = "sd", nlpar = "b1", group = "Genus:Specie
              prior(normal(0, .1), class = "sd", nlpar = "b2", group = "Genus"),
              prior(normal(0, .05), class = "sd", nlpar = "b2", group = "Genus:Speci
              prior(normal(0, .1), class = "sd", nlpar = "b3", group = "Genus"),
              prior(normal(0, .05), class = "sd", nlpar = "b3", group = "Genus:Speci


d <- readRDS("../data/age_dbh_full_noPalms.rds")


mod <- brm(form,
           data = d,
           prior = nlprior,
           family = Gamma("identity"),
           chains = 12, cores = 12, init_r = .3, iter = 6000, control = list(adapt_


saveRDS(mod, paste0("../models/genus_",genus,"_species_",species,"_cities_", cities
```

- tangle
- send to krusty

```
rsync -avz genus_species_citiesNotB1_b3climate_fulldata_noPalm.R erker@krusty:~/all
```

- run on krusty

  run from krusty terminal

```
ssh krusty
cd allo/code
nohup R CMD BATCH genus_species_citiesNotB1_b3climate_fulldata_noPalm.R &
exit

cat genus_species_citiesNotB1_b3climate_fulldata_noPalm.Rout
```

- pull back from krusty

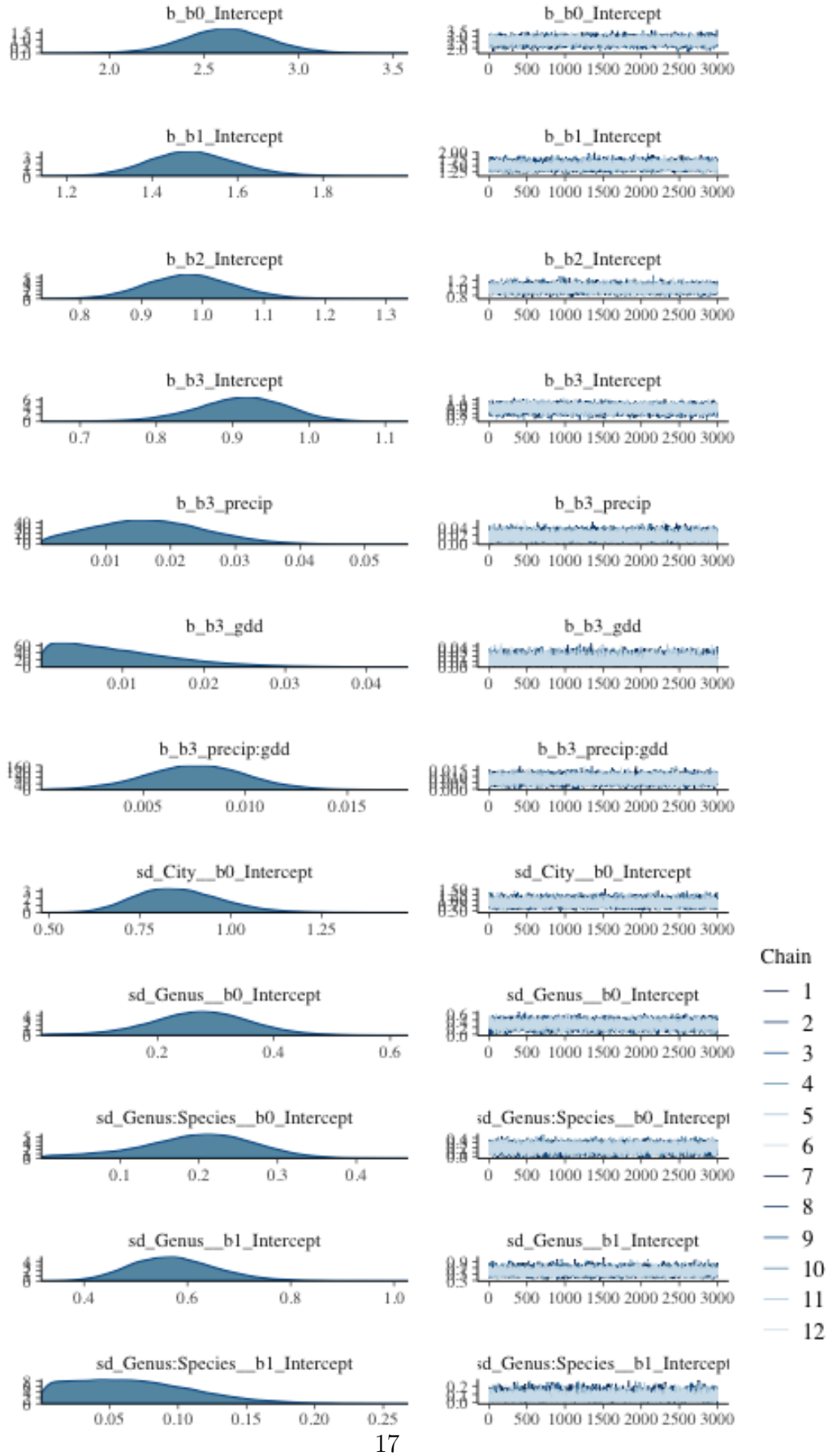
```
rsync -avz erker@krusty:~/allo/models/genus_many_species_many_cities_many_notB1_cli
```

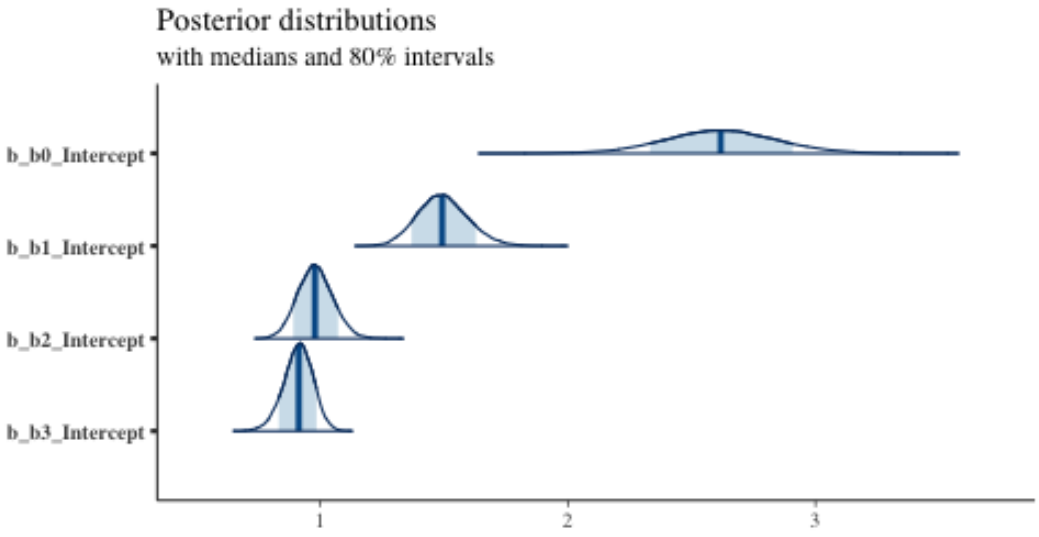
- assess model

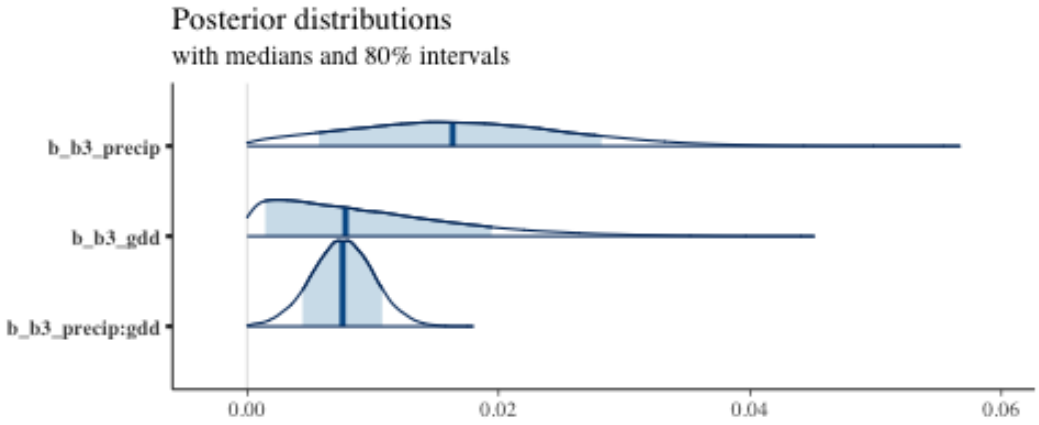
```
mod_genus_many_species_many_cities_many_notB1_climate_b3linint_hetero_no_family_Gam
mod <- mod_genus_many_species_many_cities_many_notB1_climate_b3linint_hetero_no_fam

summary(mod)

 Family: gamma
  Links: mu = identity; shape = identity
Formula: DBH ~ b0 + 100 * b1 * (1 - exp(-(b2/100) * AGE^(b3)))
         b0 ~ (1 | City) + (1 | Genus/Species)
         b1 ~ (1 | Genus/Species)
```

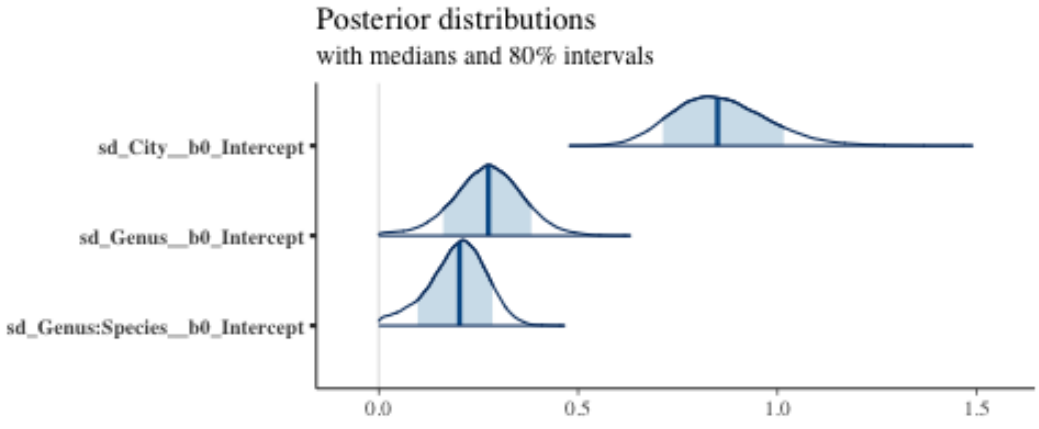
```
         b2 ~ (1 | City) + (1 | Genus/Species)
         b3 ~ precip * gdd + (1 | City) + (1 | Genus/Species)
   Data: d (Number of observations: 12156)
Samples: 12 chains, each with iter = 6000; warmup = 3000; thin = 1;
         total post-warmup samples = 36000

Group-Level Effects:
~City (Number of levels: 16)
                Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sd(b0_Intercept)     0.86      0.12     0.65     1.11      18222 1.00
sd(b2_Intercept)     0.26      0.04     0.20     0.35      16785 1.00
sd(b3_Intercept)     0.11      0.02     0.08     0.17      14946 1.00

~Genus (Number of levels: 76)
                Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sd(b0_Intercept)     0.27      0.09     0.09     0.44       6483 1.00
sd(b1_Intercept)     0.57      0.08     0.43     0.75       8746 1.00
sd(b2_Intercept)     0.22      0.05     0.12     0.32       2613 1.01
sd(b3_Intercept)     0.12      0.02     0.08     0.17       8200 1.00

~Genus:Species (Number of levels: 152)
                Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
sd(b0_Intercept)     0.20      0.07     0.04     0.33       5491 1.00
sd(b1_Intercept)     0.07      0.04     0.00     0.16       4548 1.00
sd(b2_Intercept)     0.20      0.03     0.16     0.26       6230 1.00
sd(b3_Intercept)     0.10      0.01     0.08     0.12       7096 1.00
```

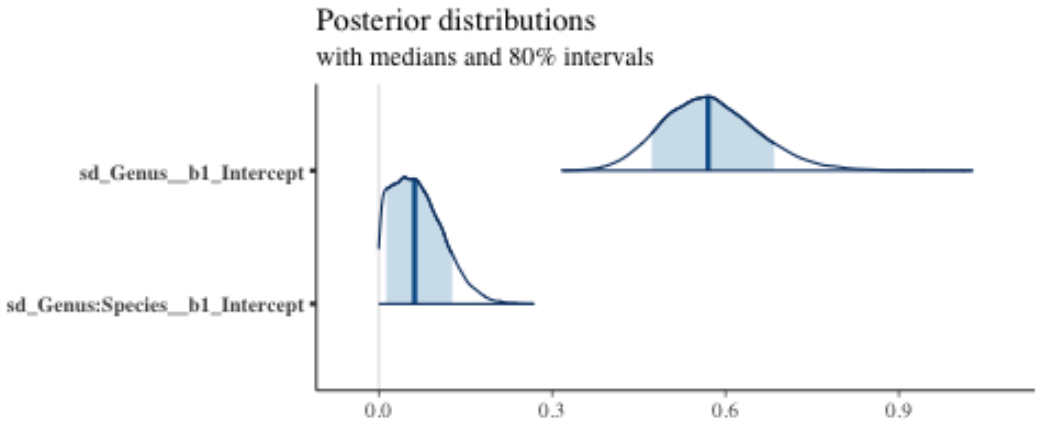
```
Population-Level Effects:
              Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
b0_Intercept      2.62      0.23     2.17     3.07       5655 1.00
b1_Intercept      1.50      0.10     1.31     1.71        947 1.02
b2_Intercept      0.98      0.07     0.85     1.12       9704 1.00
b3_Intercept      0.91      0.06     0.79     1.02       9187 1.00
b3_precip         0.02      0.01     0.00     0.03      34489 1.00
b3_gdd            0.01      0.01     0.00     0.03      25159 1.00
b3_precip:gdd     0.01      0.00     0.00     0.01      14563 1.00

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Eff.Sample Rhat
shape    13.36      0.18    13.01    13.70      47958 1.00

Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
is a crude measure of effective sample size, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
Warning message:
There were 54 divergent transitions after warmup. Increasing adapt_delta above 0.9 
See http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

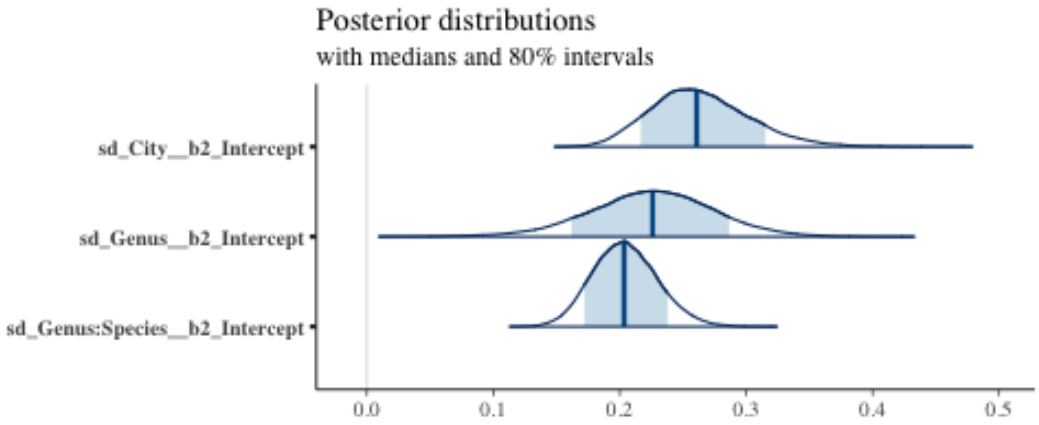

b_b0_Intercept
b_b1_Intercept
b_b2_Intercept
b_b3_Intercept
b_b3_precip
b_b3_gdd
b_b3_precip:gdd
sd_City__b0_Intercept
sd_Genus__b0_Intercept
sd_Genus:Species__b0_Intercept
sd_Genus__b1_Intercept
sd_Genus:Species__b1_Intercept
sd_City__b2_Intercept
Chain
1
2
3
4
5
6
7
8
9
10
11
12

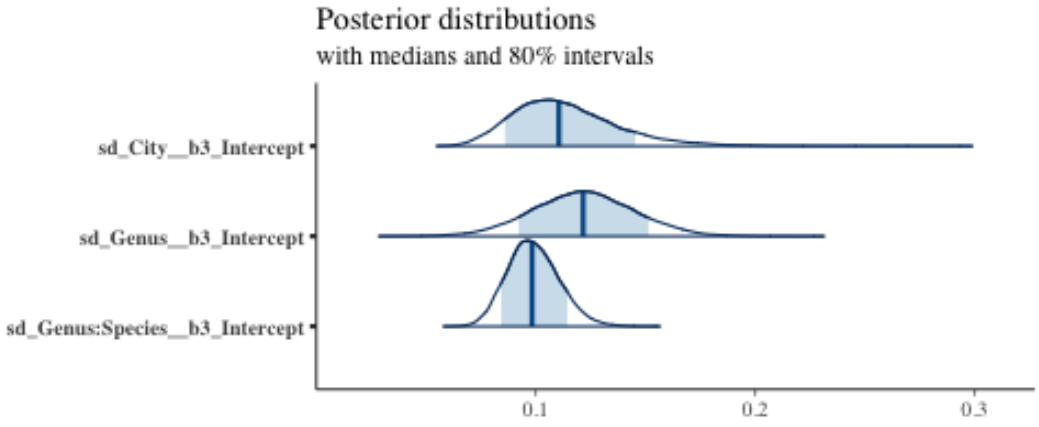

Posterior distributions
with medians and 80% intervals
b_b0_Intercept
b_b1_Intercept
b_b2_Intercept
b_b3_Intercept
1
2
3

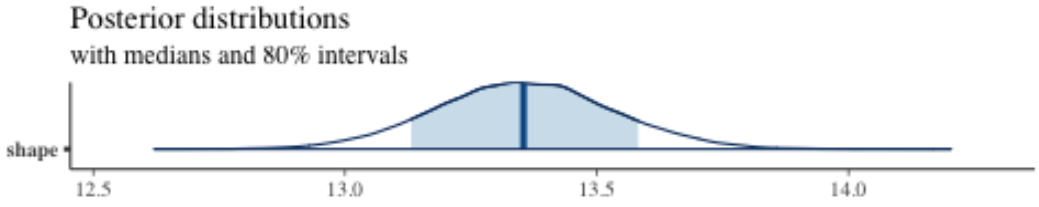

Posterior distributions
with medians and 80% intervals
b_b3_precip
b_b3_gdd
b_b3_precip:gdd
0.00
0.02
0.04
0.06

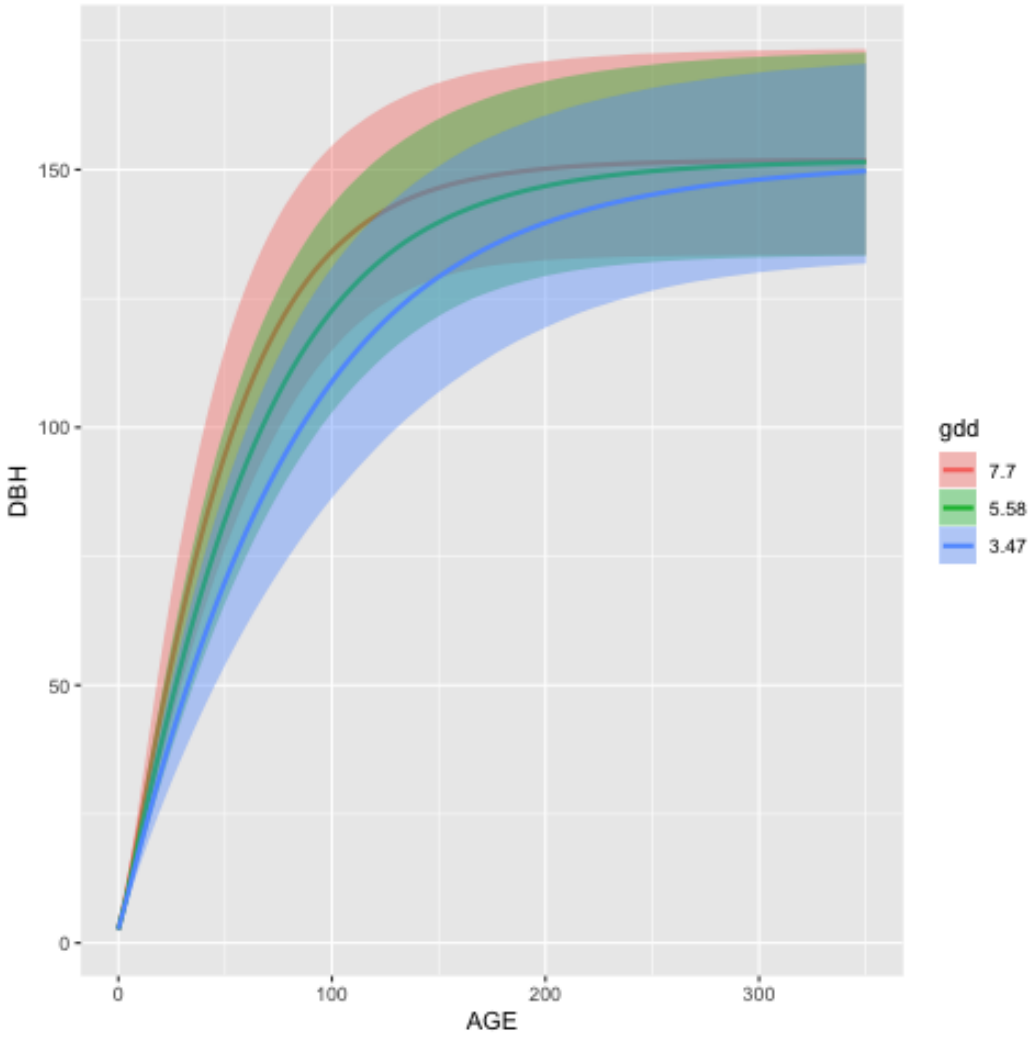

Posterior distributions
with medians and 80% intervals
sd_City__b0_Intercept
sd_Genus__b0_Intercept
sd_Genus:Species__b0_Intercept
0.0
0.5
1.0
1.5

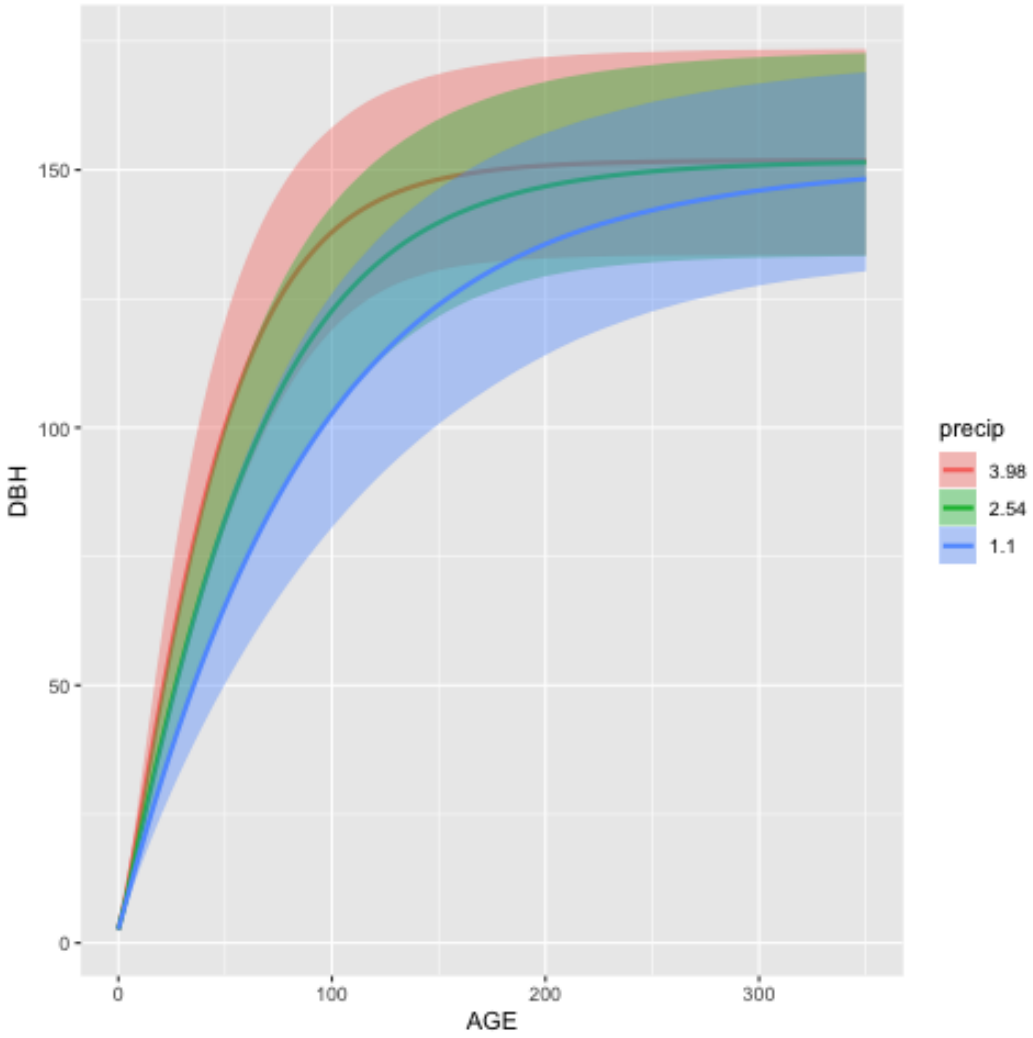

Posterior distributions
with medians and 80% intervals
sd_Genus__b1_Intercept
sd_Genus:Species__b1_Intercept
0.0
0.3
0.6
0.9

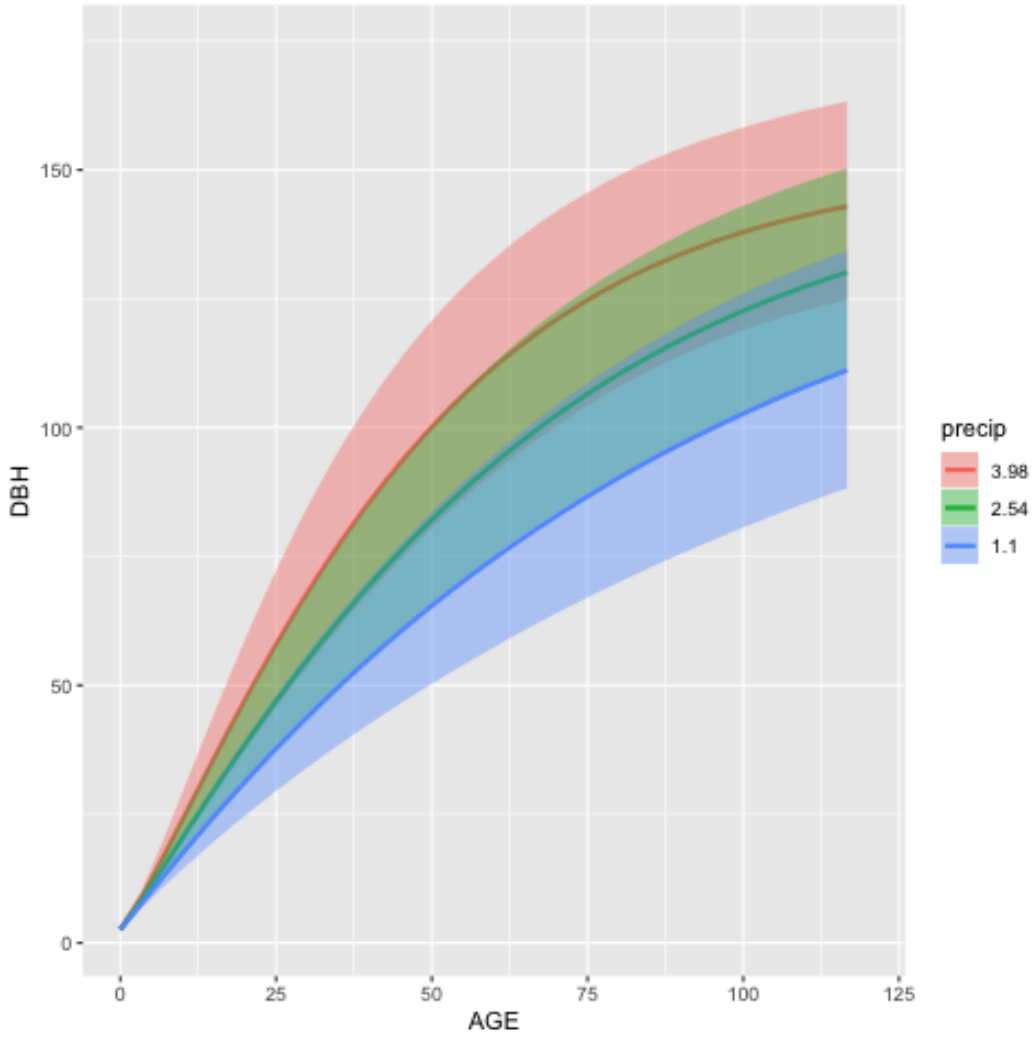

Posterior distributions
with medians and 80% intervals
sd_City__b2_Intercept
sd_Genus__b2_Intercept
sd_Genus:Species__b2_Intercept
0.0
0.1
0.2
0.3
0.4
0.5


Posterior distributions
with medians and 80% intervals
sd_City__b3_Intercept
sd_Genus__b3_Intercept
sd_Genus:Species__b3_Intercept
0.1
0.2
0.3


Posterior distributions
with medians and 80% intervals
shape
12.5
13.0
13.5
14.0

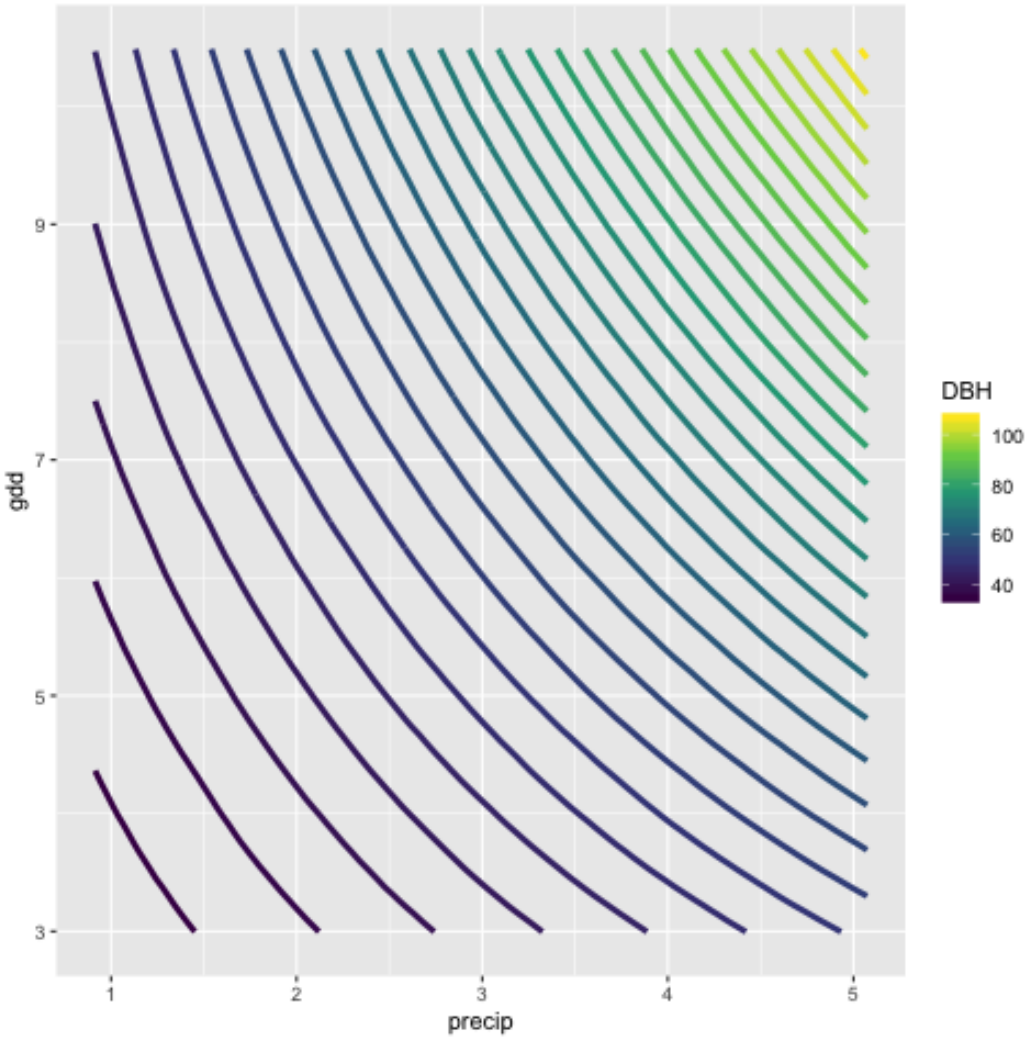

150
100
50
0
DBH
0
100
200
300
AGE
gdd
7.7
5.58
3.47


150
100
50
0
DBH
0
100
200
300
AGE
precip
3.98
2.54
1.1



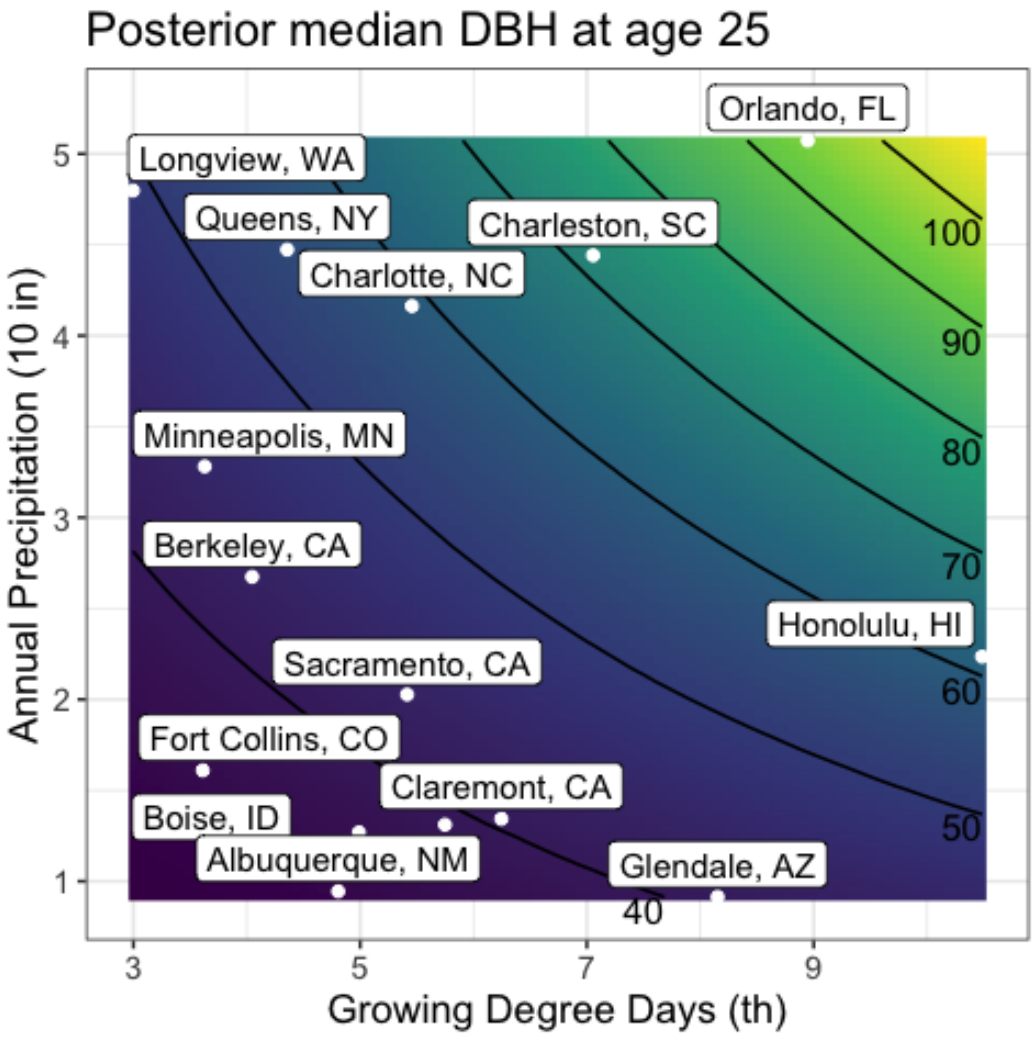

```
library(dplyr)
library(brms)
mod_genus_many_species_many_cities_many_notB1_climate_b3linint_hetero_no_family_Gam
mod <- mod_genus_many_species_many_cities_many_notB1_climate_b3linint_hetero_no_fam
precip.gdd <-   marginal_effects(mod, effects = "precip:gdd", surface = T, resoluti
saveRDS(precip.gdd, "../models/genus_many_species_many_cities_many_notB1_climate_b3


cond <- expand.grid(Species = unique(mod$data$Species),  City = unique(mod$data$Cit
cond <- left_join(cond, unique(select(mod$data, Species, Genus)))
```

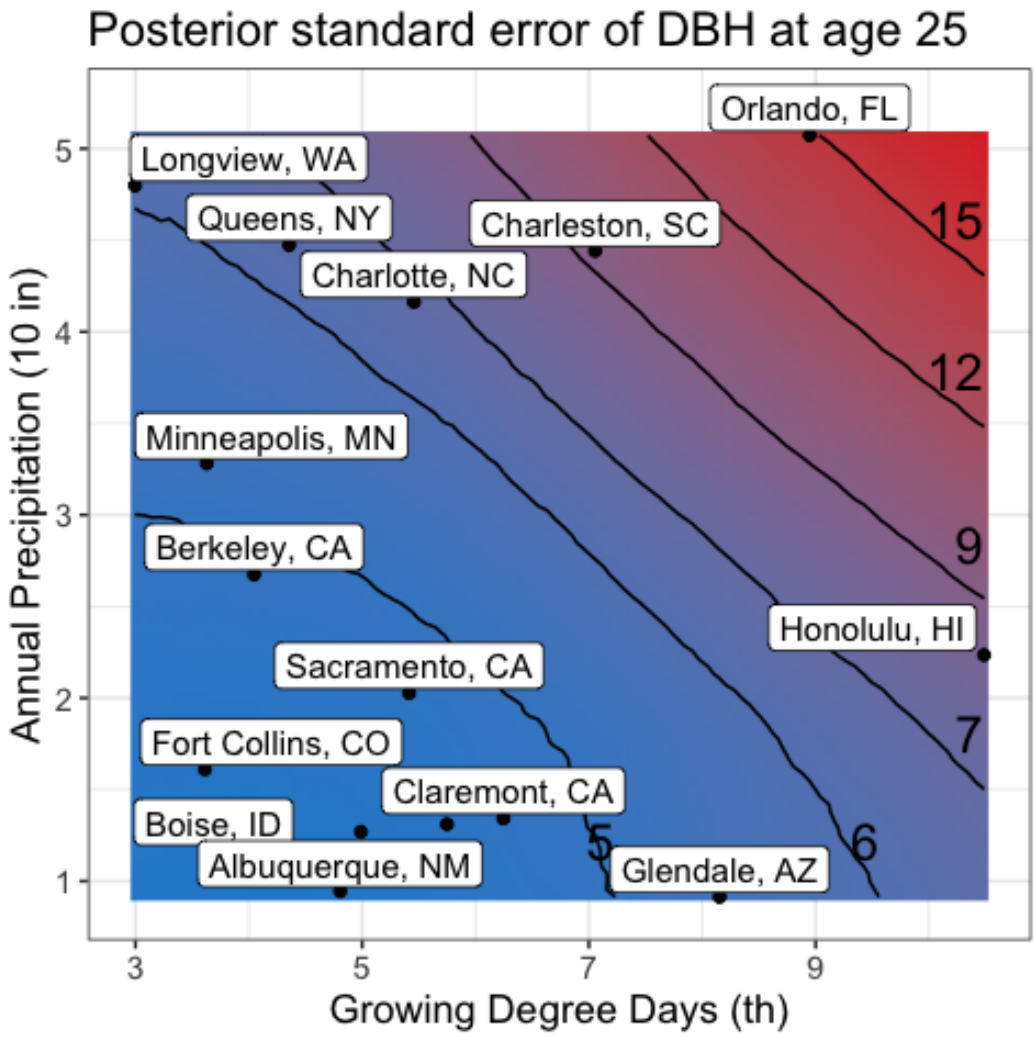
```
cond <- left_join(cond, unique(select(mod$data, City, precip, gdd)))

me <- marginal_effects(mod, effects = "AGE", conditions = cond, re_formula = NULL, r
saveRDS(me, "../models/genus_many_species_many_cities_notB1_many_climate_b3linint_h

rsync -avz calc_marginal_effects_fulldata.R erker@krusty:~/allo/code/

ssh krusty
cd allo/code
nohup R CMD BATCH calc_marginal_effects_fulldata.R &
cat calc_marginal_effects_fulldata.Rout
exit

#rsync -avz erker@krusty:~/allo/models/genus_many_species_many_cities_notB1_many_cl
rsync -avz erker@krusty:~/allo/models/genus_many_species_many_cities_many_notB1_cli

precip.gdd <- readRDS("../models/genus_many_species_many_cities_many_notB1_climate_
```

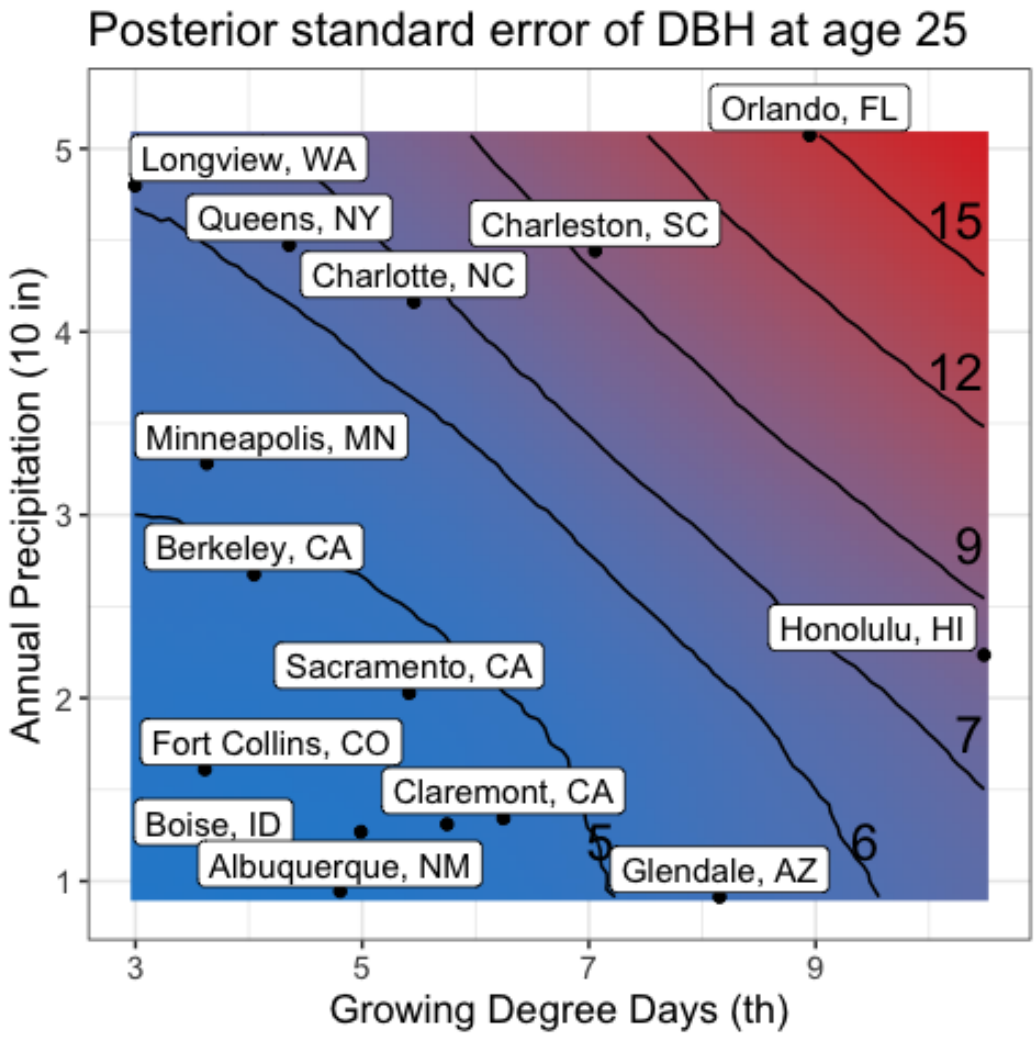


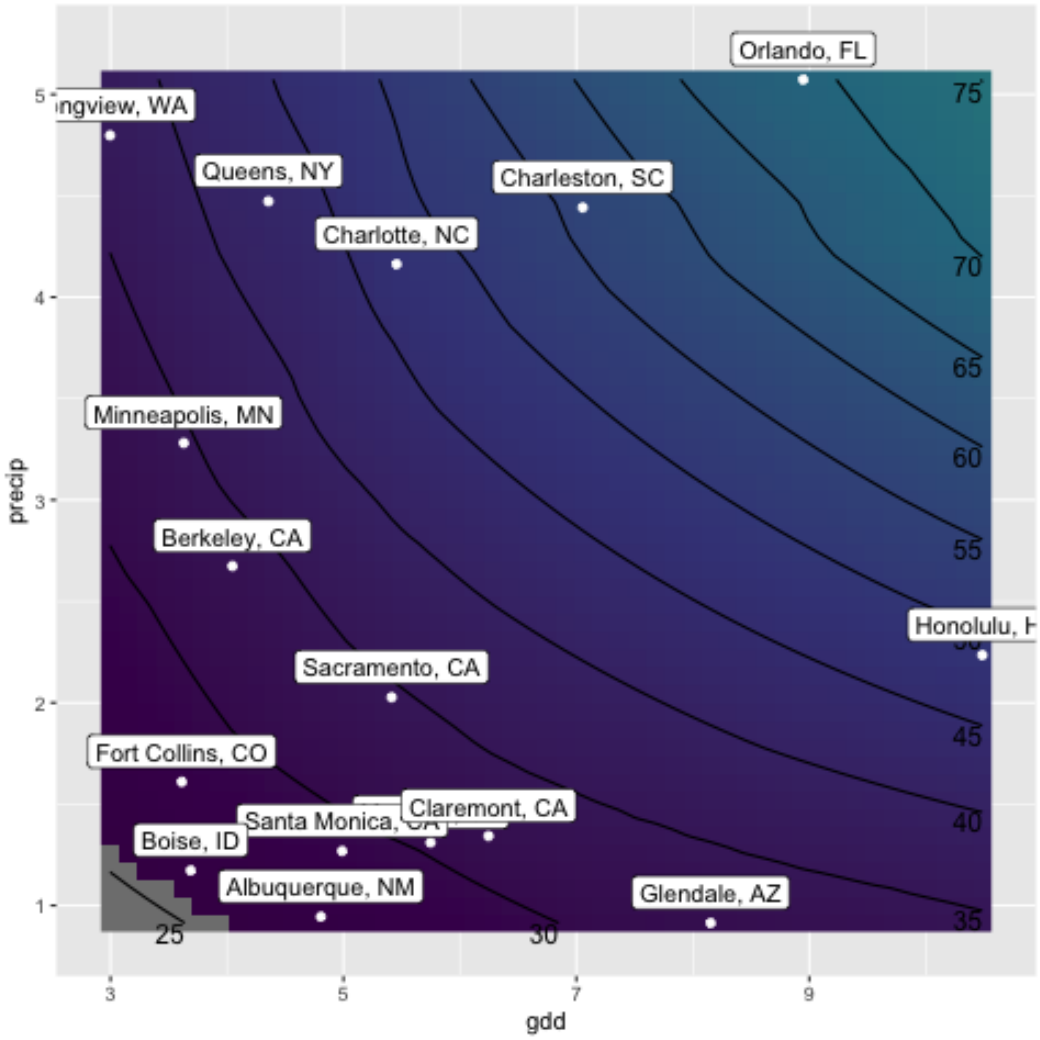
I should make these contours and label the contour lines with DBH.

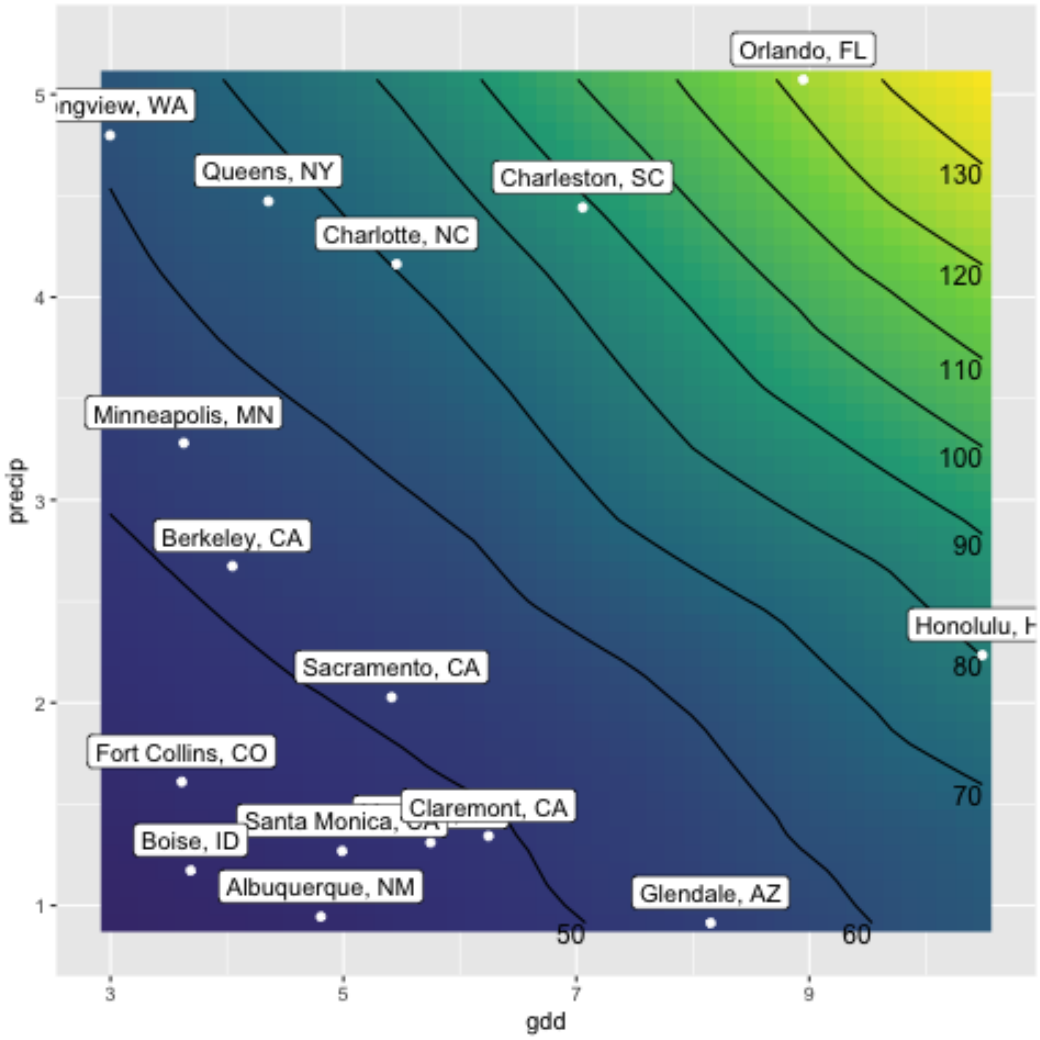
Use the direct label package

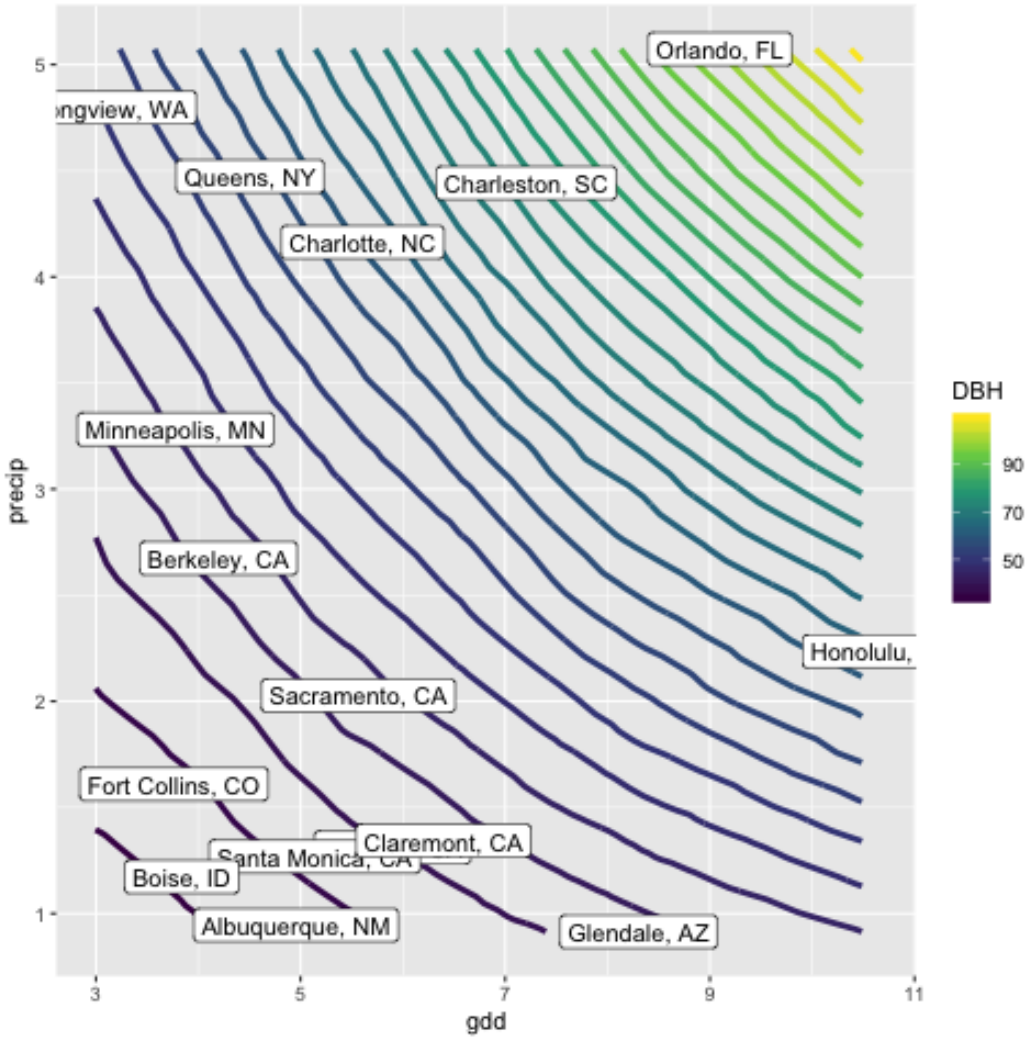

Posterior median DBH at age 25
Orlando, FL
Longview, WA
Queens, NY
Charleston, SC
Charlotte, NC
Minneapolis, MN
Berkeley, CA
Honolulu, HI
Sacramento, CA
Fort Collins, CO
Claremont, CA
Boise, ID
Albuquerque, NM
Glendale, AZ
100
90
80
70
60
50
40
Annual Precipitation (10 in)
Growing Degree Days (th)
1
2
3
4
5
3
5
7
9


Posterior standard error of DBH at age 25
Annual Precipitation (10 in)
Growing Degree Days (th)
Orlando, FL
Longview, WA
Queens, NY
Charleston, SC
Charlotte, NC
Minneapolis, MN
Berkeley, CA
Honolulu, HI
Sacramento, CA
Fort Collins, CO
Claremont, CA
Boise, ID
Albuquerque, NM
Glendale, AZ
15
12
9
7
6
5
1
2
3
4
5
3
5
7
9

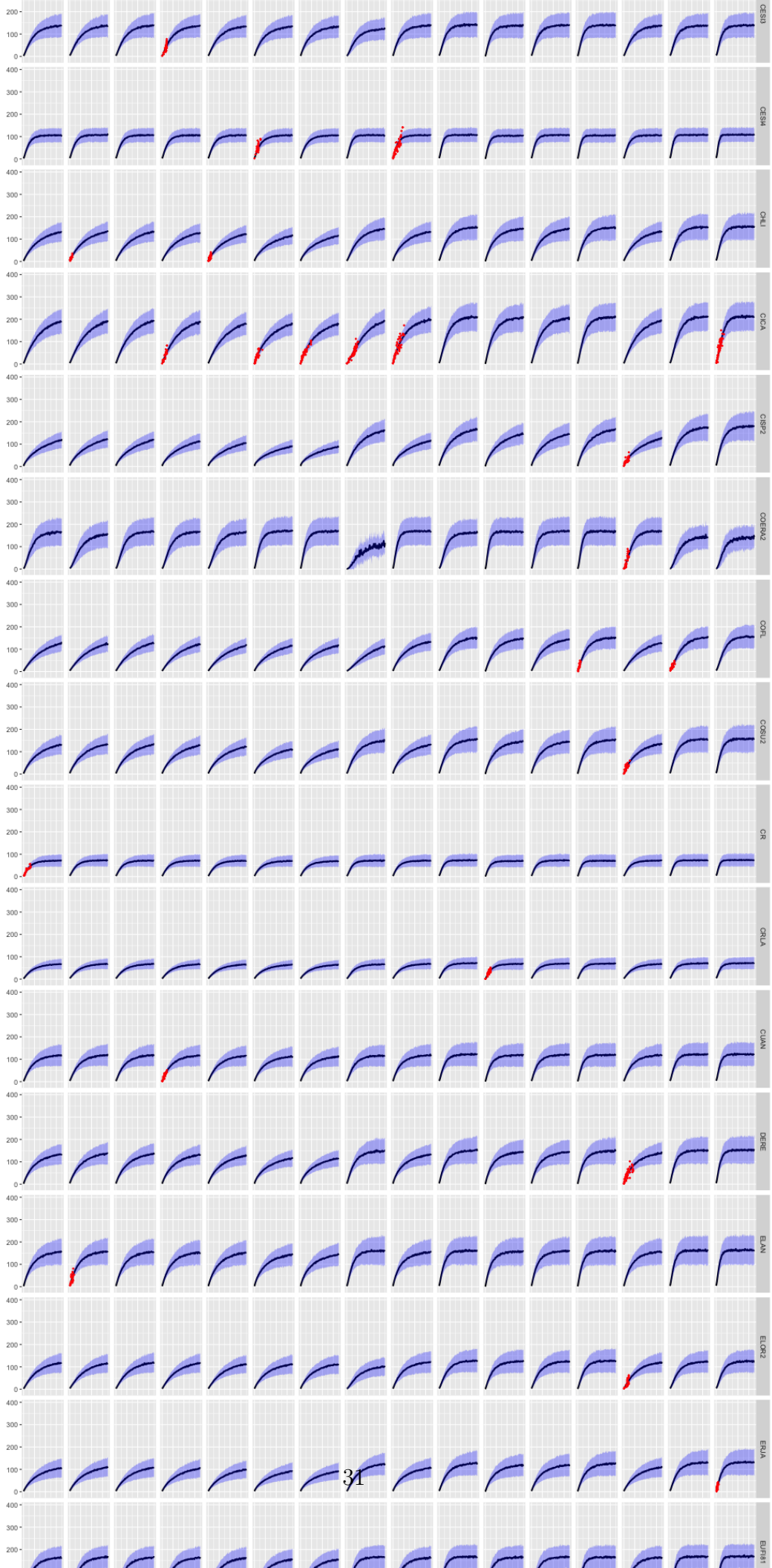

Posterior standard error of DBH at age 25
Orlando, FL
Longview, WA
Queens, NY
Charleston, SC
Charlotte, NC
Minneapolis, MN
Berkeley, CA
Honolulu, HI
Sacramento, CA
Fort Collins, CO
Claremont, CA
Boise, ID
Albuquerque, NM
Glendale, AZ
15
12
9
7
6
5
Annual Precipitation (10 in)
Growing Degree Days (th)
1
2
3
4
5
3
5
7
9


Orlando, FL
ngview, WA
Queens, NY
Charleston, SC
Charlotte, NC
Minneapolis, MN
Berkeley, CA
Honolulu, H
Sacramento, CA
Fort Collins, CO
Claremont, CA
Santa Monica, CA
Boise, ID
Albuquerque, NM
Glendale, AZ
75
70
65
60
55
45
40
35
25
30
precip
gdd
5
4
3
2
1
3
5
7
9


Orlando, FL
ngview, WA
Queens, NY
Charleston, SC
Charlotte, NC
130
120
110
100
90
80
70
60
50
Minneapolis, MN
Berkeley, CA
Honolulu, H
Sacramento, CA
Fort Collins, CO
Claremont, CA
Santa Monica, CA
Boise, ID
Albuquerque, NM
Glendale, AZ
precip
gdd
1
2
3
4
5
3
5
7
9




```
me <-readRDS("~/git/allo/models/genus_many_species_many_cities_notB1_many_climate_b
```


CES3
CES4
CHLI
CICA
CISP2
COERA2
COFL
COSU2
CR
CRLA
CUAN
DERE
ELAN
ELOR2
ERUA
EUFR61
0
100
200
300
400

- Plot Urban Tree Allometric equations on top of data

```
predict.allo <- function(x, EqName, a, b, c, d, e) {
    if(EqName == "loglogw1") {
        y = exp(a + b*log(log(x + 1) + c/2))
    }
    else if(EqName == "loglogw2") {
        y = exp(a + b*log(log(x + 1))+(sqrt(x) * (c/2)))
    }
    else if (EqName == "loglogw3") {
        y = exp(a + b*log(log(x + 1)) + x * c/2)
    }
    else if (EqName == "loglogw4") {
        y = exp(a + b*log(log(x + 1)) + x^2 * c/2)
    }
    else if (EqName == "expow1") {
        y = exp(a+ b * (x) + (c/2))
    }
    else if (EqName == "lin") {
        y = a + b * x
    }
    else if (EqName == "quad") {
        y = a + b * x + c* x^2
    }
    else if (EqName == "cub") {
        y = a+b * x+c *x^2 + d * x^3
    }
    else if (EqName == "quart") {
```

```
        y = a+b * x+c *x^2 + d * x^3 + e * x^4
    }
    return(y)
}



d <- readRDS("../data/age_dbh_full_noPalms.rds")
  eqn <- read.csv("../data/RDS-2016-0005/Data/TS6_Growth_coefficients_20180326.csv"
      mutate(a = as.numeric(a)) %>%
      rename(Species = SpCode)

  city_region <- read.csv("../data/city_climate.csv") %>%
    select(Region, City)

city_clim <- read.csv("../data/cities_gdd_precip.csv") %>%
  mutate(gdd = gdd / 1000, precip = precip /1000)

eqn <- left_join(eqn, city_region)
eqn <- left_join(eqn, city_clim)

  eqn <- eqn %>%
      filter(Predicts.component %in% c("dbh"), Independent.variable == "age")

  age_min_max = d %>%
      group_by(Region, Species) %>%
      summarize(minAGE = min(AGE, na.rm = T),
                maxAGE = max(AGE, na.rm = T))
```

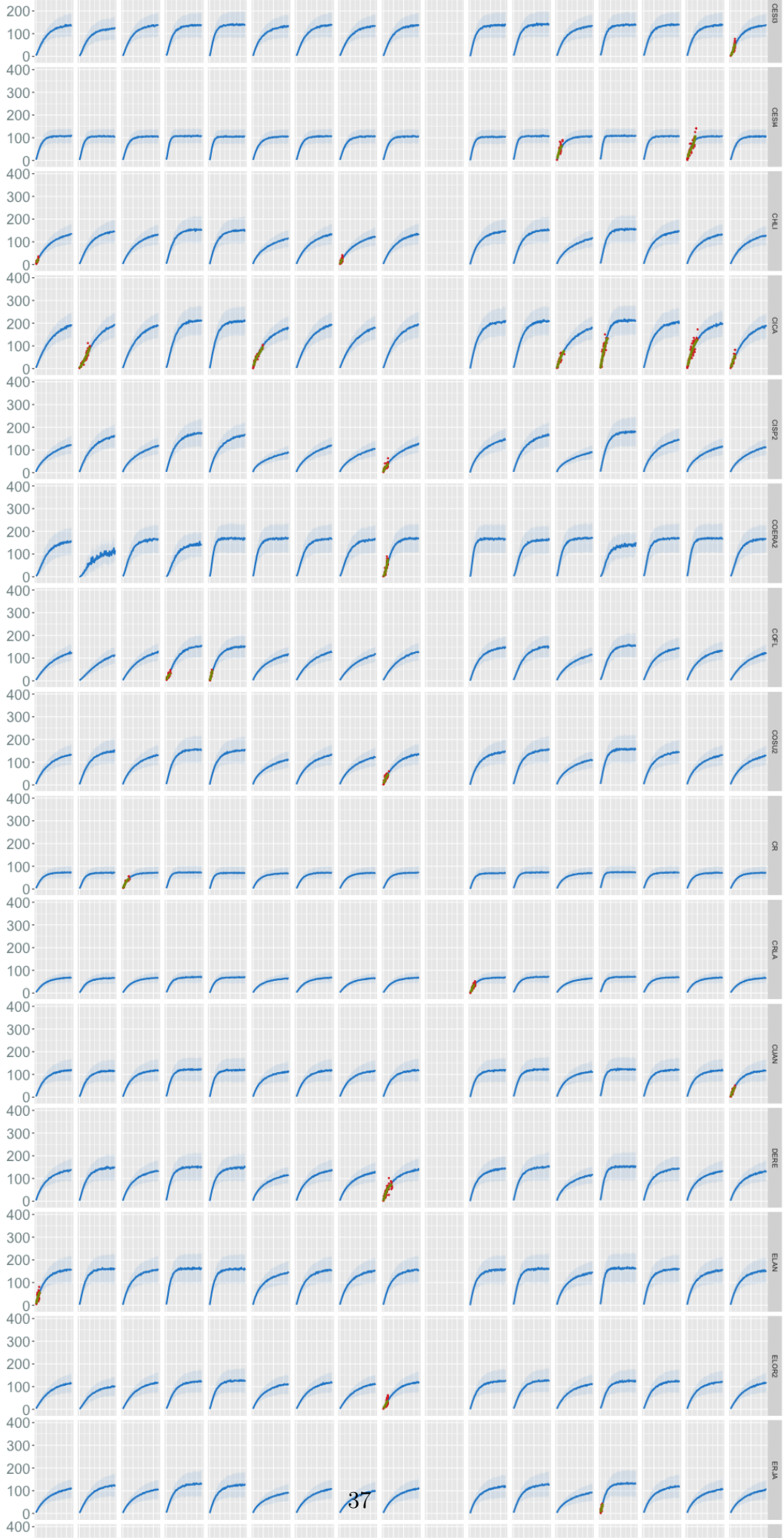
```
eqn <- left_join(eqn, age_min_max)

DBH_min_max = d %>%
    group_by(Region, Species) %>%
    summarize(minDBH = min(DBH, na.rm = T),
              maxDBH = max(DBH, na.rm = T))

eqn <- left_join(eqn, DBH_min_max)



                                        # fill in the NAs due to equations existi
eqn$minAGE[is.na(eqn$minAGE)] <- 0
eqn$maxAGE[is.na(eqn$maxAGE)] <- 100

newdata <- lapply(1:nrow(eqn), function(i) {
    x <- seq(eqn$minAGE[i], eqn$maxAGE[i],  (eqn$maxAGE[i] - eqn$minAGE[i]) / 20)
    cbind(x, eqn[i,])
})

newdata <- bind_rows(newdata)

predictions <- newdata %>% rowwise %>% mutate(predicted_dbh = predict.allo(x = x,



                                        #filter out predictions that are outside
```

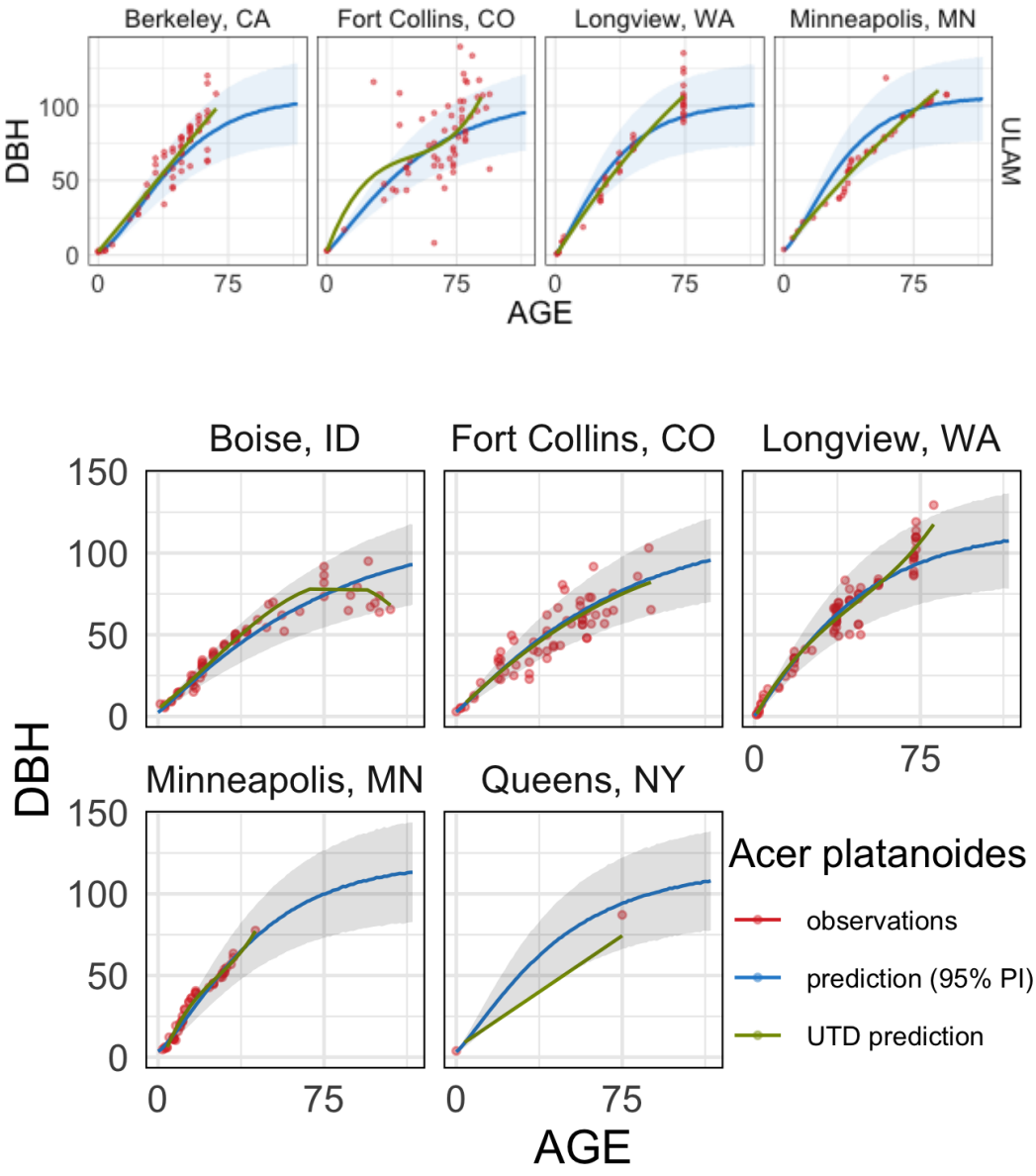
```
  predictions_apprange <- predictions %>%
      filter(predicted_dbh > Apps.min & predicted_dbh < Apps.max)

  predictions_datarange <- predictions %>%
      filter(predicted_dbh > minDBH & predicted_dbh < maxDBH)

  predictions_apprange <- predictions_apprange %>% mutate(AGE = x, DBH = predicted_
```

```
Joining, by = "Region"
Warning message:
Column ‘Region‘ joining character vector and factor, coercing into character vector
Joining, by = "City"
Joining, by = c("Region", "Species")
Joining, by = c("Region", "Species")
There were 50 or more warnings (use warnings() to see the first 50)
```

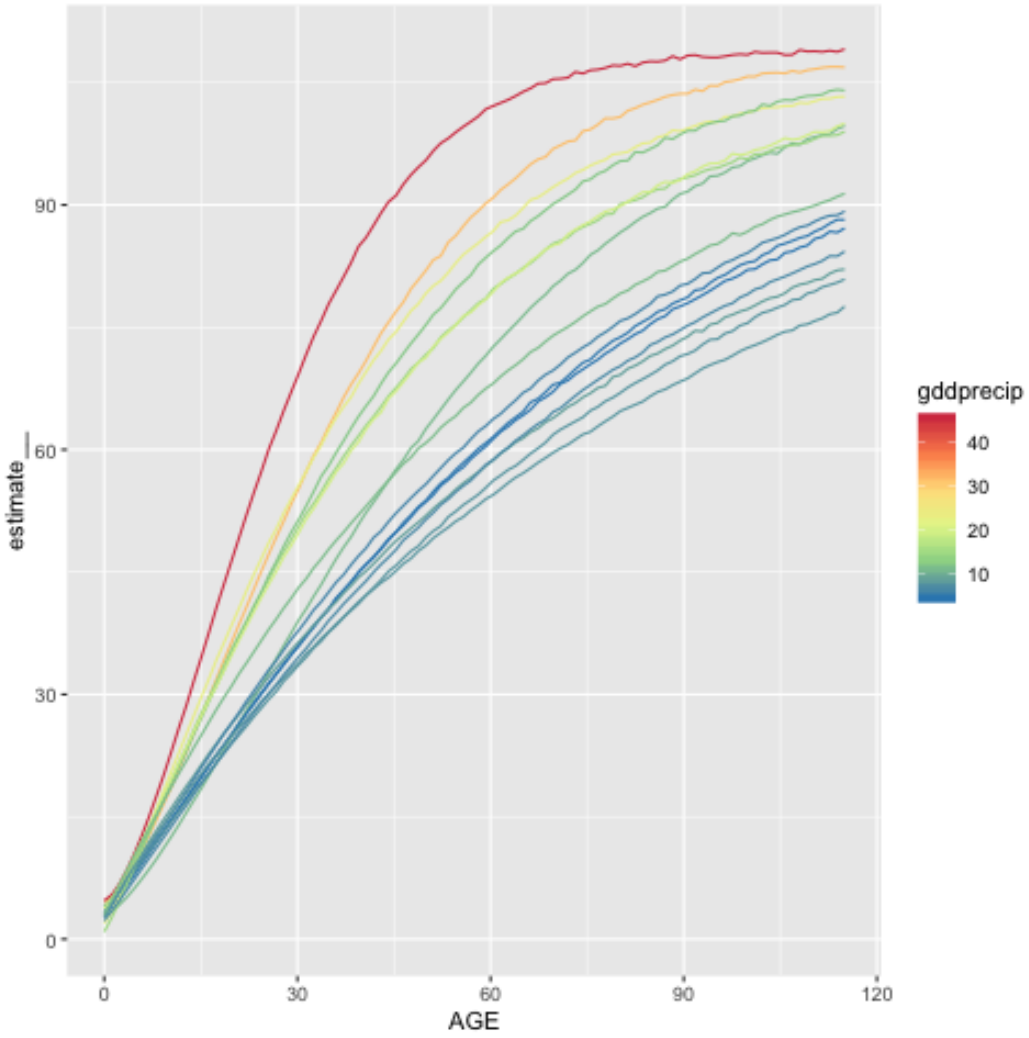
```
    me2 <- me$AGE  %>% mutate(gddprecip = gdd * precip,
                              City = factor(City, levels=unique(City[order(gddprecip)

    dd <- mod$data %>%
        mutate(gddprecip = gdd  * precip,
               City = factor(City, levels=unique(City[order(gddprecip)]), ordered=T

predictions_apprange <- predictions_apprange %>%
        mutate(gddprecip = gdd  * precip,
               City = factor(City, levels=unique(City[order(gddprecip)]), ordered=T

  dsp <- unique(dd$Species)
```

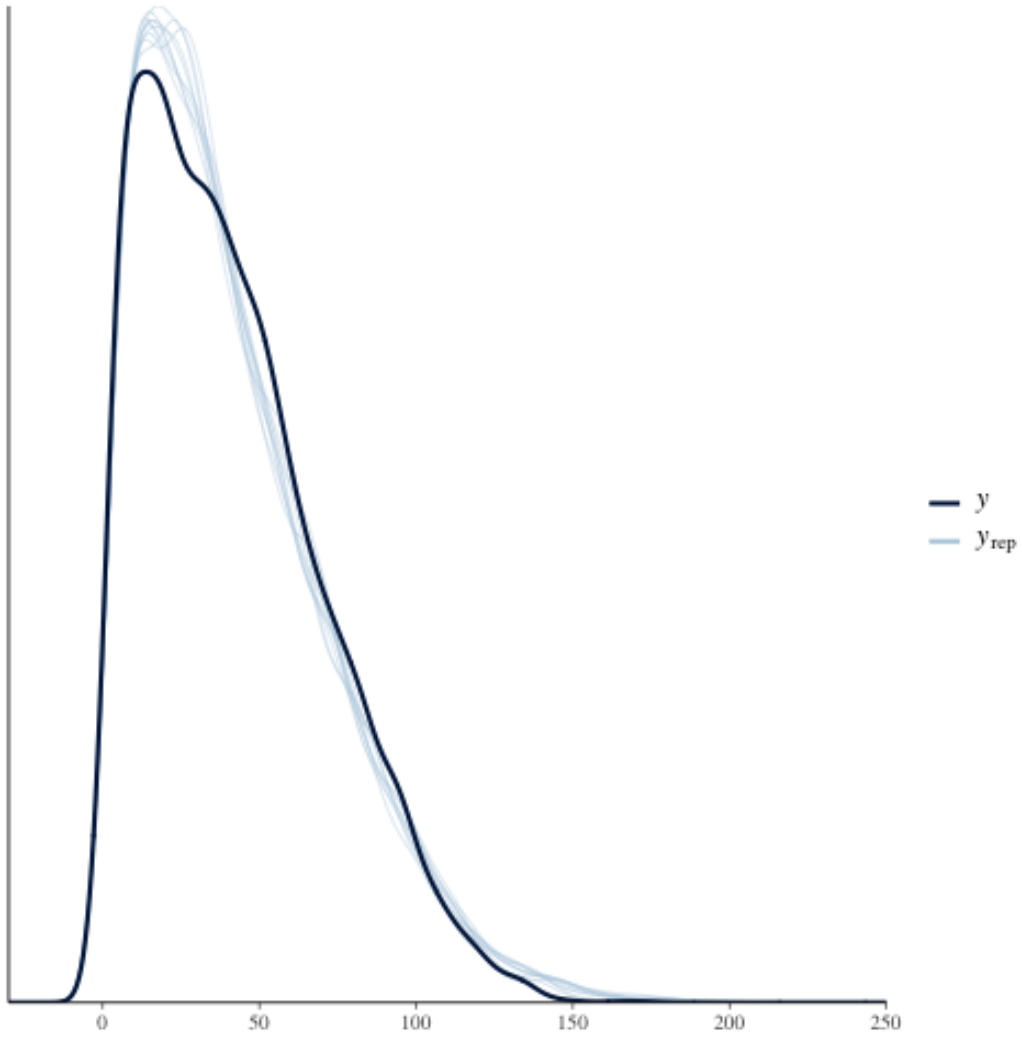
```
   me2 <- me$AGE  %>% mutate(gddprecip = gdd * precip,
                             City = factor(City, levels=unique(City[order(gddprecip)])

me2 <- left_join(me2, unique(select(d, Region, City)))

  dd <- mod$data %>%
      mutate(gddprecip = gdd  * precip,
             City = factor(City, levels=unique(City[order(gddprecip)]), ordered=TRU

  dd <- left_join(dd, unique(select(d, Region, City)))

predictions_apprange_sub <- predictions_apprange %>% filter(Species %in% unique(me2

There were 50 or more warnings (use warnings() to see the first 50)
Joining, by = "City"
Warning message:
Column ‘City‘ joining factor and character vector, coercing into character vector
Joining, by = "City"
Warning message:
Column ‘City‘ joining factor and character vector, coercing into character vector
```


CESI3
CESI4
CHLI
CICA
CISP2
COERA2
COFL
COSU2
CR
CRLA
CUAN
DERE
ELAN
ELOR2
ERJA
ELFR1
0
100
200
300
400



What city is the lower midwest? Indianapolis. There is no data for that.


estimate
90
60
30
0
0
30
60
90
120
AGE
gddprecip
40
30
20
10

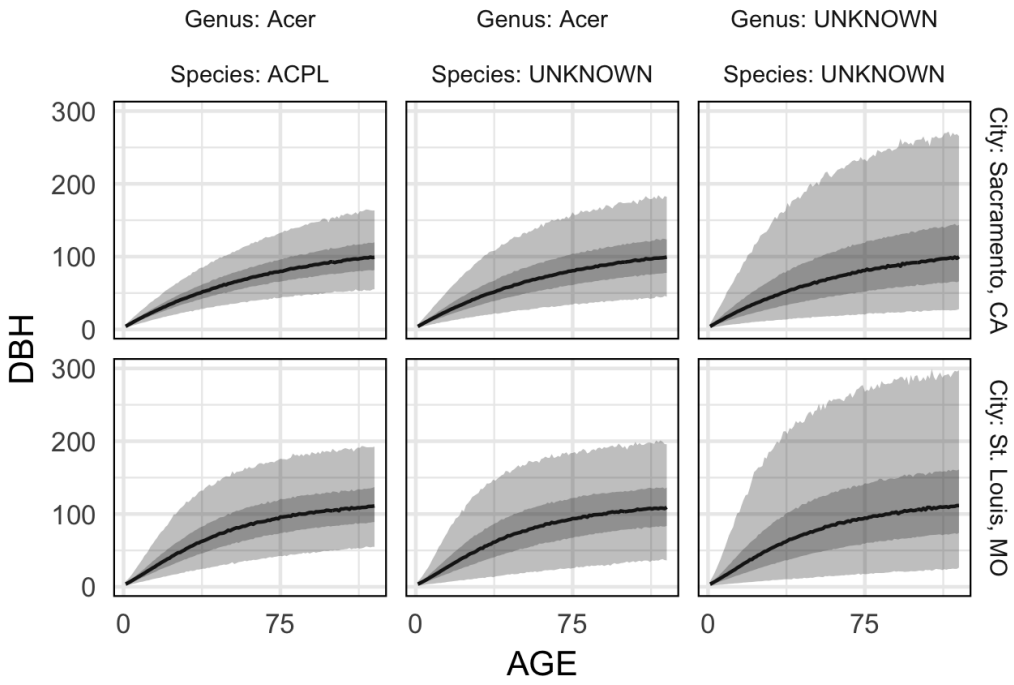


- test some predictions

```
wg  <-  read.csv("../data/precip_qt_ll.csv")
wp <-     read.csv("../data/gdd_qt_ll.csv")

w <- left_join(wg, wp, by = c("station", "lat", "long"))

stl <- c(lat = 38.6270, long = -90.1994)

w2 <- filter(w, lat > stl["lat"] - .1,
```

```
                lat < stl["lat"] + .1,
                long > stl["long"] - .1,
                long < stl["long"] + .1)


#  ggplot(w, aes(x = long, y = lat)) +
 #     geom_point() +
  #    geom_point(data = w2, color = "pink")

  Species <- c("ACPL","UNKNOWN", "UNKNOWN")
  Genus <- c("Acer", "Acer", "UNKNOWN")
  gs <- data.frame(Genus, Species)
  City <- c("Sacramento, CA", "St. Louis, MO")
  gdd <- c(5.41, 4822/1000)
  precip <- c(2.03, 4273/1000)

  nd.clim <- data.frame(City, gdd, precip)

  nd <- expand.grid(City = City, AGE = 1:120)

  nd <- left_join(nd, nd.clim)

  expand.grid.df <- function(...) Reduce(function(...) merge(..., by=NULL), list(..

  nd <- expand.grid.df(nd, gs)

  pred.nd <- predict(mod, nd, allow_new_levels = T, robust = T, probs = c(0.025, 0.
```

```
Warning message:
Column ‘station‘ joining factors with different levels, coercing to character vecto
Joining, by = "City"
There were 50 or more warnings (use warnings() to see the first 50)

pd <- cbind(nd, pred.nd)
```



Results

Model Comparisons

Short descriptions of the models tested and the brms sytax are in table **??**

Note, I removed the scaling of the parameters (multiplying and dividing by 100) in the formula for clarity. In the code they are scaled so that the parameters are on the same order of magnitude and HMC sampling is improved.

Model	Description	brms formula syntax
1	No varying parameters	DBH $\sim$ b0 + b1 * (1 - exp(-b2 * AGE^{b3}))
		b0 $\sim$ 1
		b1 $\sim$ 1
		b2 $\sim$ 1
		b3 $\sim$ 1
2	Parameters vary by city	DBH $\sim$ b0 + b1 * (1 - exp(-b2 * AGE^{b3}))
		b0 $\sim$ (1 | City)
		b1 $\sim$ (1 | City)
		b2 $\sim$ (1 | City)
		b3 $\sim$ (1 | City)
3	Parameters vary by genus and species	DBH $\sim$ b0 + b1 * (1 - exp(-b2 * AGE^{b3}))
	Species is nested in genus	b0 $\sim$ (1 | Genus / Species)
		b1 $\sim$ (1 | Genus / Species)
		b2 $\sim$ (1 | Genus / Species)
		b3 $\sim$ (1 | Genus / Species)
4	Asympotote (β_1) varies by climate	DBH $\sim$ b0 + b1 * (1 - exp(-b2 * AGE^{b3}))
		b0 $\sim$ 1
		b1 $\sim$ gdd * precip
		b2 $\sim$ 1
		b3 $\sim$ 1
5	Growth rate (β_3) varies by climate	DBH $\sim$ b0 + b1 * (1 - exp(-b2 * AGE^{b3}))
		b0 $\sim$ 1
		b1 $\sim$ 1
		b2 $\sim$ 1
		b3 $\sim$ gdd * precip
6	Parameters vary by city, genus, and species.	DBH $\sim$ b0 + b1 * (1 - exp(-b2 * AGE^{b3}))
	Growth rate varies by climate.	b0 $\sim$ (1 | City) + (1 | Genus/Species)
		b1 $\sim$ (1 | City) + (1 | Genus/Species)
		b2 $\sim$ (1 | City) + (1 | Genus/Species)
		b3 $\sim$ precip * gdd + (1 | City) + (1 | Genus/Species)
7	Parameters vary by city, genus, and species	DBH $\sim$ b0 + b1 * (1 - exp(-b2 * AGE^{b3}))
	(but asympote does not vary by city).	b0 $\sim$ (1 | City) + (1 | Genus/Species)
	Growth rate varies by climate.	b1 $\sim$ (1 | Genus/Species)
		b2 $\sim$ (1 | City) + (1 | Genus/Species)
		b3 $\sim$ precip * gdd + (1 | City) + (1 | Genus/Species)

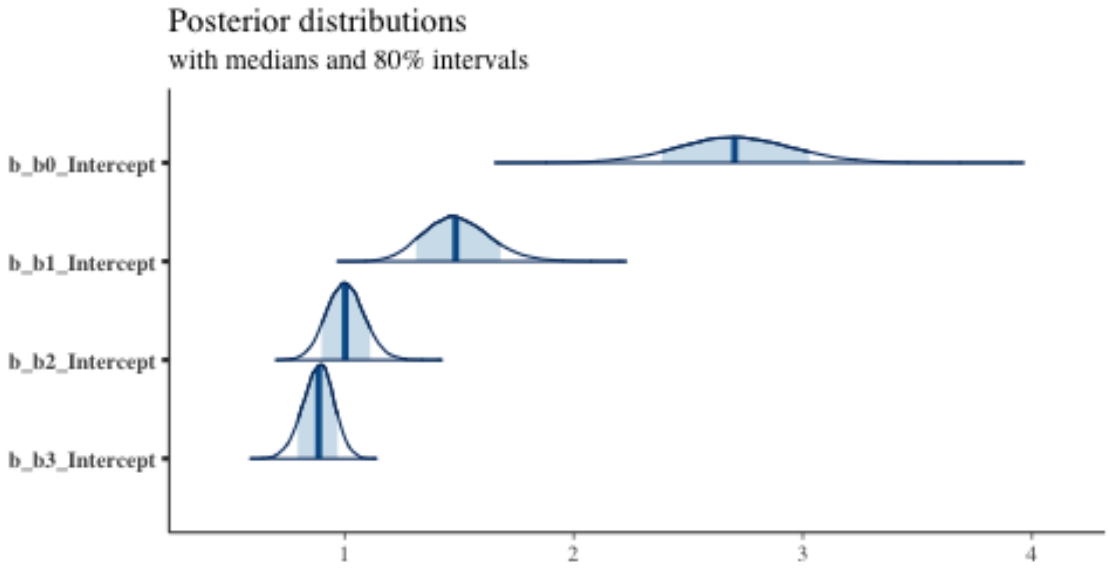
Table 1: $\widehat{elpd}_{\text{loo}}$ is the estimated expected log pointwise predictive density. elpd diff is the difference from the $\widehat{elpd}_{\text{loo}}$ of the top model. se elpd loo is standard error of Vehtari et al. (2017) for descriptions

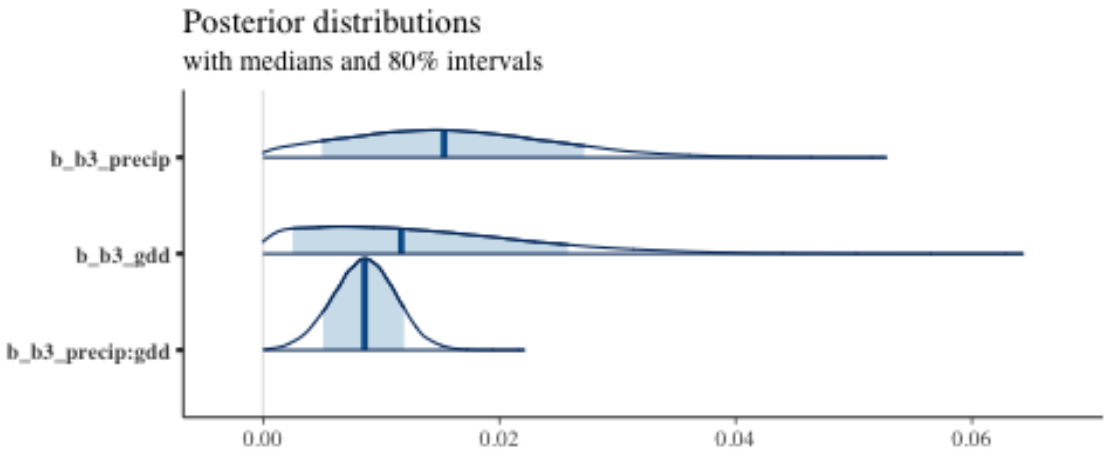
	Model	$\widehat{elpd}_{\text{loo}}$	difference
Best	6	-18845.41	0.00
	7	-18976.38	-130.97
	3	-18989.24	-143.83
	2	-19764.48	-919.06
	5	-20180.41	-1334.99
	4	-20195.21	-1349.80
Worst	1	-20513.12	-1667.70

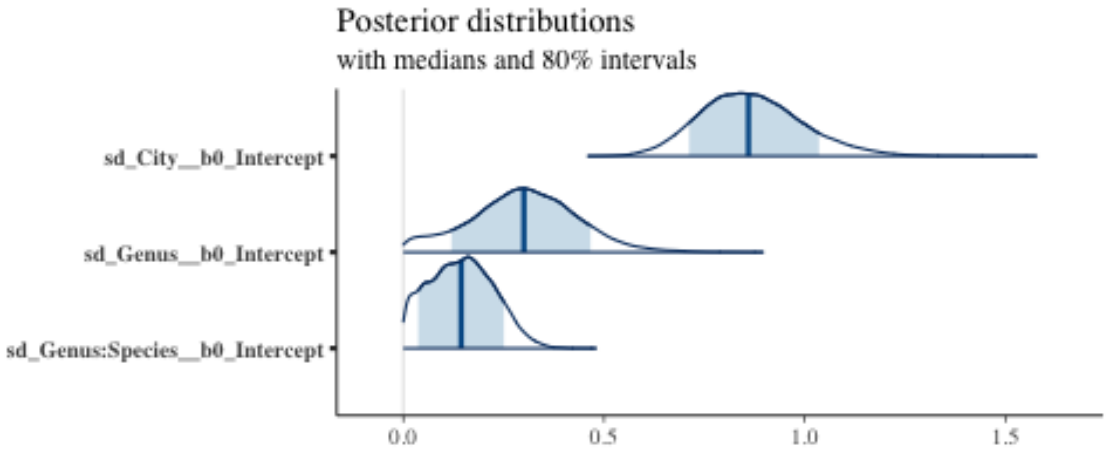
The standard error for the elpd difference of 131 between model 6 and model 7 is 21.4. Therefore, there is strong evidence that model 6 has higher out of sample predictive accuracy than model 7.

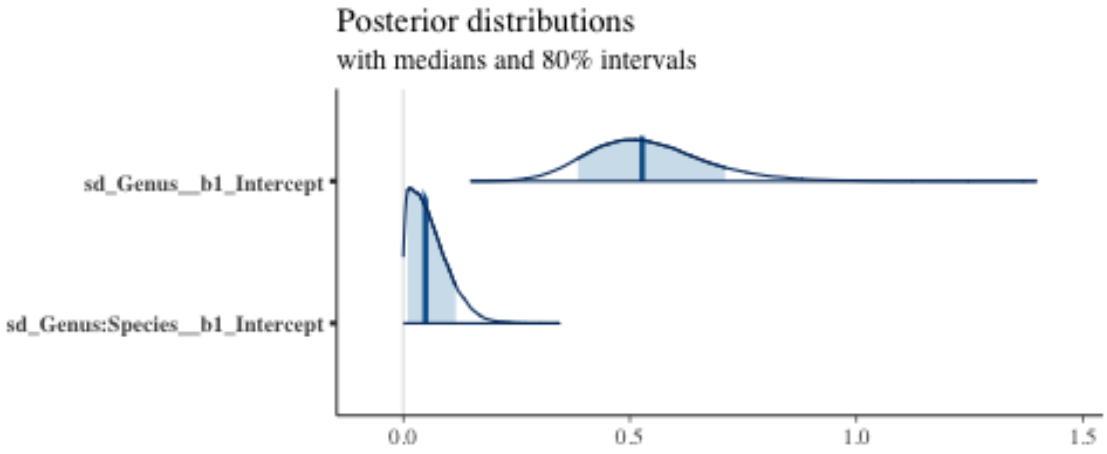
parameter estimates

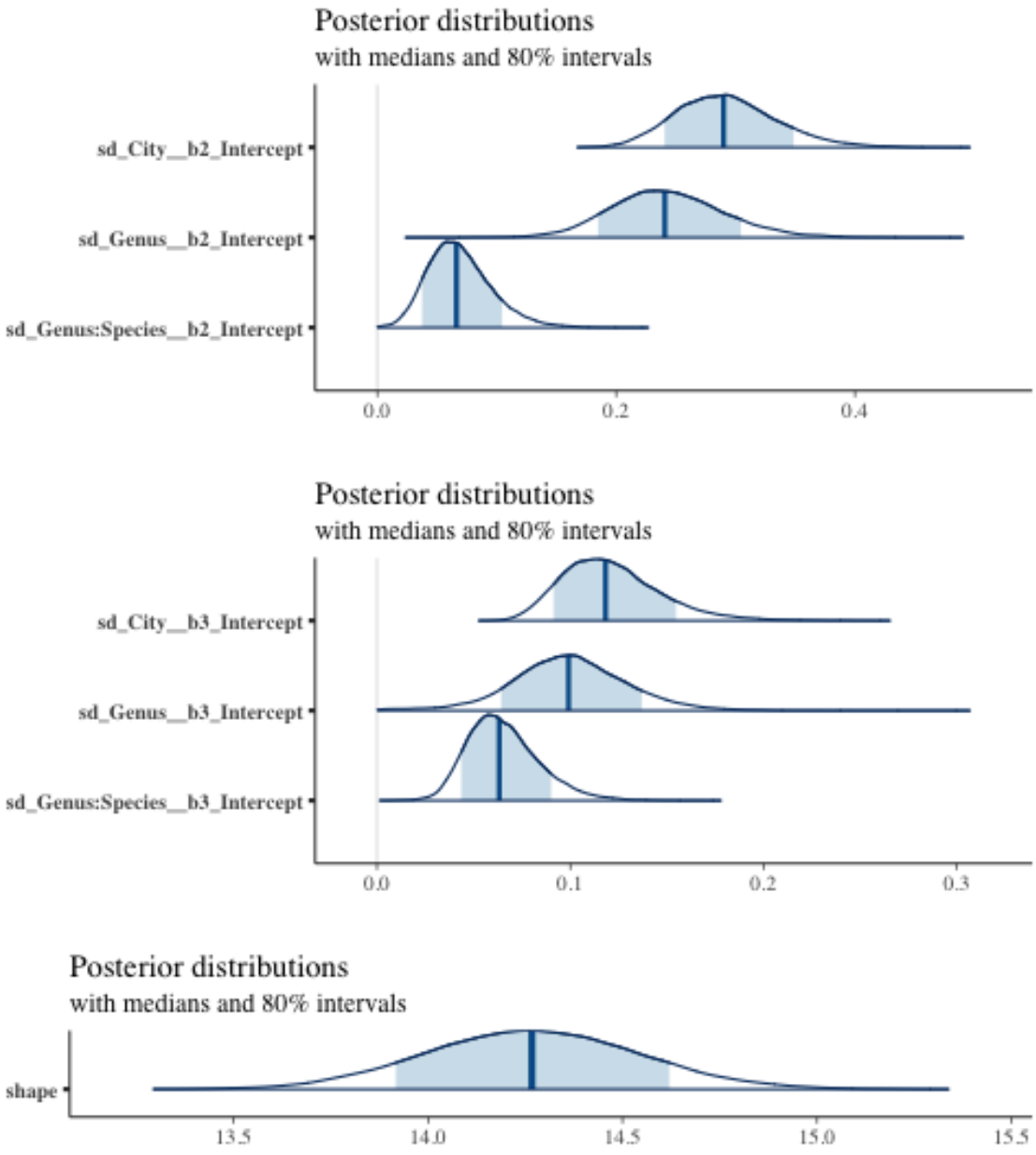
Posterior distributions with 80% interval and median for parameters in




Posterior distributions
with medians and 80% intervals
b_b3_precip
b_b3_gdd
b_b3_precip:gdd
0.00
0.02
0.04
0.06

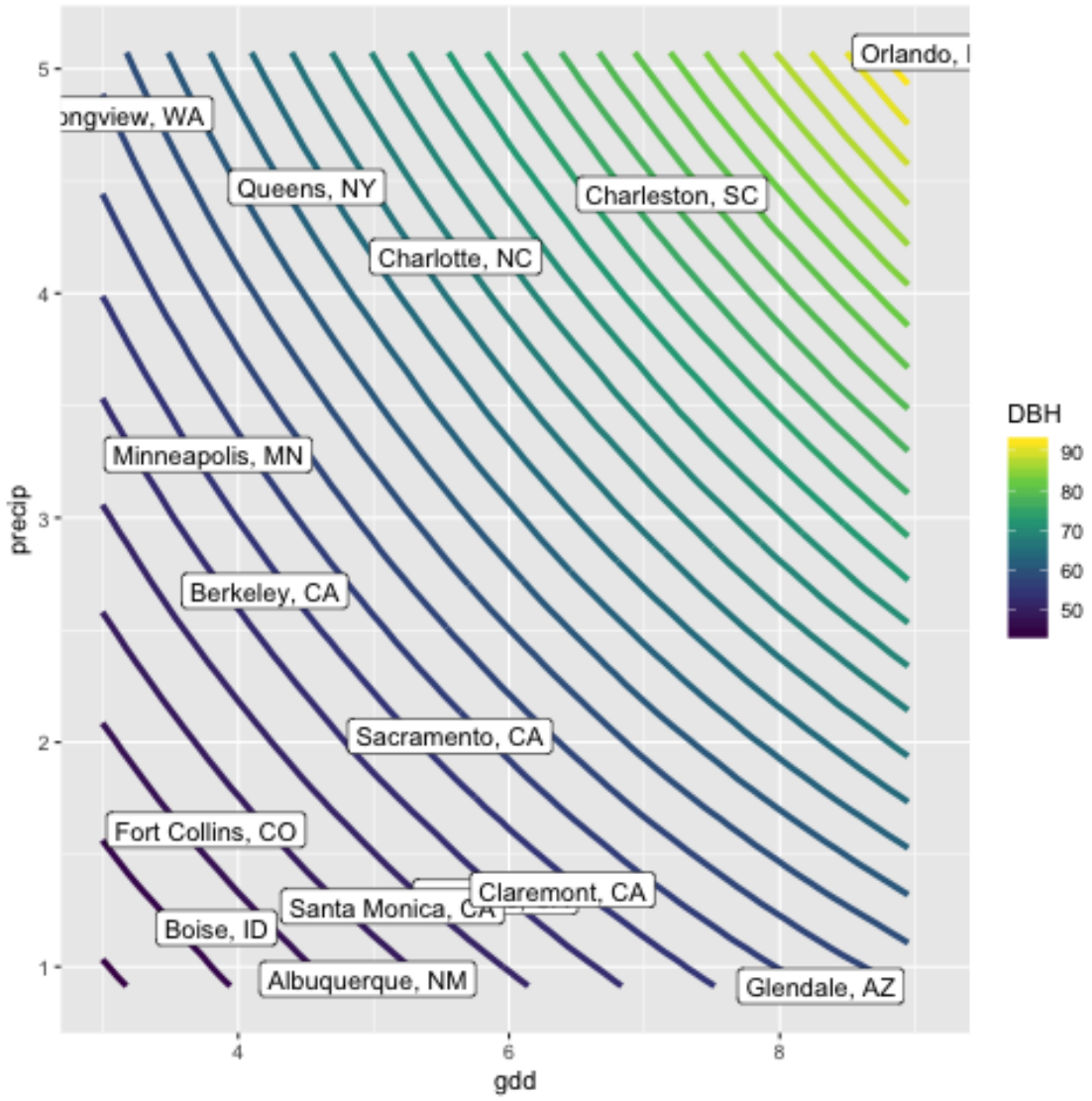

Posterior distributions
with medians and 80% intervals
sd_City__b0_Intercept
sd_Genus__b0_Intercept
sd_Genus:Species__b0_Intercept
0.0
0.5
1.0
1.5


Posterior distributions
with medians and 80% intervals
sd_Genus__b1_Intercept
sd_Genus:Species__b1_Intercept
0.0
0.5
1.0
1.5

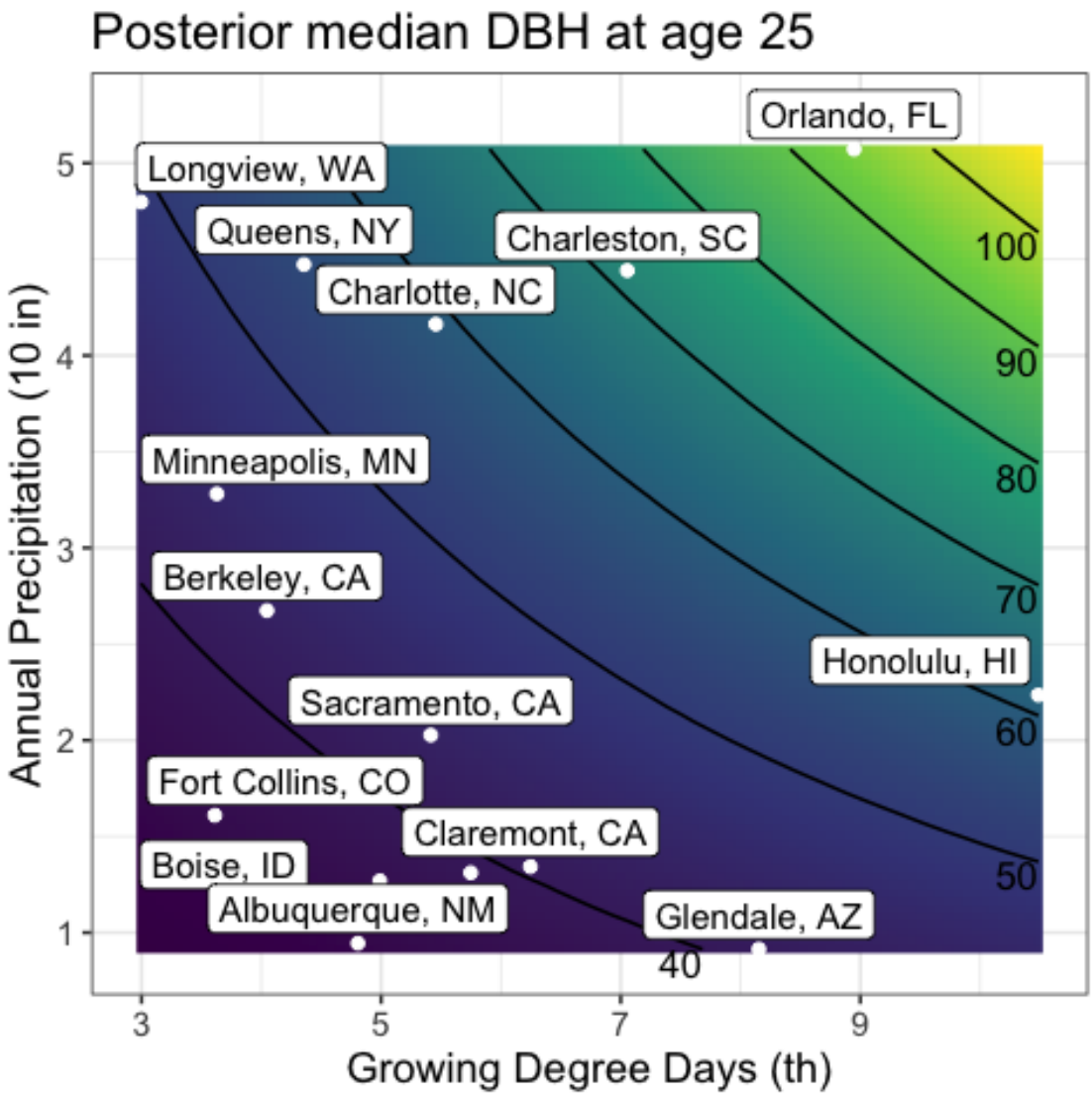


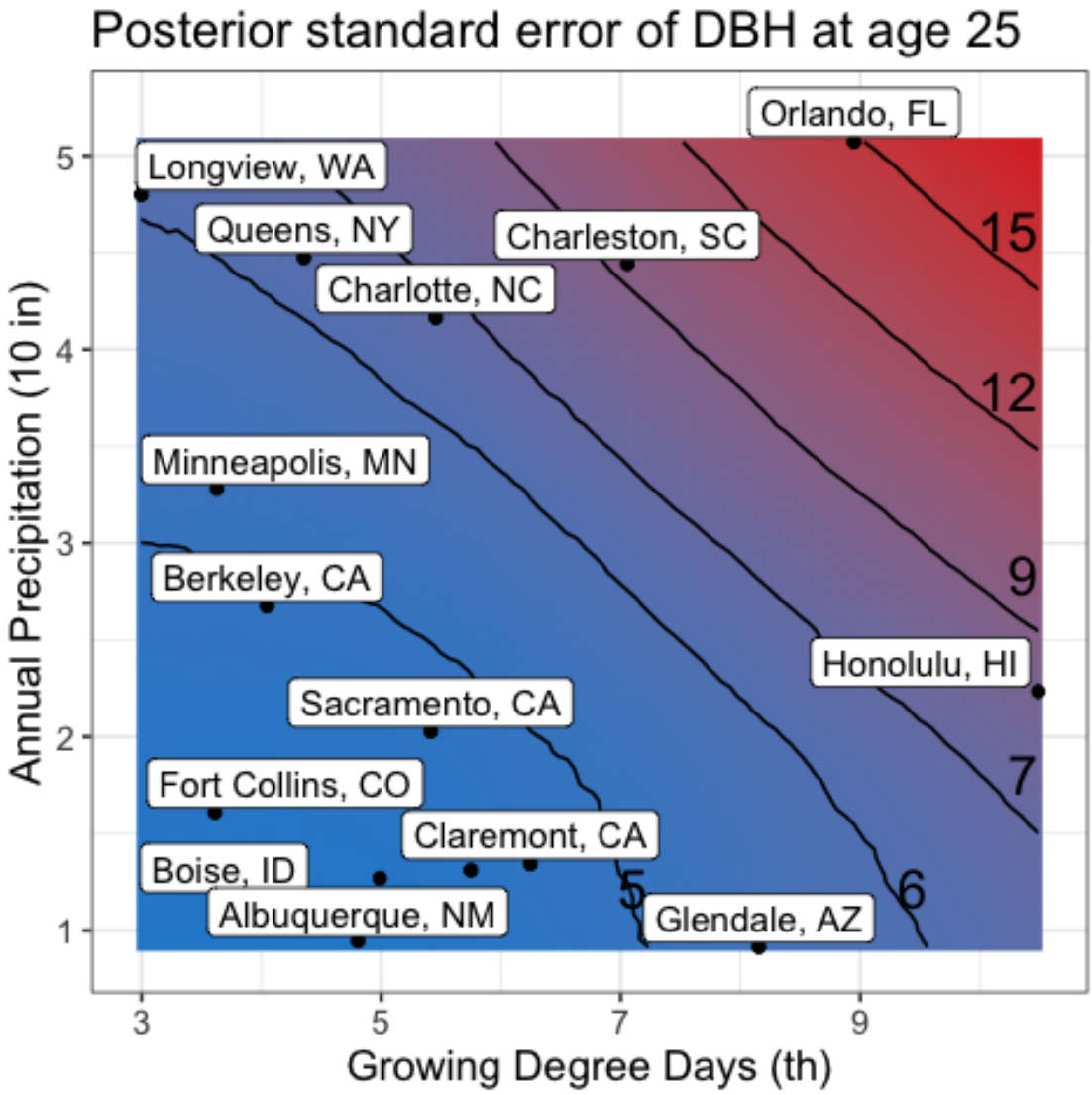
climate effects

There is a positive effect of growing degree days (gdd) and annual precipitation (precip) on tree diameter (dbh), and a postive interaction between the two. Marginal effects of climate on DBH in . There is an estimated 40cm difference in dbh between an average tree in Orlando, FL and one in Boise, ID.



Include an uncertainty / standard error panel also.


Posterior median DBH at age 25
Orlando, FL
Longview, WA
Queens, NY
Charleston, SC
Charlotte, NC
Minneapolis, MN
Berkeley, CA
Honolulu, HI
Sacramento, CA
Fort Collins, CO
Claremont, CA
Boise, ID
Albuquerque, NM
Glendale, AZ
100
90
80
70
60
50
40
Annual Precipitation (10 in)
1
2
3
4
5
3
5
7
9
Growing Degree Days (th)


Posterior standard error of DBH at age 25
Orlando, FL
Longview, WA
Queens, NY
Charleston, SC
Charlotte, NC
Minneapolis, MN
Berkeley, CA
Honolulu, HI
Sacramento, CA
Fort Collins, CO
Claremont, CA
Boise, ID
Albuquerque, NM
Glendale, AZ
15
12
9
7
6
5
Annual Precipitation (10 in)
Growing Degree Days (th)
1
2
3
4
5
3
5
7
9

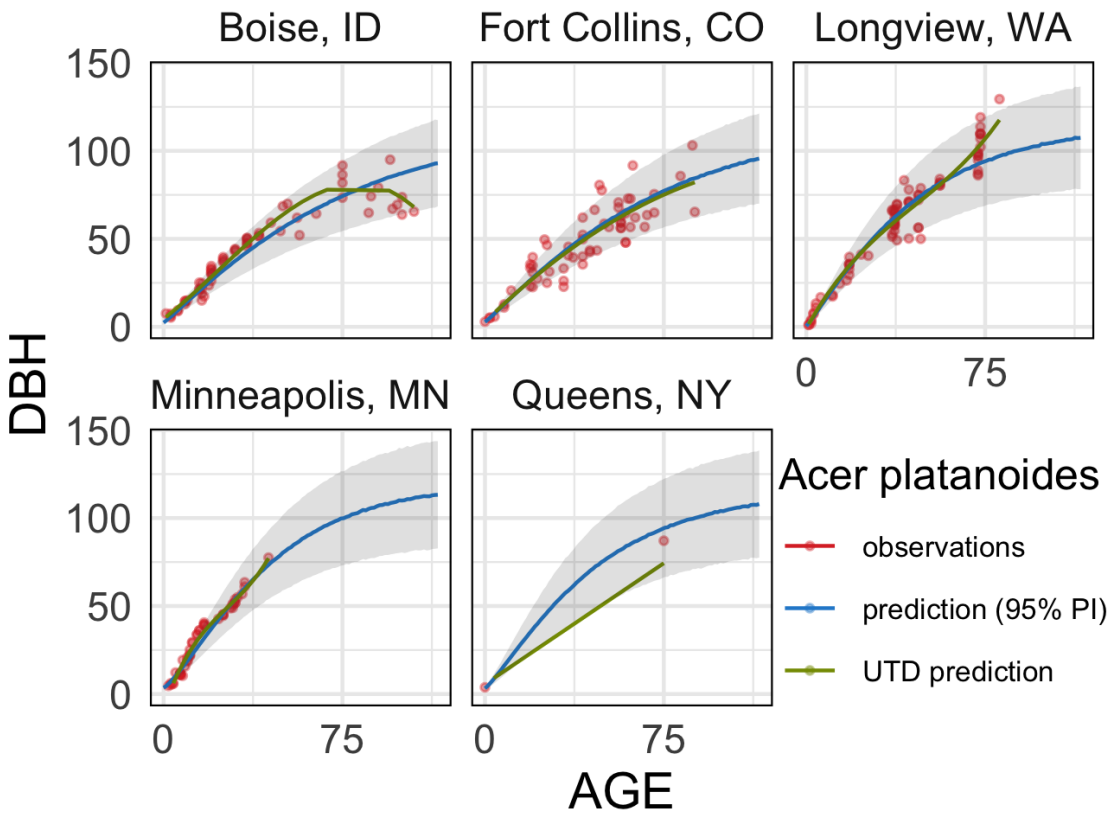
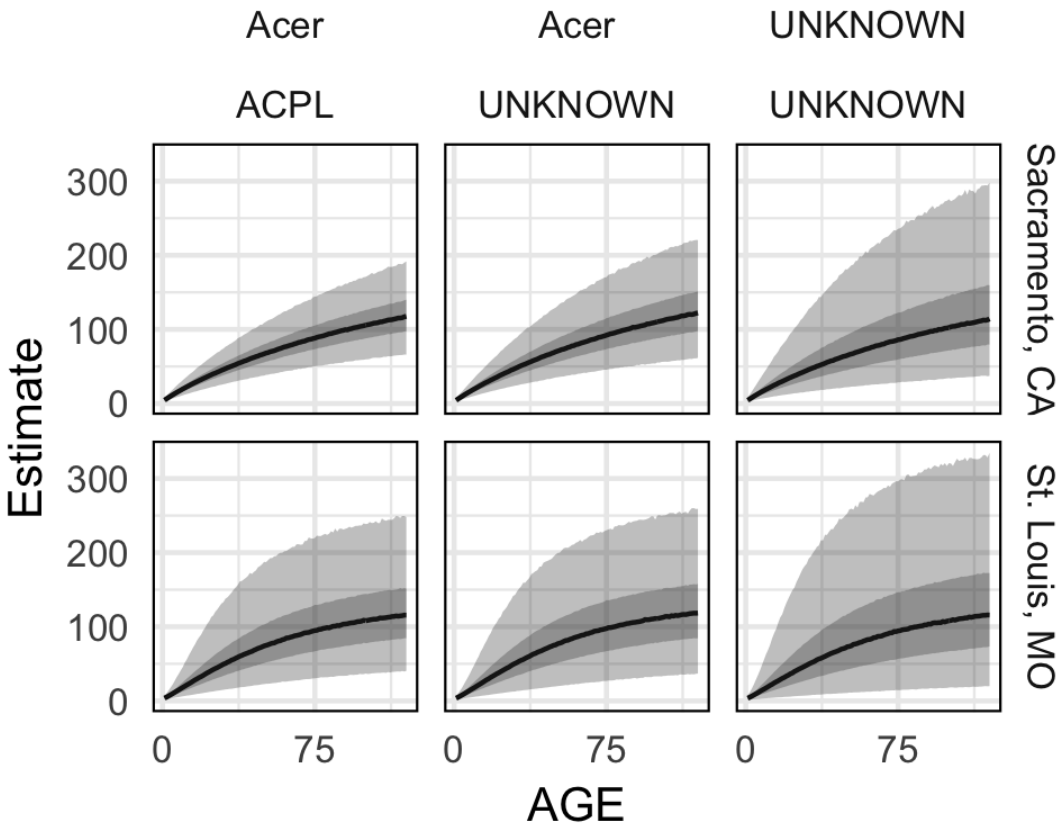


Figure 4: red points are observations, blue lines are predictions of the model and shading indicates 95% prediction interval, green lines show UTD equations

Comparing to existing equations

illustrative comparisons

Uncertainty increases when predicting out of sample cities, genera, and species



Discussoins

interpreting parameter estimates: For most species there are no data near the true asymptote and so this parameter should be considered a highly uncertain estimate of the real maximum dbh of a tree. Pragmatically, it models/causes the slowing of the diameter growth of a tree as it ages.

An early version of the UTD equations didn't have as much data, but their approach

modified parameters based on the number of frost free days Frelich (1992)

Peper et al. (2001a) - tested modified weibull following Frelich (1992), but went with logarithm regression model because it had the best in-sample fit. We think weibull would have the best out of sample fit.

Peper et al. (2001a) noted that differences in the dimensions of sweetgum and camphor in Modesto and in Santa Monica were due to different pruning regimes, cultural practices. This shows the challenges in modelling. There are some difficult to capture human cultural elements.

several trees

- what is the distribution of maximum age by species? many have very young
- or the distribution of apps max?
- We need to be able to predict to older ages if we want to make realistic predictions.

make better

- more data, duh, perhaps used results to identify where to sample
- more cities, this is important for interpolation across climate space. Could allow for nonlinear relationships and for more variables.
- better climate predictors
- interactions between climate and species.
- use phylogenetic distance, gaussian process, instead of multiple levels of taxonomy.
- an easy one is to nest genus with tree functional type (broad leaf deciduous, evergreen conifer, etc.)
- extend species with species level predictors (leaf morphology, wood characteristics, shade tolerance, etc).

- smarter priors (e.g. max dbh based on champion trees?) is this possible? I think it would be a very neat extension, but need to think about how these champions are not urban trees most the time. They provide the upper limit on the asymptote, but for urban trees the asymptote could be quite lower.
- incorporate uncertainty in AGE
- There were only 4 trees in Queens NY sampled.
- repeat measures on the same individuals would help much.
- Get more trees in the database, UFIA effort?
- add varying intercept for location of tree in sidewalk/underpowerlines etc.
- There are some funny things with minneapolis data. is it from the same individual for each species?
- challenges of separating β_1 and β_3 without old trees.

this different? Well the urban part, management practices like topping, pollarding, pruning, drastically alter growth and tree dimensions. Others have discussed the difference between forest and urban/open grown trees a fair bit.

because of the richness of the dataset, the geographic range of that these equations can be applied to is larger than many other equations which are usually for a particular speices in a particular region.

Unlike many past studies we do not have repeat measures on individual trees, but it would be straightforward to incorporate such data in the model.

References

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1):nil.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):nil.

Frelich, L. E. (1992). Predicting dimensional relationships for twin cities shade trees.

Hall, D. B. and Bailey, R. L. (2001). Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *Forest science*, 47(3):311–321.

Lappi, J. and Bailey, R. L. (1988). A height prediction model with random stand and tree parameters: An alternative to traditional site index methods. *Forest Science*, 34(4):907–927.

Li, R., Stewart, B., and Weiskittel, A. (2011). A bayesian approach for modelling non-linear longitudinal/hierarchical data with random effects in forestry. *Forestry*, 85(1):17–25.

McPherson, E. G. and Simpson, J. R. (1999). Carbon dioxide reduction through urban forestry. *Gen. Tech. Rep. PSW-171, USDA For. Serv., Pacific Southwest Research Station, Albany, CA*.

McPherson, E. G., van Doorn, N. S., and Peper, P. J. (2016a). Urban tree database.

McPherson, E. G., van Doorn, N. S., and Peper, P. J. (2016b). Urban tree database and allometric equations.

Nothdurft, A., Kublin, E., and Lappi, J. (2006). A non-linear hierarchical mixed model to describe tree height growth. *European Journal of Forest Research*, 125(3):281–289.

Peper, P. J., Alzate, C. P., McNeil, J. W., and Hashemi, J. (2014). Allometric equations for urban ash trees (fraxinus spp.) in oakville, southern ontario, canada. *Urban Forestry & Urban Greening*, 13(1):175–183.

Peper, P. J., McPherson, E. G., and Mori, S. M. (2001a). Equations for predicting diameter, height, crown width, and leaf area of san joaquin valley street trees. *Journal of Arboriculture*, 27(6):306–317.

Peper, P. J., McPherson, E. G., and Mori, S. M. (2001b). Predictive equations for dimensions and leaf area of coastal southern california street trees. *Journal of Arboriculture*, 27(4):169.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Troxel, B., Piana, M., Ashton, M. S., and Murphy-Dunning, C. (2013). Relationships between bole and crown size for young urban trees in the northeastern usa. *Urban forestry & urban greening*, 12(2):144–153.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432.

Weiskittel, A. R., Hann, D. W., Kershaw, J. A., and Vanclay, J. K. (2011). *Forest Growth and Yield Modeling*. []. John Wiley & Sons, Ltd.

to address

- fitting multilevel models in the frequentist way (lme4) seems quite complex and is much over my head. Talk to Jun about this concern.
- why model correlations of parameters within groups?