

Multilevel Urban Tree Allometric equations

Burkner

October 26, 2018

Introduction

Multilevel models have been used for decades in tree growth equations (Lappi and Bailey 1988). Indeed the test dataset, "orange", in the statistical programming language, R, is used to demonstrate the fitting of nonlinear mixed effects models cite. 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

BRMS

limitations of existing approach and how our model addresses them.

Existing Approach

Can predict negative diameters

Tests several model forms

extrapolation to higher ages than measured is fraught because equation form determined by best fit, not a separate model for each species and city combination. in a case with 2 observations the model had to what equation to use for unobserved species and city combinations is not clear. (e.g. should species or c
Equations derived in the 17 cities are meant to apply to work in 17 large and heterogeneous regions.

maybe the framework should be broader, just growth equations for all trees with urban tree as a case study/example of application.

There are two story lines.

1. use of newerish statistical package to fit a pretty complex model
2. that growth equations for urban trees are newish, can be improved, and are important because they are fundamental to predicting ecosystem services.

Big problem - urban ecosystem services depends on trees. predicting depends on growth narrower within - urban trees growth differently, there had been a lack of data. yet narrower summary of approach and results

1. theoretically sound model form (through use of weibull)
2. expansion of predictive envelope/potential (through use of multilevel modeling).

Objective is to show how multilevel models can be used to predict tree growth

1. Information about species functional groups and phylogeny to predict growth for unobserved species

2. Information about
3. Information about climate to predict growth in unobserved cities
- 4.

Urban trees have important ecological effects in cities that can affect human wellbeing. Forecasting ecosystem service provision by trees requires a model of tree growth, since most services are directly related to tree size (e.g. carbon storage in biomass, air filtration by leaf area). Until recently urban tree growth equations have been rare because of lacking data. In 2016 the US Forest Service published the "Urban Tree Database and Allometric Equations", a valuable dataset of the most common tree species in a dozen cities across the US. In addition to allometric equations, they also provide growth equations.

While they are great step forward and an incredible resource, their growth equations have limitations. First, their method of model selection simply chose the curve with the lowest AIC. This meant that any amount of extrapolation was highly suspect. In the extreme case where a species only had two observations in a city, the only possible fit was a line. Second, they only fit equations for the most common species in each of the cities they sampled. It is not clear how to modify an equation to fit a unobserved species or a unobserved city. If red maple was not a species sampled in my city, should I use the red maple equation from a climatically dissimilar city? Or should I use the equation from the species most similar to red maple from my city? What if I want to make a prediction about an unobserved species in an unobserved city?

We remove these limitations by fitting a multilevel weibull model to the data. The weibull curve has a long history of use in forestry and captures asymptotic growth well. While it may not provide the best fit to every dataset, it realistically models growth given what we know about biological constraints, is flexible, and is less unreasonable to extrapolate from. The multilevel model allows information to be shared across species and across cities, and, when combined with species' phylogenetic data and cities' climate data, allows for predictions at

combinations of cities and species with little or no individual level data, provided we know genus/functional group of the species and the climate of the city.

DBH for illustrative purposes.

on selecting a model form Weiskittel et al. (2011): "This reality should lead to model forms that are complex enough to accurately and adequately characterize the expected major behaviors of the population, but simple enough to avoid being overly influenced by the sample's peculiarities."

past work

Nothdurft et al. (2006)

Yang et al. (2009)

Hall and Bailey (2001)

Weiskittel et al. (2011)

Growth estimate of Eucalyptus stands based on nonlinear multilevel mixed-effects model theory

Miina, J., Heinonen, J., 2008. Stochastic simulation of forest regeneration establishment using a multilevel multivariate model. *Forest Science* 54, 206–219.

Miina, J., Saksa, T., 2006. Predicting regeneration establishment in Norway spruce plantations using a multivariate multilevel model. *New Forests* 32, 265–283.

Uzoh, F.C.C., Oliver, W.W., 2008. Individual tree diameter increment model for managed even-aged stands of ponderosa pine throughout the western United States using a multilevel linear mixed effects model. *Forest Ecology and Management* 256, 438–445.

Kershaw Jr., J.A., Benjamin, J.G., Weiskittel, A.R., 2009. Approaches for modeling vertical distribution of maximum knot size in black spruce: A comparison of fixed- and mixed-effects nonlinear models. *Forest Science* 55, 230–237.

Gregoire, T.G., Schabenberger, O., 1996. A non-linear mixed-effects model to predict cumulative bole volume of standing trees. *Journal of Applied Statistics* 23, 257–271.

Leites, L.P., Robinson, A.P., 2004. Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. *Forest Science* 50, 204–212.

Sharma, M., Parton, J., 2007. Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach. *Forest Ecology and Management* 249, 187–198.

Wang, Y., LeMay, V.M., Baker, T.G., 2007b. Modelling and prediction of dominant height and site index of *Eucalyptus globulus* plantations using a nonlinear mixed-effects model approach. *Canadian Journal of Forest Research* 37, 1390–1403.

Urban Tree Growth Modeling https://www.fs.fed.us/psw/publications/mcpherson/psw_2012_mcpherson001.pdf

Urban Tree Database and Allometric Equations E. Gregory McPherson, Natalie S. van Doorn, and Paula J. Peper

The economic, social, and ecological benefits of trees are directly related to their size, as indicated by leaf area, crown volume, and bio- mass (Scott et al. 1998, Stoffberg et al. 2010, Xiao et al. 2000a). Growth equa- tions underpin the calculations produced by many computer models used in urban forestry, such as i-Tree, National Tree Benefit Calculator, OpenTreeMap, and ecoSmart Landscapes (fig. 1).

Quotes from the utd publication:

The development of allometric equations for urban open-grown trees has been sporadic. Fleming (1988) measured trees in New Jersey having full healthy crowns to develop linear relationships between d.b.h., height, crown spread, and age. Frelich (1992) measured only healthy trees (12 species, 221 trees total) growing in Minneapolis and St. Paul, Minnesota, to predict linear size relationships. Nowak (1994b) developed an allometric equation for leaf area based on data from park trees in Chicago. Tree dimensions and leaf area were predicted for the most abundant street tree species in Modesto and Santa Monica, California (Peper et al. 2001a, 2001b). In New Haven, Connecticut, Troxel et al. (2013)

developed allometric equations for predicting d.b.h. from age and height, crown diameter, and crown volume from d.b.h. for early growth (15 years) of 10 street tree species. Outside of North America, growth equations have been developed for street-side *Tilia* species in Copenhagen, Denmark (Larsen and Kristoffersen 2002), and *T. cordata* Mill., *Fraxinus excelsior* L. and *Aesculus hippocastanum* L. in Warsaw, Poland (Lukaszkiewicz and Kosmala 2008, Lukaszkiewicz et al. 2005). Predictive models were developed from allometric data for five street tree species in northeastern Italy by Semenzato et al. (2011). Stoffberg et al. (2008) used allometric relationships between age and d.b.h., height, and crown diameter to estimate dimensions at 10, 15, and 30 years after planting for three street tree species in Tshwane, South Africa. The allometric equations from all these studies reflect the effects of local site conditions, management practices, and growing season on growth, limiting application outside their region of origin (fig. 5).

urban trees equations

"Growth equations for forest trees may not be directly transferable to open-grown urban trees because they grow and partition bole, branch, twig, and leaf biomass differently (Anderegg et al. 2015, Nowak 1994a, Peper and McPherson 1998) (fig. 4). For example, in forests, tree crowns compete for limited space and may not reach their maximum expansion potential (Martin et al. 2012)." from the

past modelling

This report presents the third, most recent and most complete sets of growth equations. The equations presented in this report were developed using more sophisticated statistical methods than before. For example, in the first studies, logarithmic regression and exponential models predominantly provided the best fits to measured data (Peper et al. 2001a, 2001b). In these equations, the best

model fits ranged from polynomials (from simple linear to quartic) to logarithmic and exponential models (Peper et al. 2014). The newest equations have been integrated with numerical models of tree benefits in the ecoSmart Landscapes software (McPherson et al. 2014).

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



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

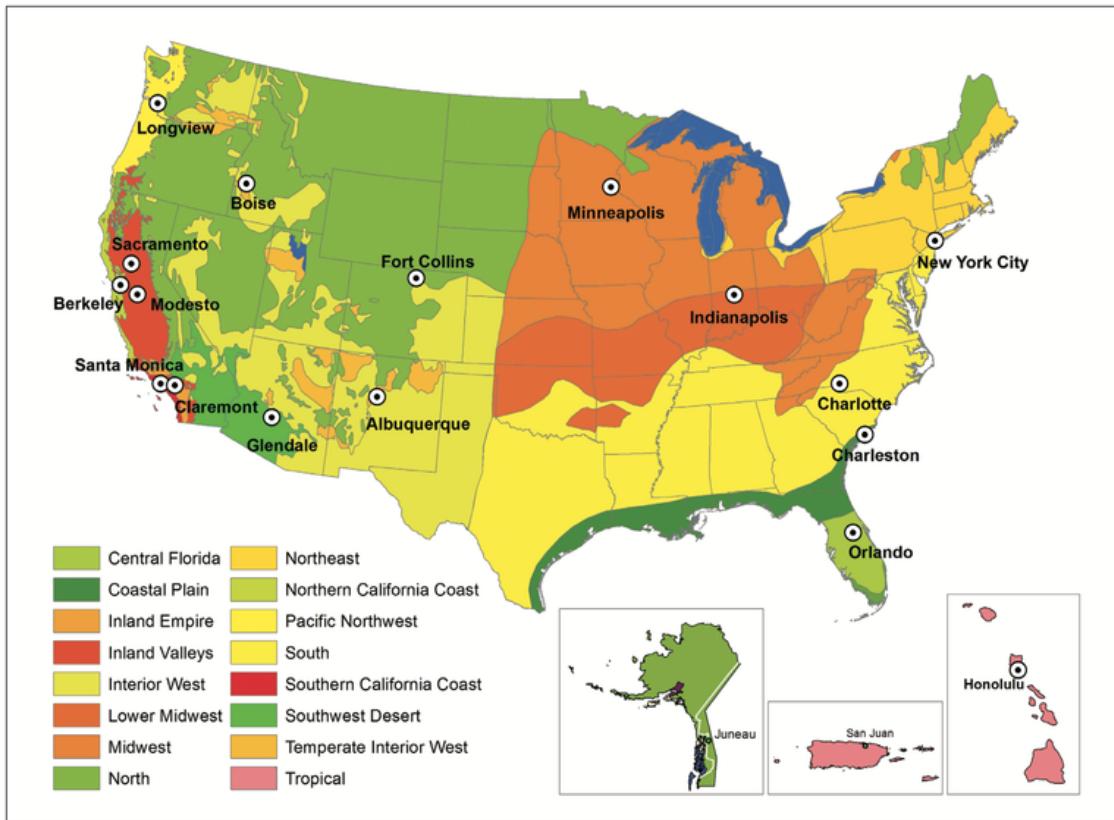


Figure 9—Climate zones were aggregated from 45 Sunset climate zones into 16 zones. Each zone has a reference city where tree growth data were collected. Sacramento, California, was added as a second reference city (with Modesto) to the Inland Valleys zone.

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

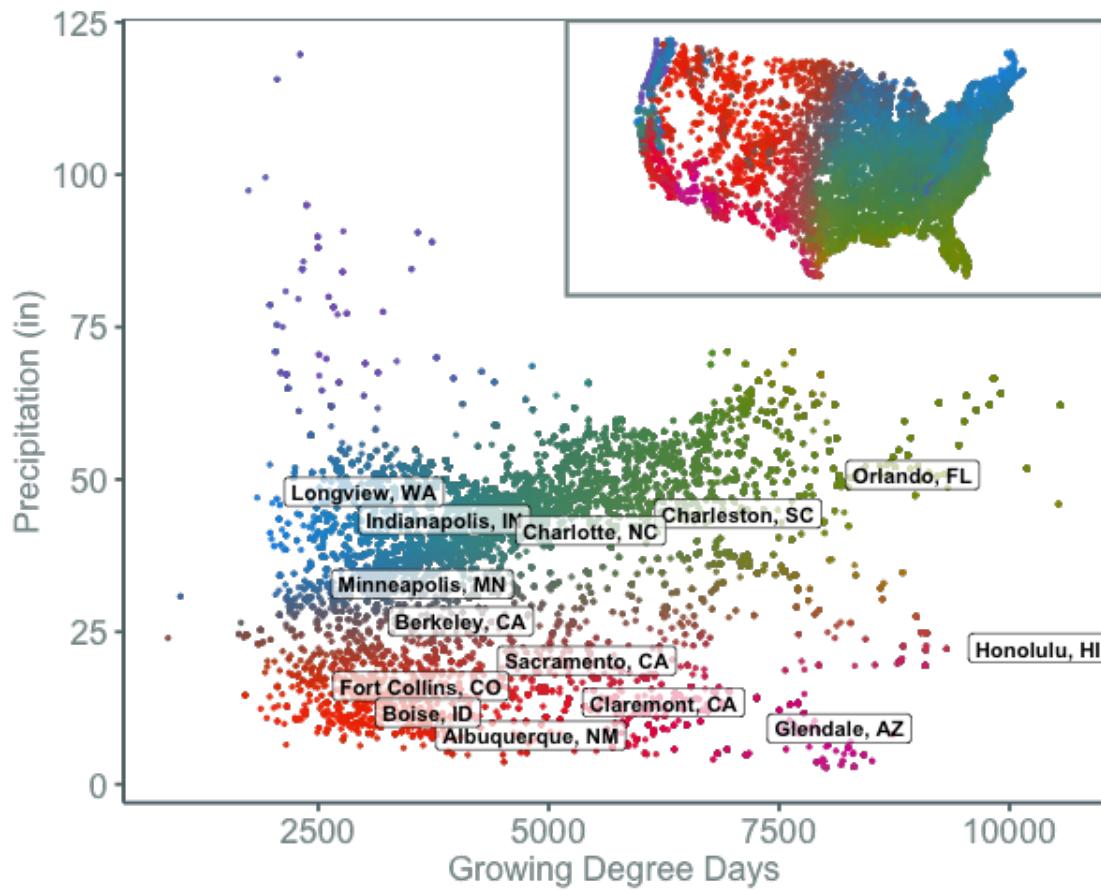


Figure 3: US census tract centroids with UTD reference cities overlaid in growing degree day (GDD) and precipitation climate space and matching color gradient in geographic space. The reference cities cover climate space well, and variation in precipitation and growing degree days is continuous.

conterminous US plotted in GDD and precipitation space. We approximated 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 continuously across geographic space in a way that better captures the natural gradients of climate.

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 asymptote. Instead the asymptote in our curve could be considered a pragmatic way to constrain diameter growth, or a practical asymptote. While trees could theoretically continue to increase in diameter indefinitely, they don't in reality. The asymptote 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 that meet these criteria and could be used.

Another characteristic of tree growth curves is heteroscedasticity, 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.

Model

(Notation below. I don't use conventional subscript letters. The nesting of species within genus isn't obvious either until lower levels. I did this because genus could also end up being nested. There is a fair bit going on here, suggestions for how to improve are much appreciated.)

$$y_i \sim \text{Gamma}(\mu_i, \alpha_y)$$

$$\mu_i = \beta_{0sc[i]} + \beta_{1sc[i]}(1 - \exp(-\beta_{2sc[i]}x_i^{\beta_{3sc[i]}}))$$

where:

y_i is the diameter at breast height of tree i and has a gamma distribution with mean, μ_i , and shape, α_y .

β_0 is the intercept, or the diameter of a tree at time of transplanting.

β_1 (plus β_0) is the asymptote of the sigmoidal weibull curve. 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.

β_2 and β_3 affect the rate of growth. β_2 provides flexibility to have slow or fast growth at young ages (small x).

All β 's must be positive and they are likely correlated with one another, especially β_1 and β_3 . Without very old trees that are close to their asymptotic dbh, it is harder to separate these two parameters.

for each beta, $j = 0, 1, 2$, species, s , and each city, c .

$$\beta_{jsc[i]} = \beta_j + \gamma_{js[i]} + \delta_{jc[i]}$$

for beta₃

$$\beta_{3sc[i]} = \beta_3 + \tau_1 * Precip_i + \tau_2 * GDD_i + \tau_3 * (Precip_i * GDD_i) + \gamma_{3s[i]} + \delta_{3c[i]}$$

where β_j is the mean for β coefficient β_j . γ_{js} is the contribution of genetic (species) effect for species s on β_j . δ_{jc} is the city effect for city c on β_j .

Species effect:

$$\gamma_{js[i]} \sim N(\gamma_{jg[i]}, \sigma_{js[i]})$$

Genus effect:

$$\gamma_{jg[i]} \sim N(0, \sigma_{jg[i]})$$

for each j. City random effect:

$$\delta_{jc} \sim N(0, \sigma_{jc[i]})$$

Priors:

Priors were selected to be slightly informative and make very biologically unreasonable parameters improbable. The quantity of data overwhelms the priors

$$\text{for gamma } \alpha = \mu^2/\sigma^2 \quad \beta = \mu/\sigma^2$$

parameter	mu	sd	alpha (of gamma)	beta (of gamma)
β_0	3	1	9	3
β_1	1.75	.3	34.027778	19.444444
β_2	1.25	.15	69.444444	55.555556
β_3 (no climate)	1	.15	44.444444	44.444444
β_3 intercept (climate)	.6	.15	16.	26.666667
β_3 intercept (climate,genus,species)	.3	.15	4.	13.333333

$$\beta_0 \sim Gamma(4, 1.33)$$

$$\beta_1 \sim \text{Gamma}(34, 19.4)$$

$$\beta_2 \sim \text{Gamma}(69.4, 55.5)$$

$$\beta_3 \sim \text{Gamma}(44.4, 44.4)$$

these priors selected because they create a wide range of possible mean curves, but they are physically possible.

variability by cities and species, something realistic for each parameter all sigmas: for β_0 , the intercept. realistic values for this range from 1 to maybe 10 (that would be a very large average size to plant). Therefore, the sd for δ , $\sigma_{0c[i]}$, is likely less than 1.5. I'll set the prior so that there's a 95% chance it's less than 1.5. Variation between genus/species, the sd for γ , is probably similar.

$((10 - 1) / 2) / 3 = 1.5$. (take the range of possible values, assume it is ~99.7% range of normal, find the sd by cutting in half and dividing by 3). This is the highest sd I would expect. Set the prior so that I'm saying I think there is a 99.7% chance I think the the sd is less than this. So I make the sd on the prior, 1/3 the highest sd I think is possible.

species is half of genus

$$\sigma_{0c[i]} \sim \text{Normal}(0, .5)$$

$$\sigma_{0g[i]} \sim \text{Normal}(0, .5)$$

$$\sigma_{0s[i]} \sim \text{Normal}(0, .25)$$

When fit in stan, β_1 is multiplied by 100, so that it is on a similar order of magnitude as the other parameters (around 1) and the interpretation is the asymptotic diameter in meters (rather than centimeters). Possible values for β_1 might range from .5 to 3 between different genera. Differences between species within genera will likely be less. I expect there to be less difference between cities in the average asymptote.

$((3 - .5) / 2) / 3 = .4167$ (I'll set genus to slightly less than this because there is also species variation, which I'll set to .2 since I expect most the time species within the same genus to be quite similar, probably within about 1m of one another). I'll set city variation to the same as species.

$$.4 / 3 = .133 \quad .2 / 3 = .067$$

for β_1

$$\sigma_{1_g[i]} \sim \text{half-Normal}(0, .13)$$

$$\sigma_{1_s[i]} \sim \text{half-Normal}(0, .06)$$

$$\sigma_{1_c[i]} \sim \text{half-Normal}(0, .06)$$

for β_2 and β_3 I don't have as good of intuition for β_2 or β_3 , but from looking at curves created with different values, I expect their variation to be smaller than for β_0 and β_1 . For a start I'll set the variation to about half of what I set it for β_1 .

$$\sigma_{2_g[i]} \sim \text{half-Normal}(0, .06)$$

$$\sigma_{2_s[i]} \sim \text{half-Normal}(0, .03)$$

$$\sigma_{2_c[i]} \sim \text{half-Normal}(0, .03)$$

$$\sigma_{3_g[i]} \sim half-Normal(0, .06)$$

$$\sigma_{3_s[i]} \sim half-Normal(0, .03)$$

$$\sigma_{3_c[i]} \sim half-Normal(0, .03)$$

The coefficients for climate variables on β_1 gdd are in thousands. range in data is about 6000 gdd, or 6 thousands gdd. I expect the effects of gdd, precip, and their interaction to be positive. I expect the increase in asymptote across this range to be not too big, maybe as little as .1m and as much as 1m. So the coefficient might be 0.0167 to 0.167. The interaction term will also be positive, but smaller.

precip

$$\tau_1 \sim Normal(.01, .01)$$

gdd

$$\tau_2 \sim Normal(.03, .01)$$

gdd:preci

$$\tau_3 \sim Normal(.025, .01)$$

genus: many; species: many; cities: many; climate: b3linint; hetero: no; family: Gamma

- model R code

```
library(dplyr)
library(brms)
```

```

genus <- "many"
species <- "many"
cities <- "many"
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 | City) + (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), nlpars = "b0", lb = 0),
              prior(gamma(34, 19.4), nlpars = "b1", lb = 0),
              prior(gamma(69.4, 55.5), nlpars = "b2", lb = 0),
              prior(gamma(16, 26), nlpars = "b3", lb = 0),
              prior(normal(0.01, 0.015), nlpars = "b3", coef = "gdd"),
              prior(normal(0.01, 0.01), nlpars = "b3", coef = "precip"),
              prior(normal(0.005, 0.005), nlpars = "b3", coef = "precip:gdd"),
              prior(gamma(20, 1), class = "shape"),
              prior(normal(0, .3), class = "sd", nlpars = "b0", group = "City"),

```

```

prior(normal(.1, .3), class = "sd", nlpar = "b1", 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:Species"),
prior(normal(.1, .4), class = "sd", nlpar = "b1", group = "Genus"),
prior(normal(0, .1), class = "sd", nlpar = "b1", group = "Genus:Species"),
prior(normal(0, .1), class = "sd", nlpar = "b2", group = "Genus"),
prior(normal(0, .05), class = "sd", nlpar = "b2", group = "Genus:Species"),
prior(normal(0, .1), class = "sd", nlpar = "b3", group = "Genus"),
prior(normal(0, .05), class = "sd", nlpar = "b3", group = "Genus:Species")

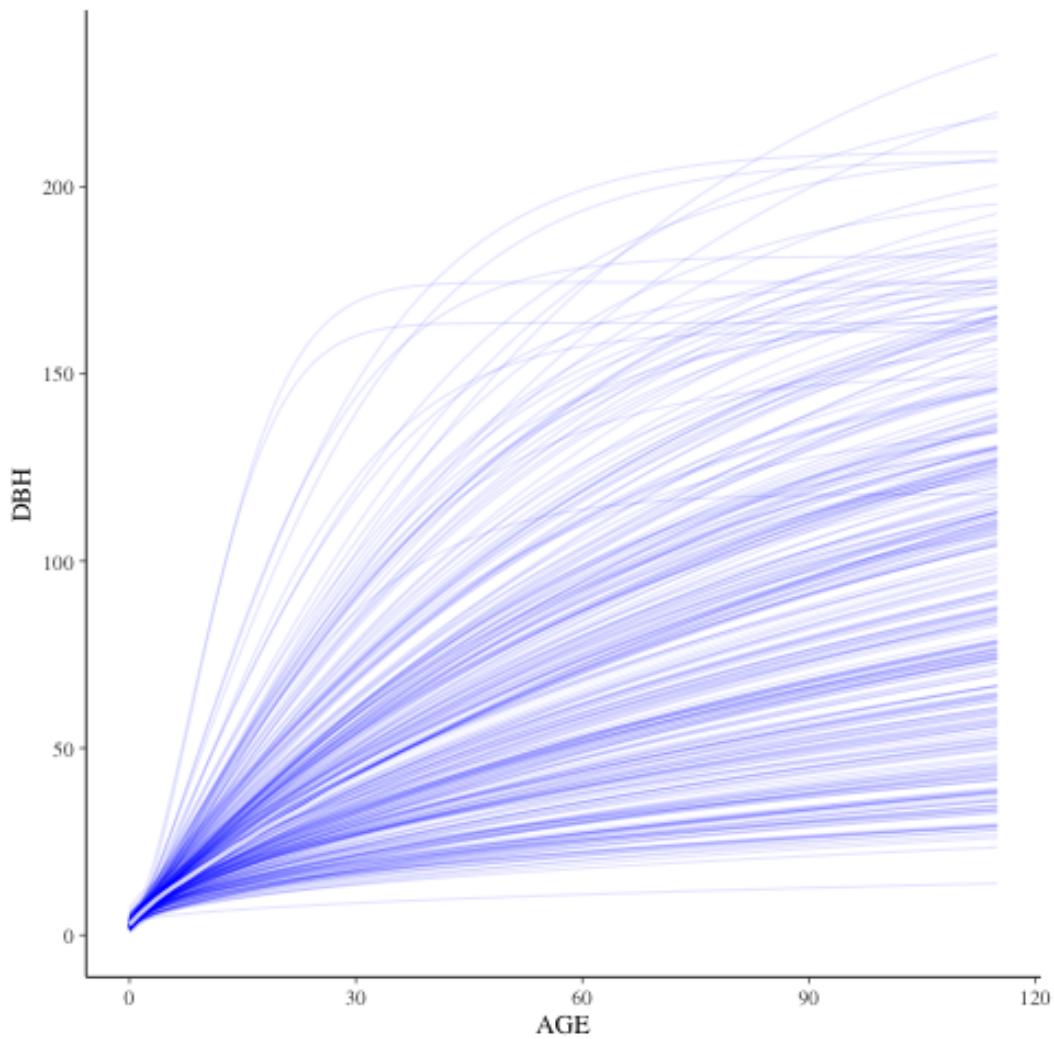
d <- readRDS("../data/age_dbh_testing_noWAR0.rds")

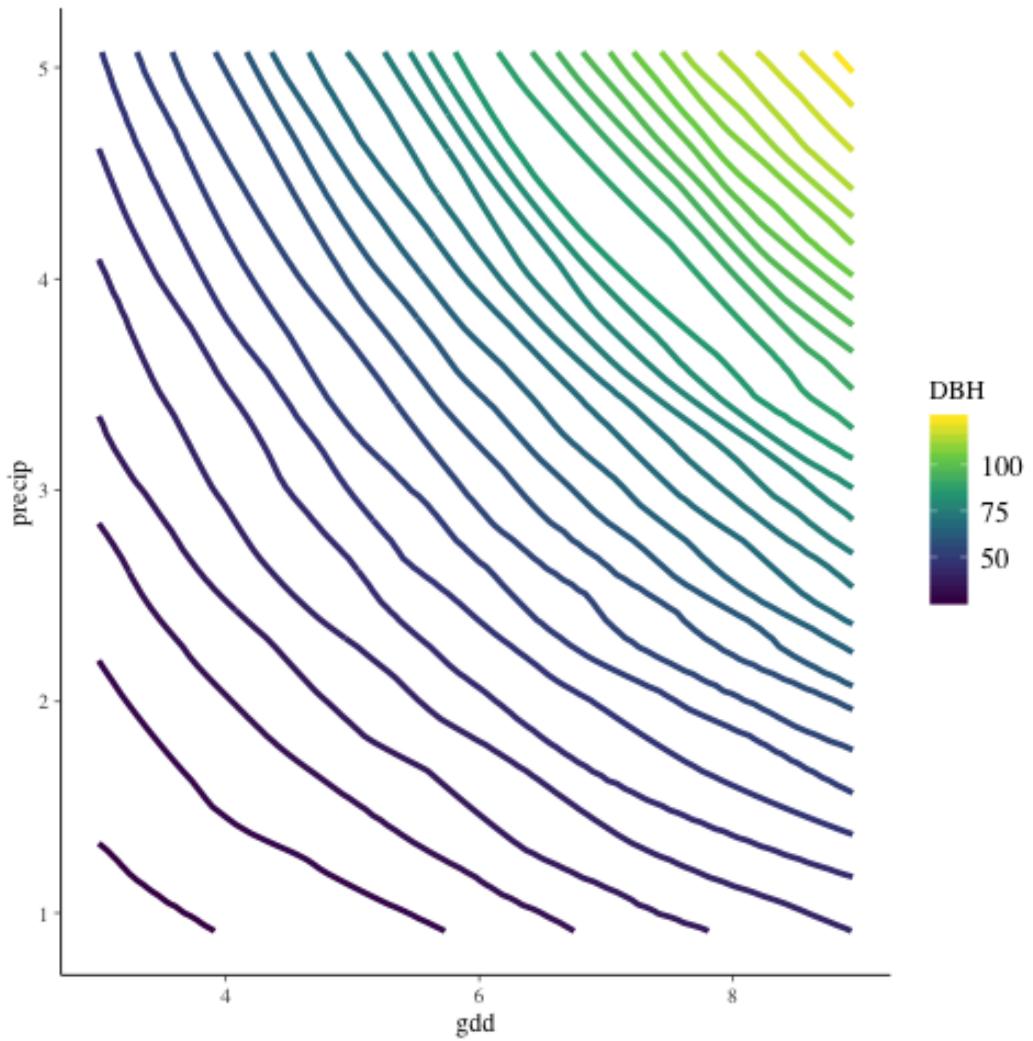
## prior_mod <- brm(form,
##                   data = d,
##                   prior = nlprior,
##                   family = Gamma("identity"),
##                   sample_prior = "only",
##                   chains = 2, cores = 2, init_r = .3, iter = 300, control = list()

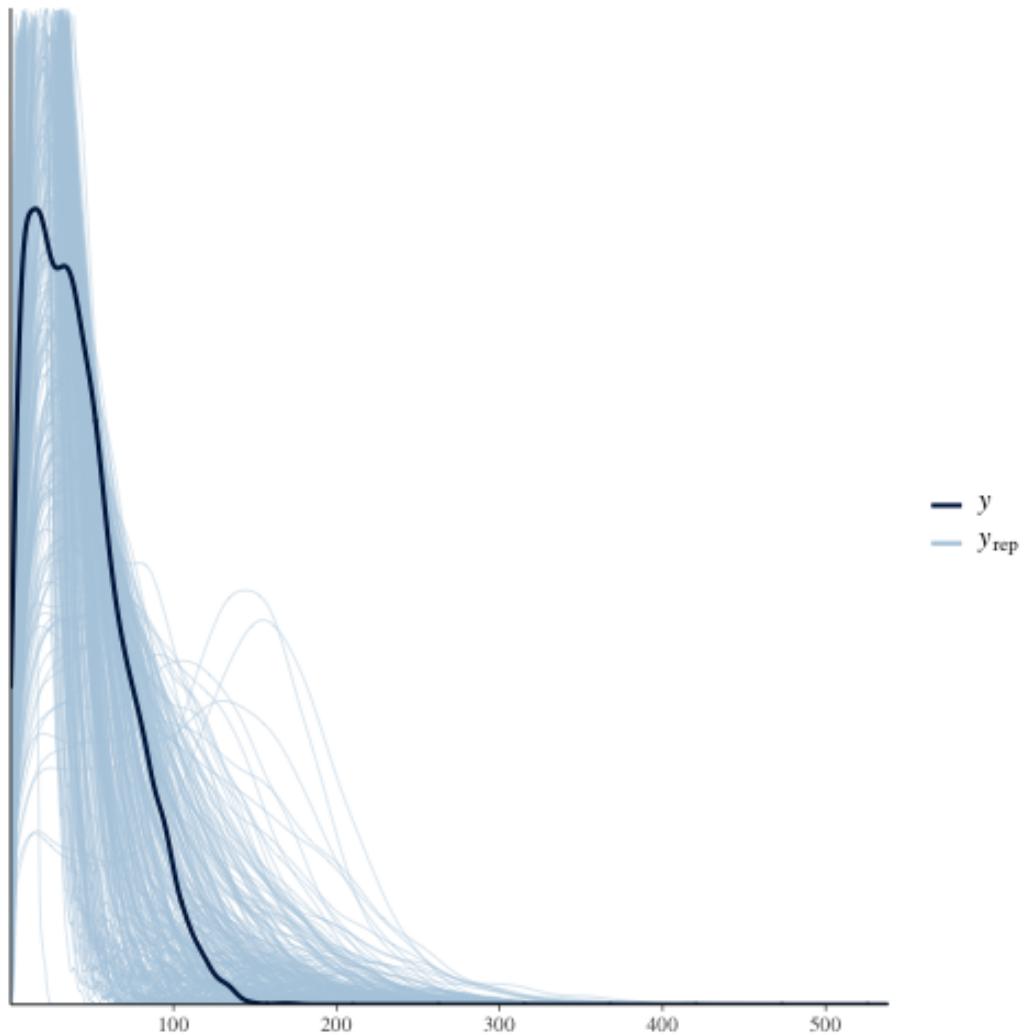
mod <- brm(form,
            data = d,
            prior = nlprior,

```

```
family = Gamma("identity"),  
chains = 12, cores = 12, init_r = .3, iter = 6000, control = list(adapt_<br/>  
saveRDS(mod, paste0("../models/genus_",genus,"_species_",species,"_cities_", cities
```







- tangle
- send to krusty

```
rsync -avz genus_species_cities_b3climate.R erker@krusty:~/allo/code/
```

- run on krusty
- run from krusty terminal

```

ssh krusty
cd allo/code
nohup R CMD BATCH genus_species_cities_b3climate.R &
exit

cat genus_species_cities_b3climate.Rout

```

- pull back from krusty

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

- assess model

```

mod_genus_many_species_many_cities_many_climate_b3linint_hetero_no_family_Gamma <-
mod <- mod_genus_many_species_many_cities_many_climate_b3linint_hetero_no_family_Gam

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 | City) + (1 | Genus/Species)

b2 ~ (1 | City) + (1 | Genus/Species)

b3 ~ precip * gdd + (1 | City) + (1 | Genus/Species)

Data: d (Number of observations: 5381)

Samples: 12 chains, each with iter = 6000; warmup = 3000; thin = 1;

total post-warmup samples = 36000

Group-Level Effects:

~City (Number of levels: 15)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(b0_Intercept)	0.83	0.13	0.61	1.11	24073	1.00
sd(b1_Intercept)	0.38	0.08	0.25	0.58	17614	1.00
sd(b2_Intercept)	0.28	0.04	0.21	0.37	20967	1.00
sd(b3_Intercept)	0.15	0.03	0.10	0.22	19786	1.00

~Genus (Number of levels: 17)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(b0_Intercept)	0.42	0.13	0.17	0.69	10239	1.00
sd(b1_Intercept)	0.61	0.13	0.39	0.90	19152	1.00
sd(b2_Intercept)	0.15	0.05	0.04	0.25	7541	1.00
sd(b3_Intercept)	0.11	0.03	0.04	0.17	9574	1.00

~Genus:Species (Number of levels: 24)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(b0_Intercept)	0.12	0.08	0.01	0.29	10158	1.00
sd(b1_Intercept)	0.12	0.06	0.01	0.25	9660	1.00
sd(b2_Intercept)	0.09	0.03	0.05	0.16	7914	1.00
sd(b3_Intercept)	0.06	0.02	0.03	0.11	9304	1.00

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
b0_Intercept	2.46	0.25	1.96	2.95	12399	1.00
b1_Intercept	1.65	0.17	1.34	1.99	14097	1.00
b2_Intercept	0.95	0.08	0.81	1.11	15411	1.00
b3_Intercept	0.93	0.08	0.78	1.07	15675	1.00

b3_precip	0.01	0.01	0.00	0.03	45951	1.00
b3_gdd	0.01	0.01	0.00	0.03	37118	1.00
b3_precip:gdd	0.00	0.00	0.00	0.01	26547	1.00

Family Specific Parameters:

	Estimate	Est.Error	1-95% CI	u-95% CI	Eff.Sample	Rhat
shape	15.03	0.29	14.46	15.61	64154	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 6 divergent transitions after warmup. Increasing adapt_delta above 0.9 m
See <http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup>

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 | City) + (1 | Genus/Species)
b2 ~ (1 | City) + (1 | Genus/Species)
b3 ~ precip * gdd + (1 | City) + (1 | Genus/Species)

Data: d (Number of observations: 5381)

Samples: 12 chains, each with iter = 500; warmup = 250; thin = 1;
total post-warmup samples = 3000

Group-Level Effects:

~City (Number of levels: 15)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(b0_Intercept)	0.48	0.05	0.38	0.59	3000	1.01
sd(b1_Intercept)	0.29	0.05	0.20	0.41	23	1.16
sd(b2_Intercept)	0.16	0.02	0.13	0.19	3000	1.00
sd(b3_Intercept)	0.09	0.01	0.07	0.12	325	1.02

~Genus (Number of levels: 17)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(b0_Intercept)	0.33	0.09	0.17	0.52	447	1.03
sd(b1_Intercept)	0.32	0.10	0.05	0.47	19	1.21
sd(b2_Intercept)	0.14	0.03	0.09	0.19	1671	1.01
sd(b3_Intercept)	0.07	0.02	0.01	0.11	25	1.17

~Genus:Species (Number of levels: 24)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(b0_Intercept)	0.05	0.04	0.00	0.13	2275	1.00
sd(b1_Intercept)	0.17	0.09	0.01	0.36	177	1.07
sd(b2_Intercept)	0.06	0.02	0.02	0.10	136	1.05
sd(b3_Intercept)	0.06	0.02	0.03	0.09	751	1.02

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
b0_Intercept	2.53	0.17	2.19	2.86	1554	1.00
b1_Intercept	1.56	0.13	1.33	1.81	1566	1.01
b2_Intercept	0.88	0.06	0.77	1.00	314	1.03
b3_Intercept	0.97	0.05	0.88	1.07	91	1.04
b3_precip	0.01	0.01	0.00	0.03	3000	1.00

b3_gdd	0.01	0.00	0.00	0.02	3000	1.01
b3_precip:gdd	0.00	0.00	0.00	0.01	48	1.07

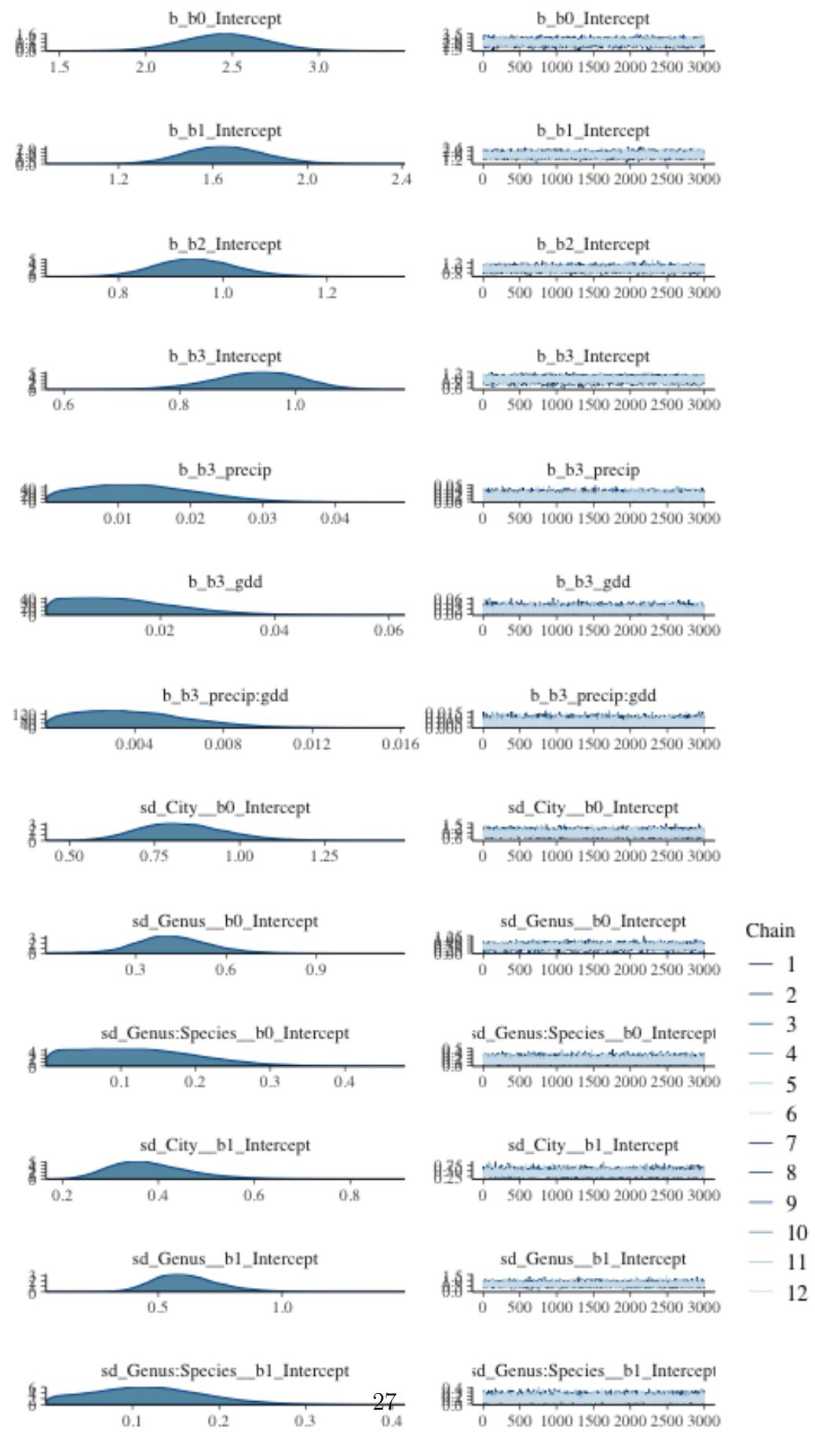
Family Specific Parameters:

	Estimate	Est.Error	1-95% CI	u-95% CI	Eff.Sample	Rhat
shape	14.87	0.30	14.29	15.47	3000	1.00

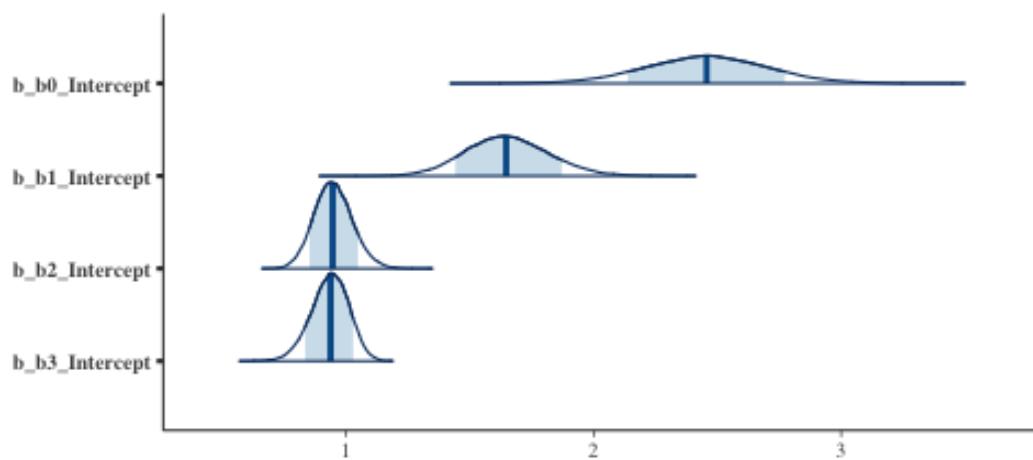
Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample is a crude measure of effective sample size, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

Warning messages:

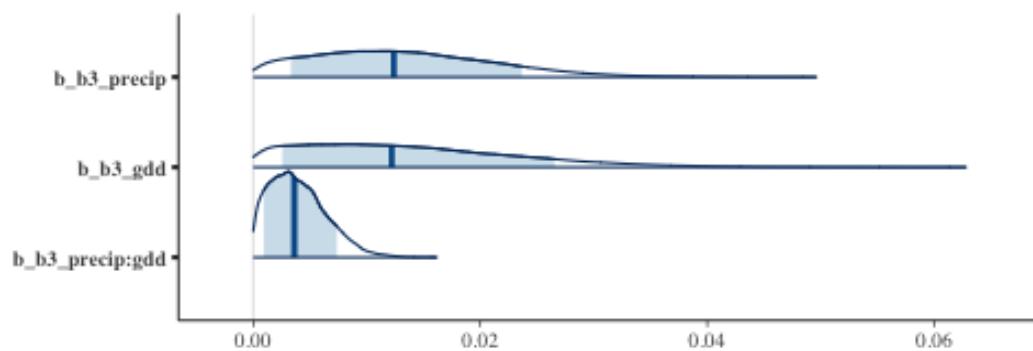
- 1: The model has not converged (some Rhats are > 1.1). Do not analyse the results!
We recommend running more iterations and/or setting stronger priors.
- 2: There were 4 divergent transitions after warmup. Increasing adapt_delta above 0. See <http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup>



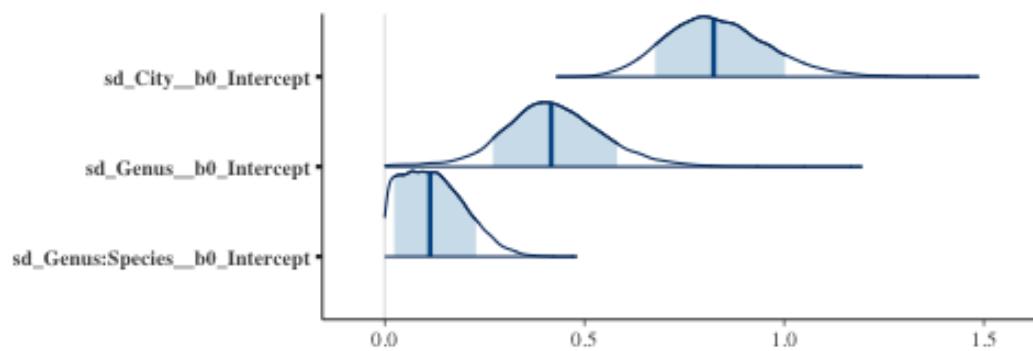
Posterior distributions
with medians and 80% intervals

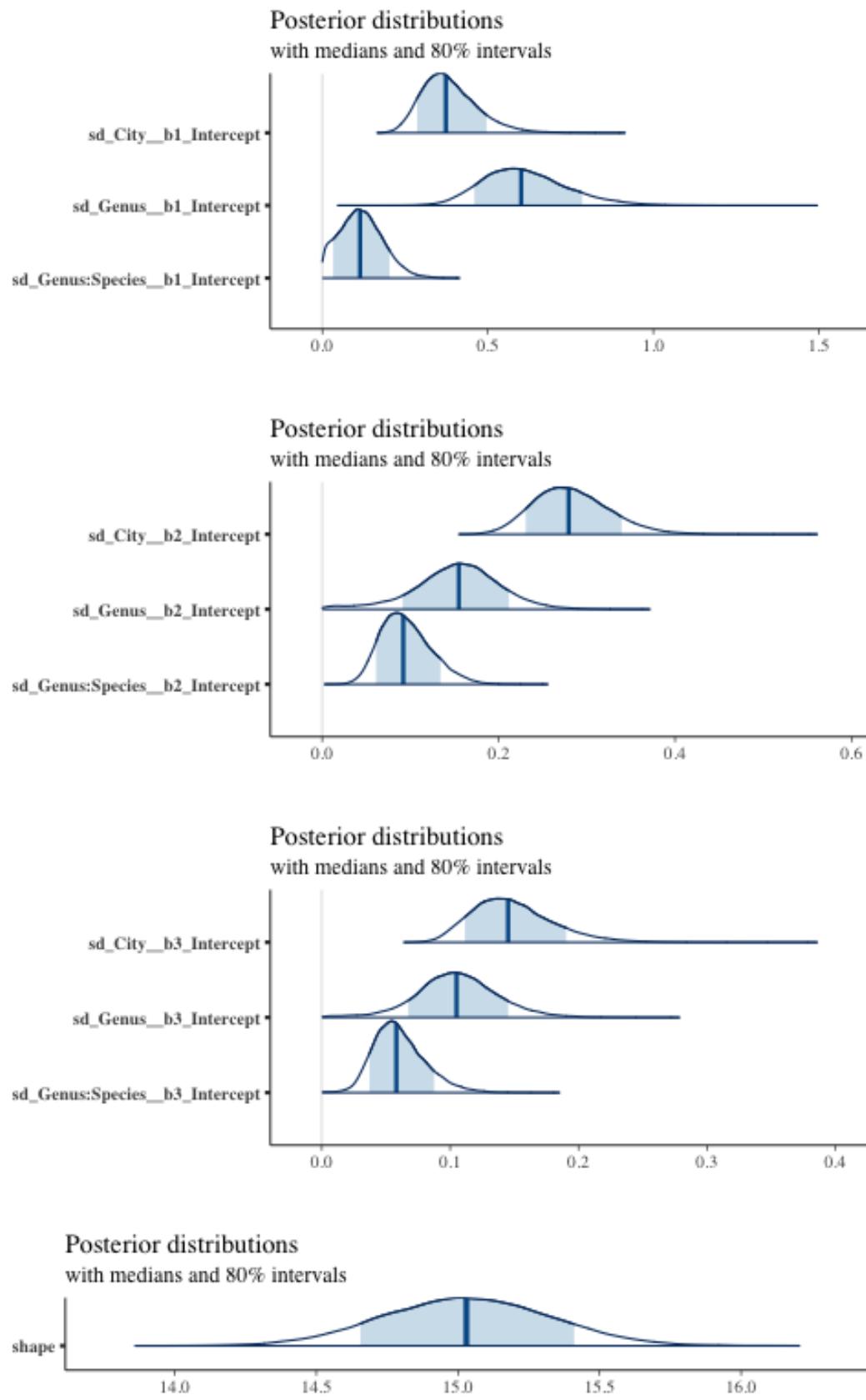


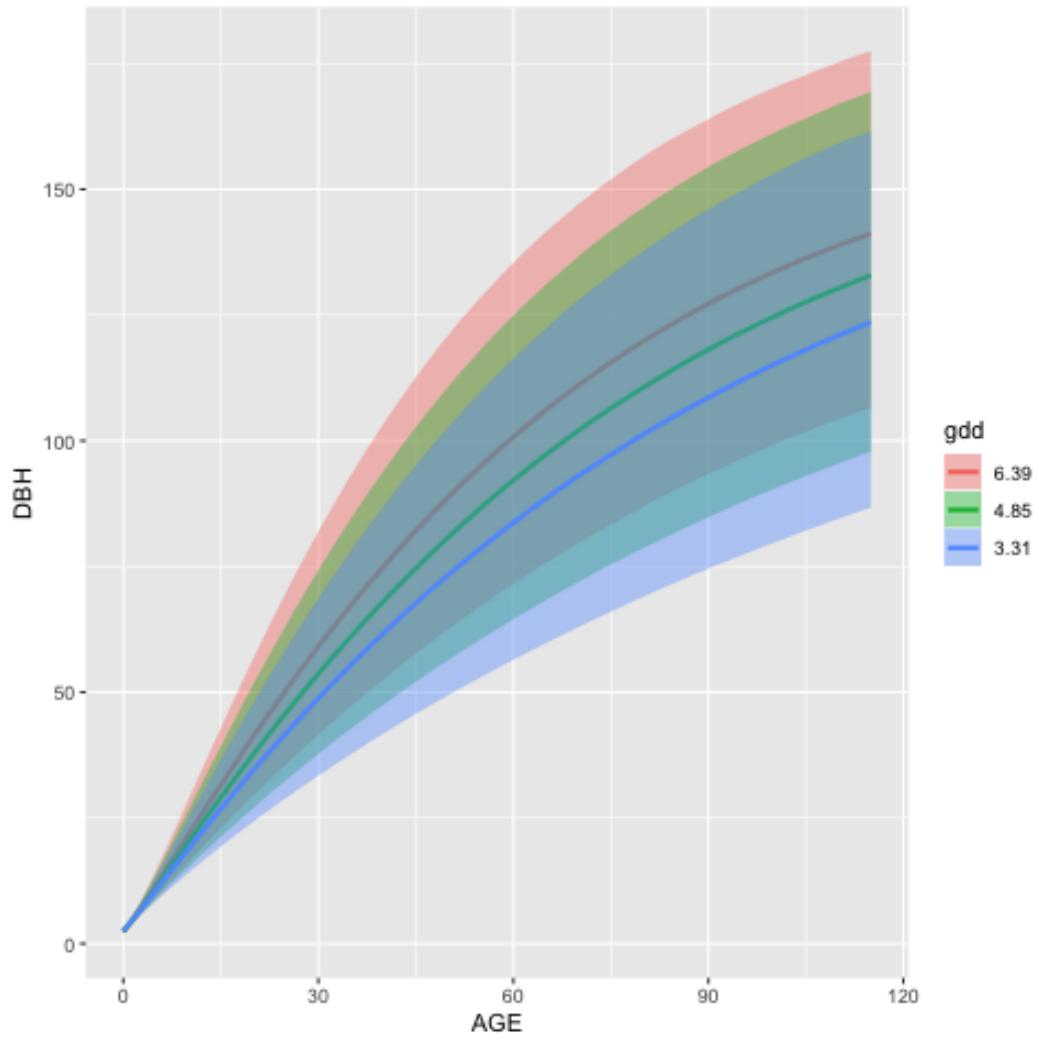
Posterior distributions
with medians and 80% intervals

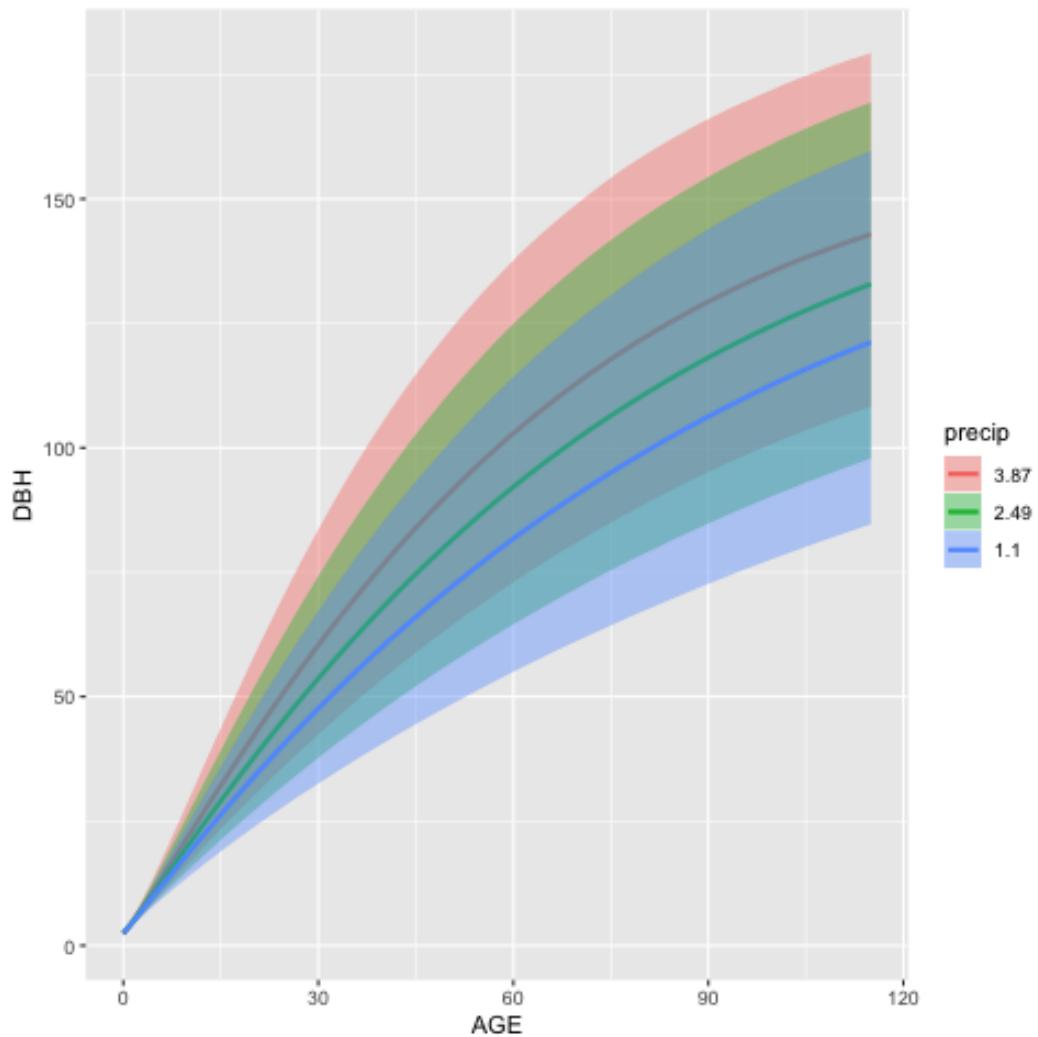


Posterior distributions
with medians and 80% intervals

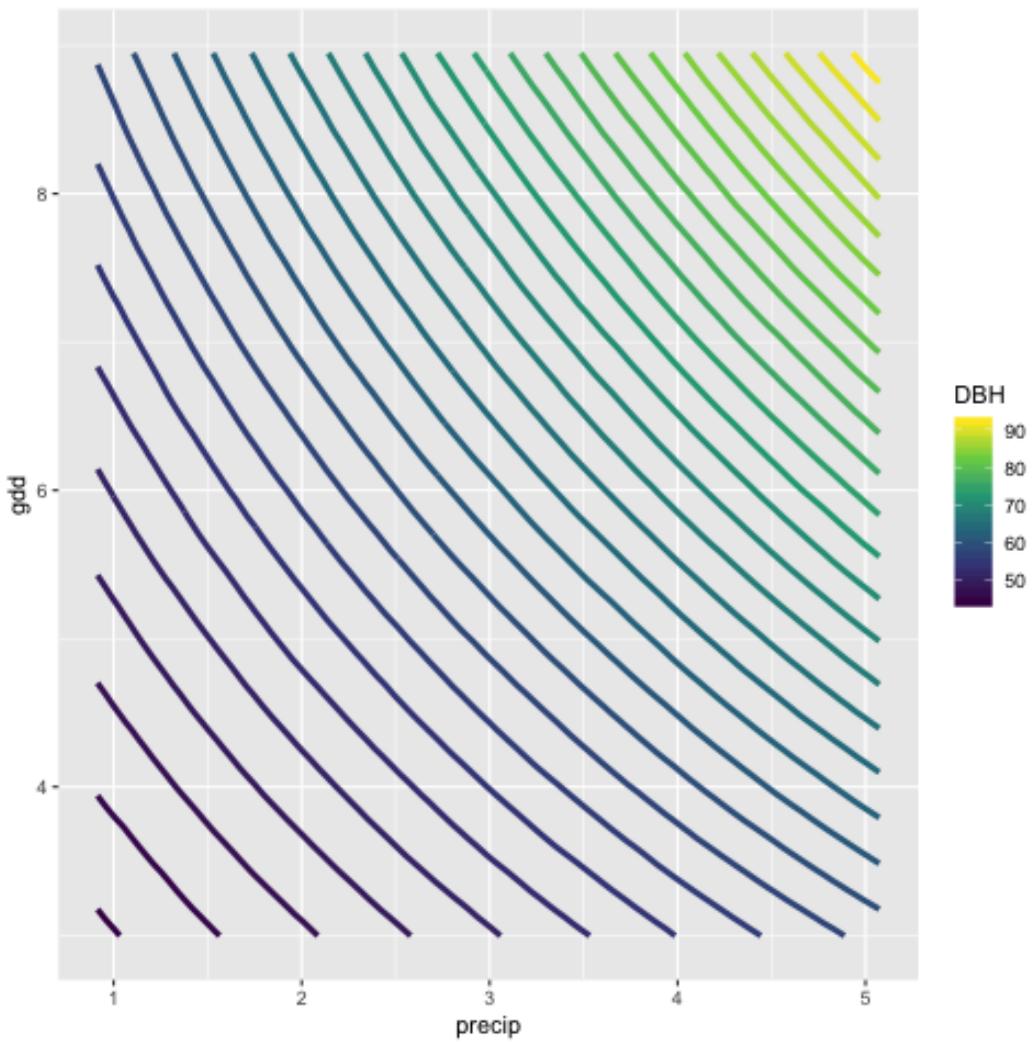


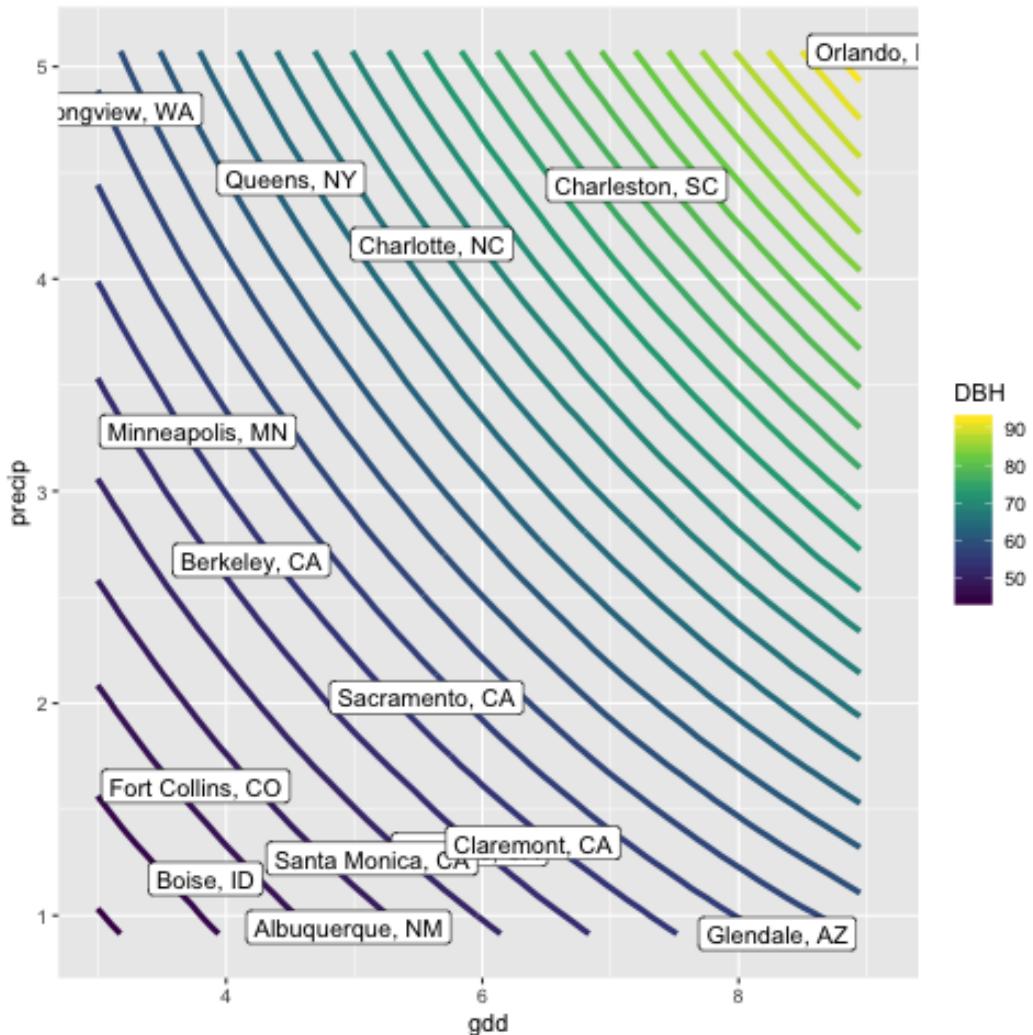






```
precip.gdd <- marginal_effects(mod, effects = "precip:gdd", surface = T)
```





```
to.include <- c("LIST", "PLAC", "FRPE")
```

```
cond <- expand.grid(Species = unique(mod$data$Species), City = unique(mod$data$City))
cond <- left_join(cond, unique(select(mod$data, Species, Genus)))
cond <- left_join(cond, unique(select(mod$data, City, precip, gdd)))
cond <- filter(cond, Species %in% to.include)

me <- marginal_effects(mod, effects = "AGE", conditions = cond, re_formula = NULL)
```

Joining, by = "Species"

Warning message:

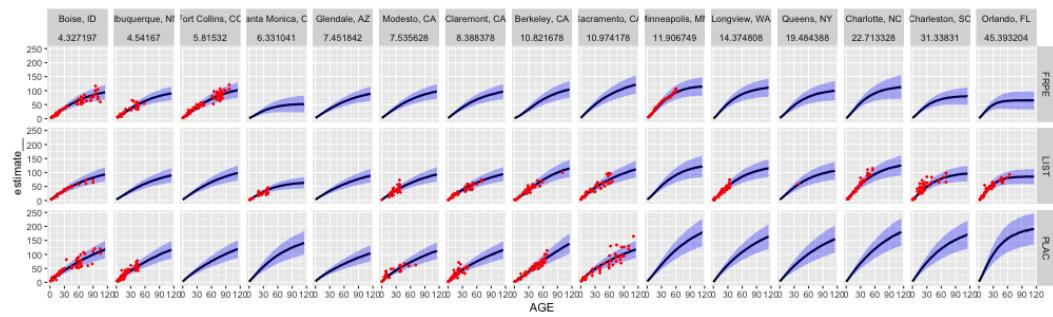
Column ‘Species’ joining factor and character vector, coercing into character vector

Joining, by = "City"

Warning message:

Column ‘City’ joining factor and character vector, coercing into character vector

There were 50 or more warnings (use warnings() to see the first 50)



```
library(brms)
library(dplyr)

mod <- readRDS("../models/genus_many_species_many_cities_many_climate_b3linint_hetero.RDS")

cond <- expand.grid(Species = unique(mod$data$Species), City = unique(mod$data$City))

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 = NA)
saveRDS(me, "../models/genus_many_species_many_cities_many_climate_b3linint_hetero.RDS")

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

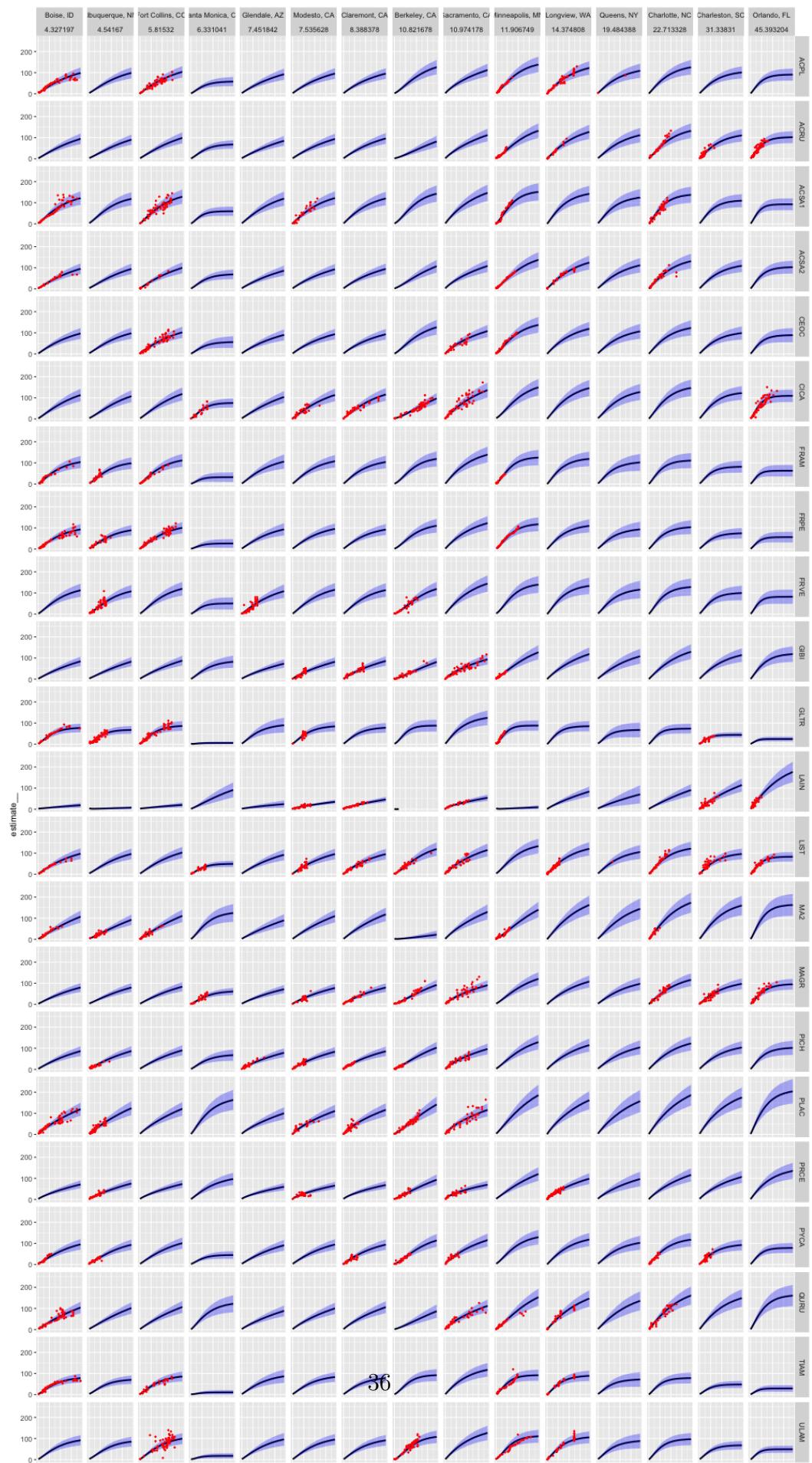


```
ssh krusty
cd allo/code
nohup R CMD BATCH marginal_effects_fullconditions.R &
```

```
cat marginal_effects_fullconditions.Rout  
exit
```

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

```
me <-readRDS("~/git/allo/models/genus_many_species_many_cities_many_climate_b3linin
```



- 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_testing_noWAR0.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 existing
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

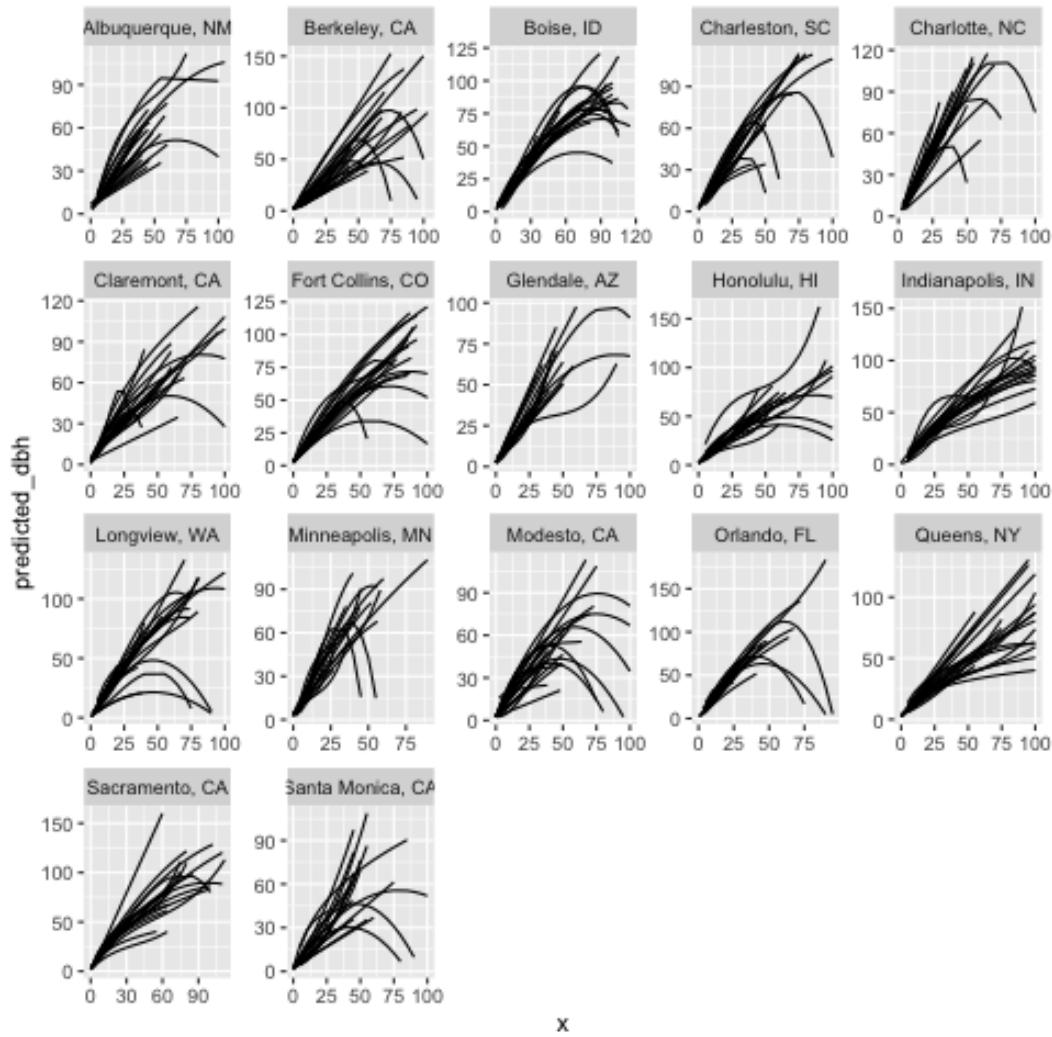
```

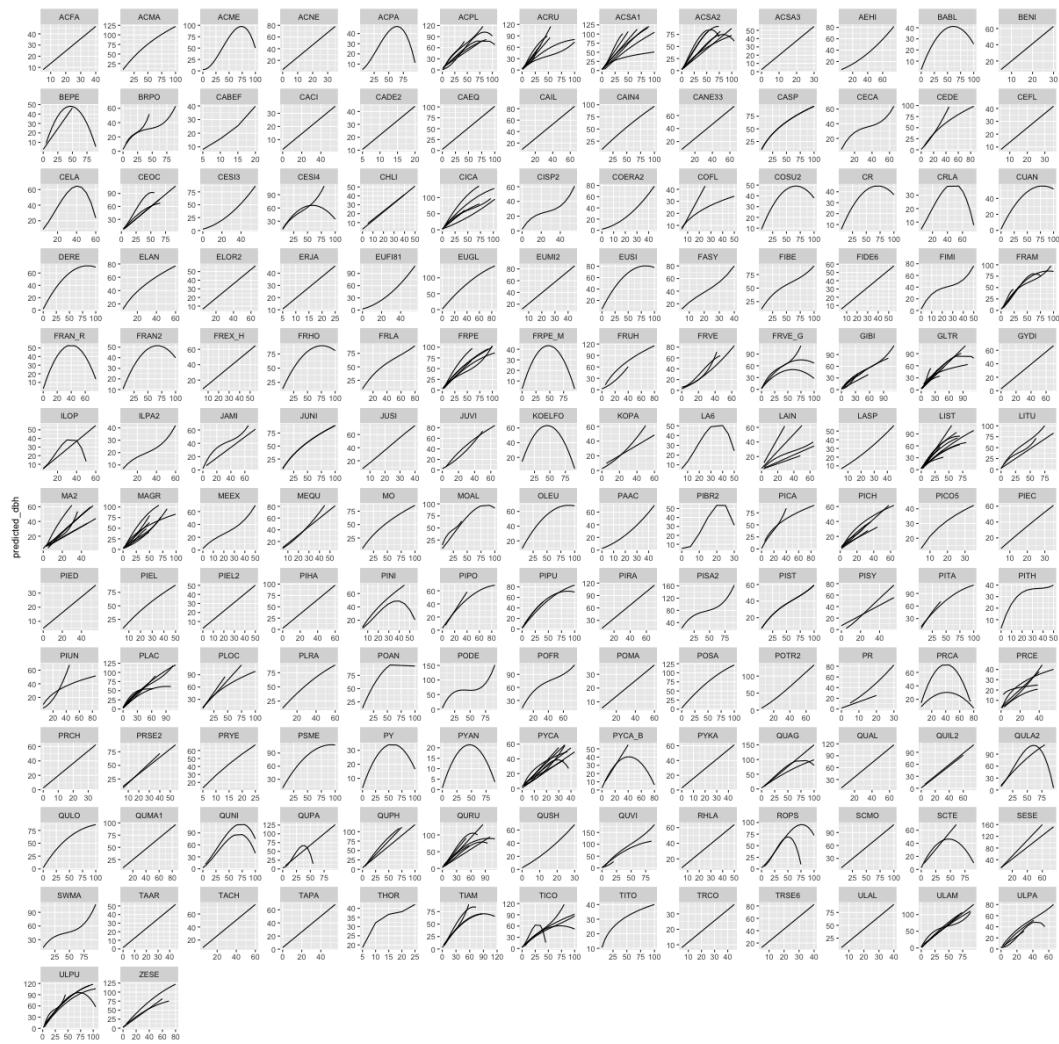
```
predictions_apprange <- predictions %>%
  filter(predicted_dbh > Apps.min & predicted_dbh < Apps.max)

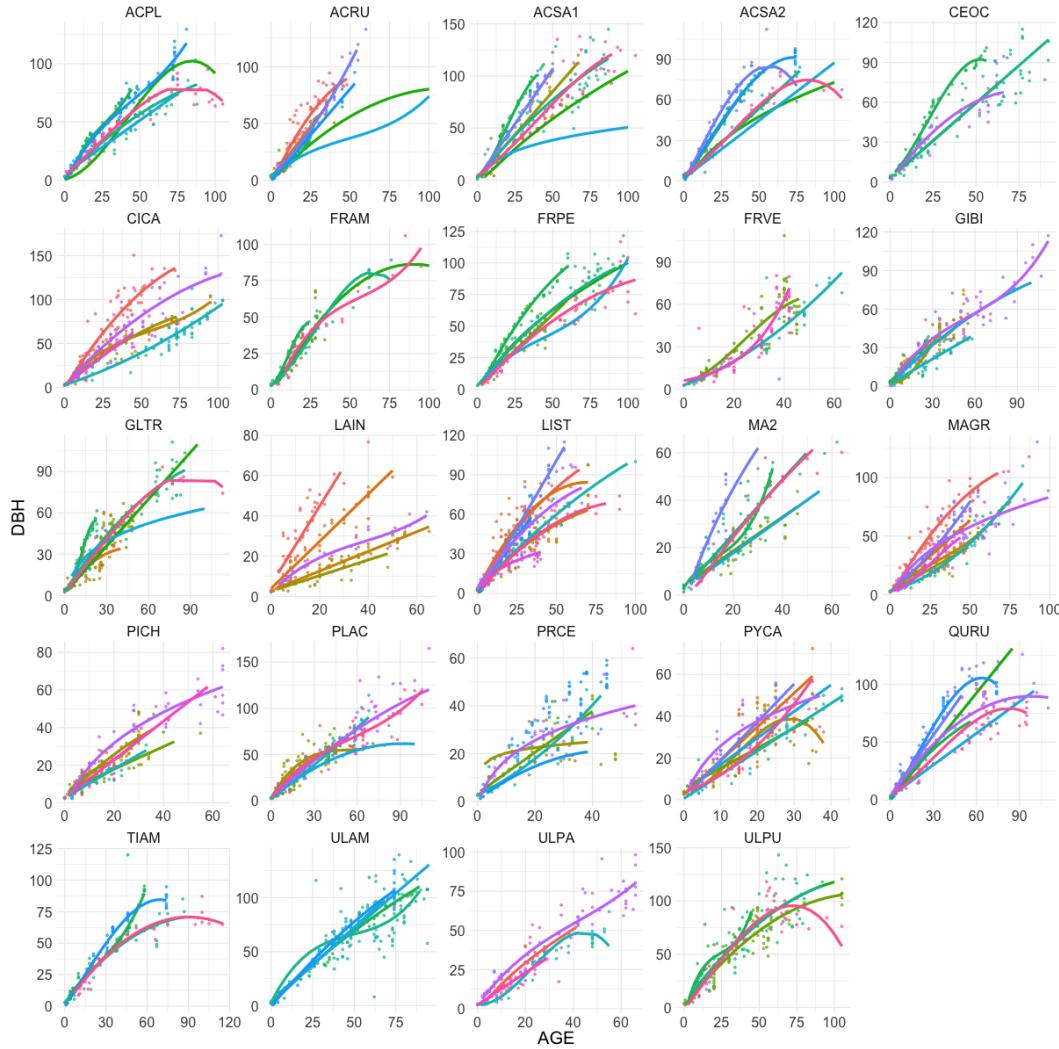
predictions_datarange <- predictions %>%
  filter(predicted_dbh > minDBH & predicted_dbh < maxDBH)

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)

blue <- "blue"

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=TRUE))

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:

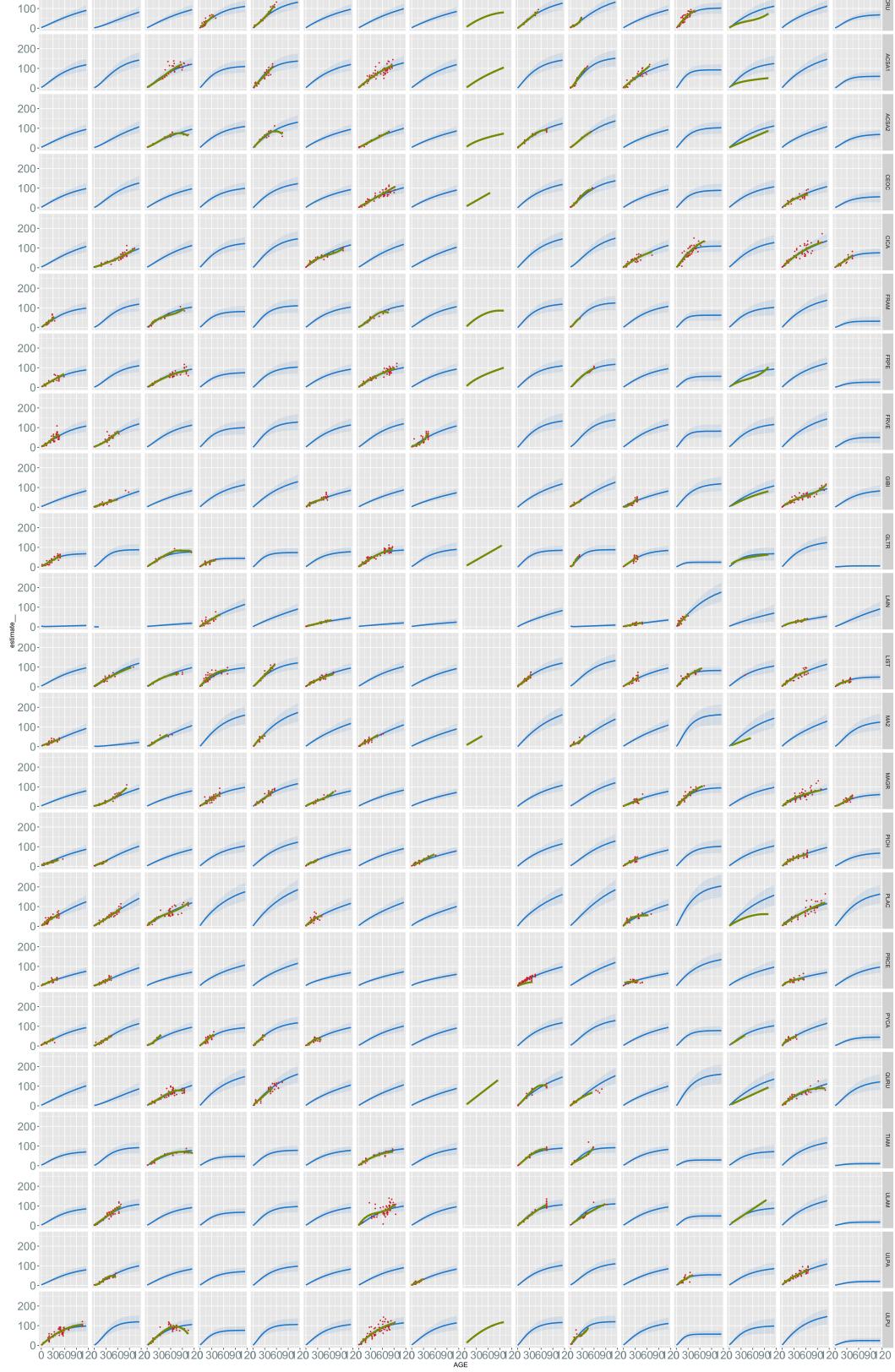
Column ‘City’ joining factor and character vector, coercing into character vector

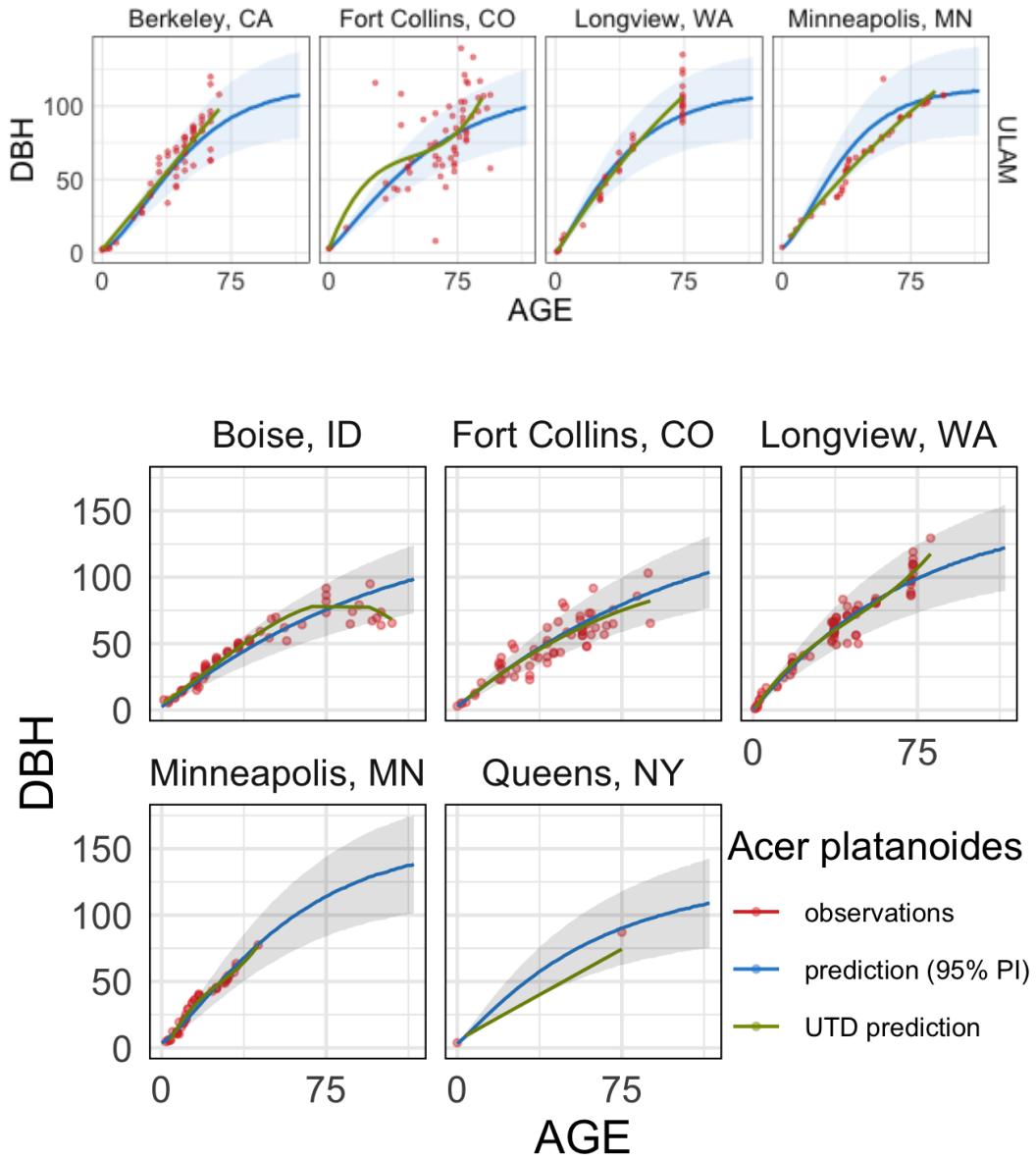
Joining, by = "City"

Warning message:

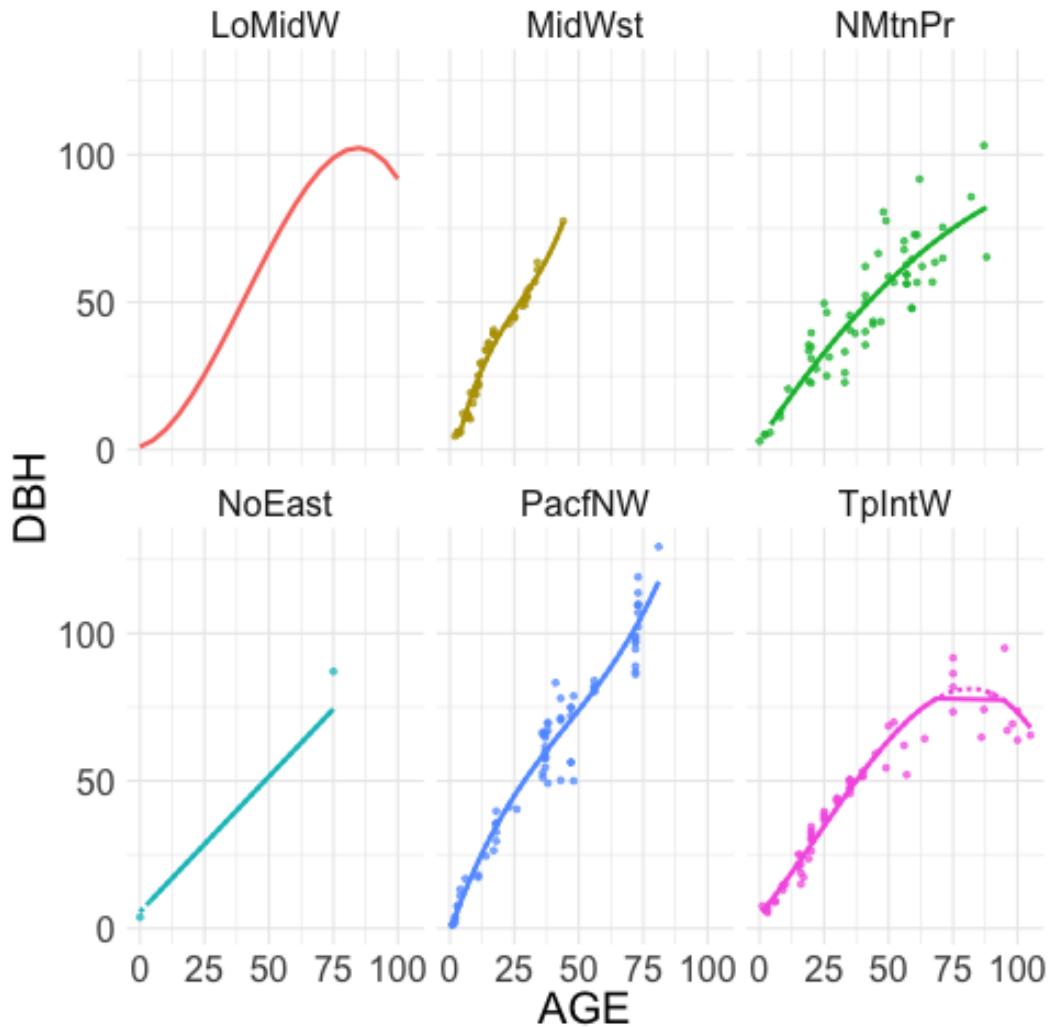
Column ‘City’ joining factor and character vector, coercing into character vector



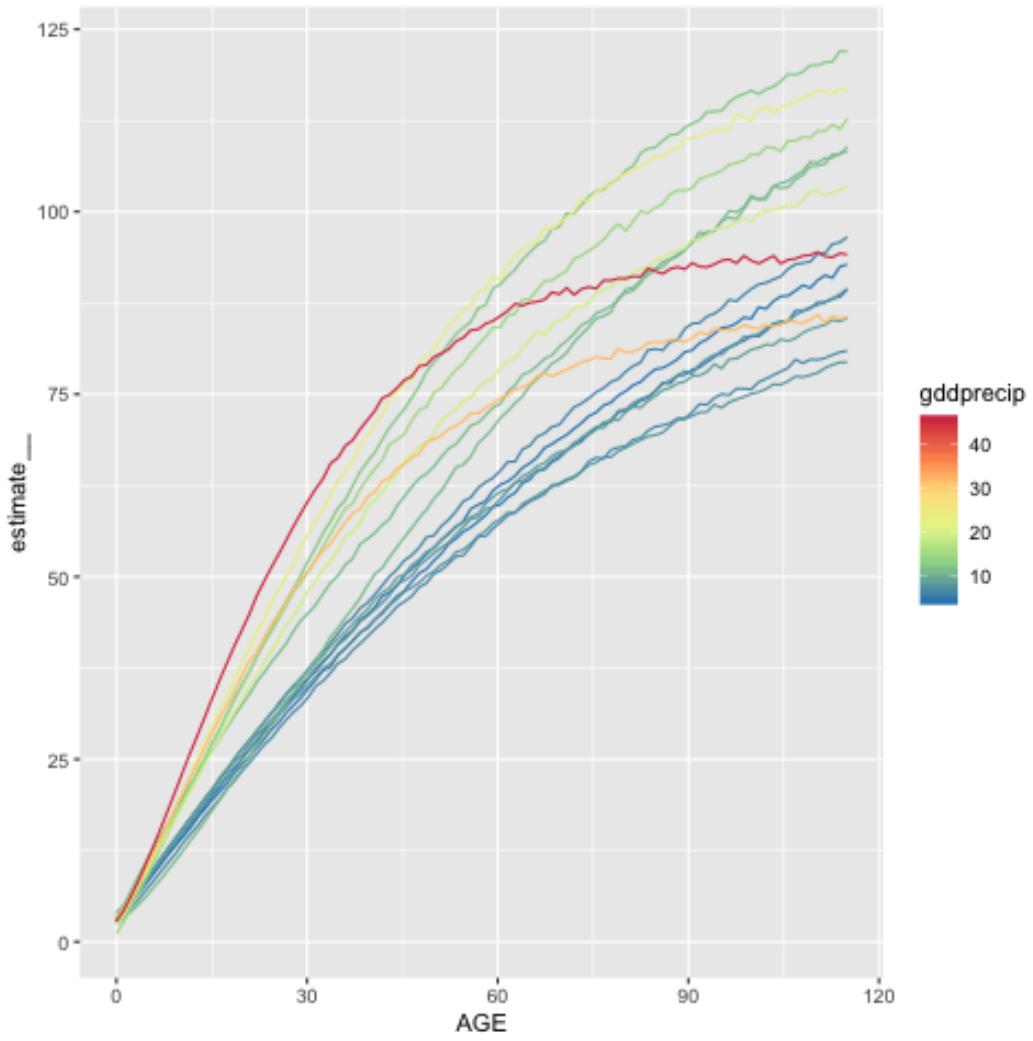


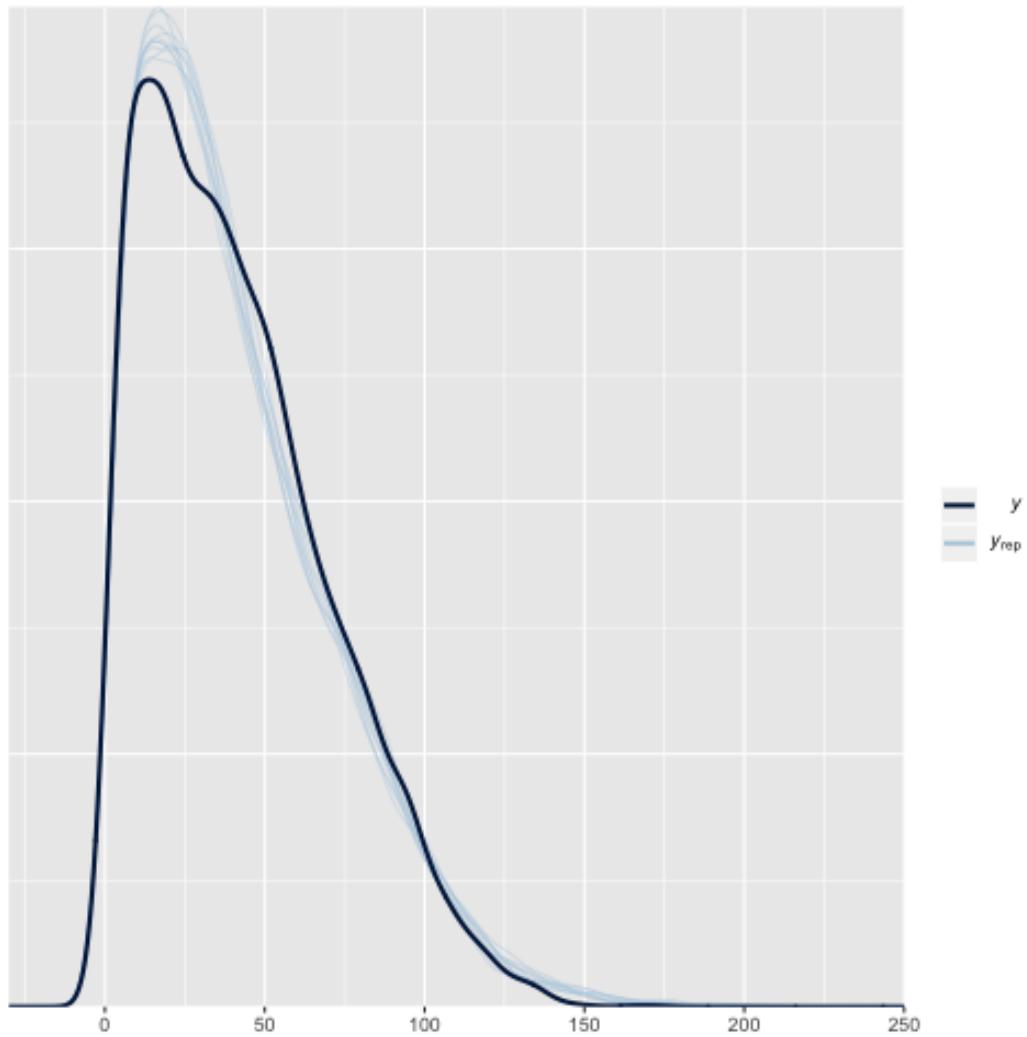


What city is the lower midwest? Indianapolis. There is no data for that.



Funny thing about the Apps range is that for functions that eventually decrease, they don't properly stop inference. Look at the pink line above. The application range should be given for both the predictor and the response variable.





- 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)

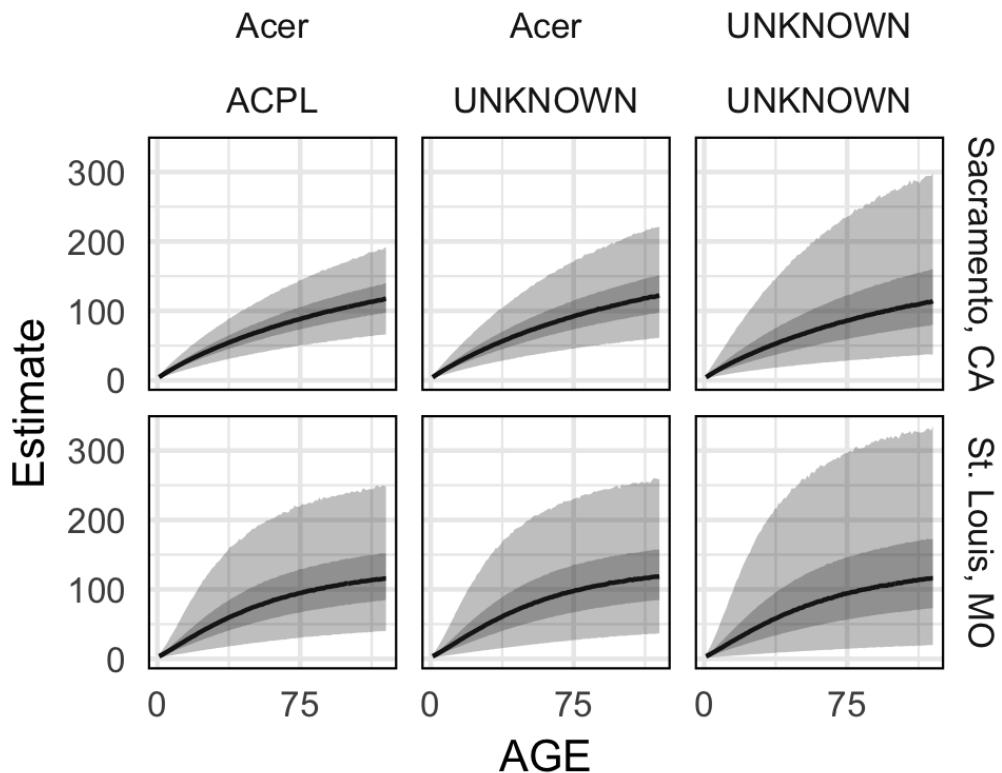
expand.grid.df <- function(...) Reduce(function(...) merge(..., by=NULL), list(...))

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

pred.nd <- predict(mod, nd, allow_new_levels = T, robust = T, probs = c(0.025, 0.975))

```

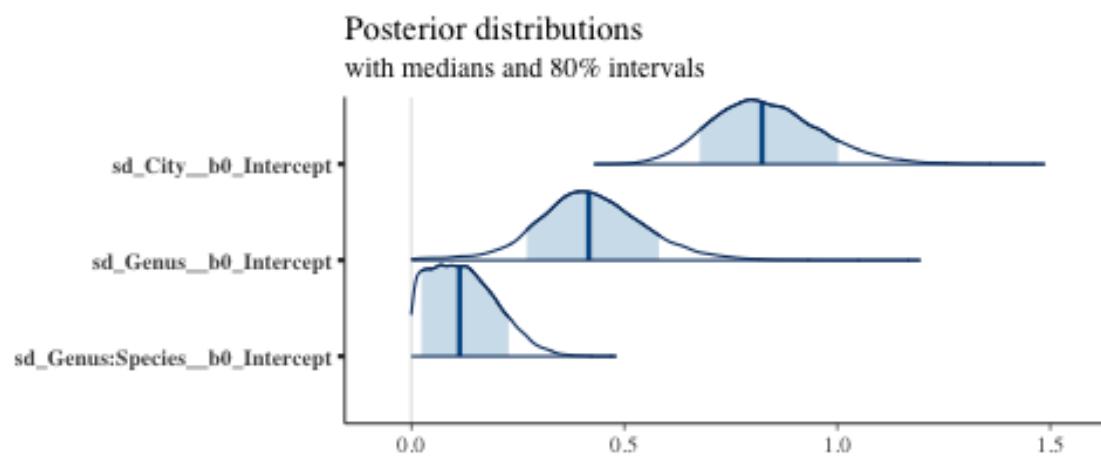
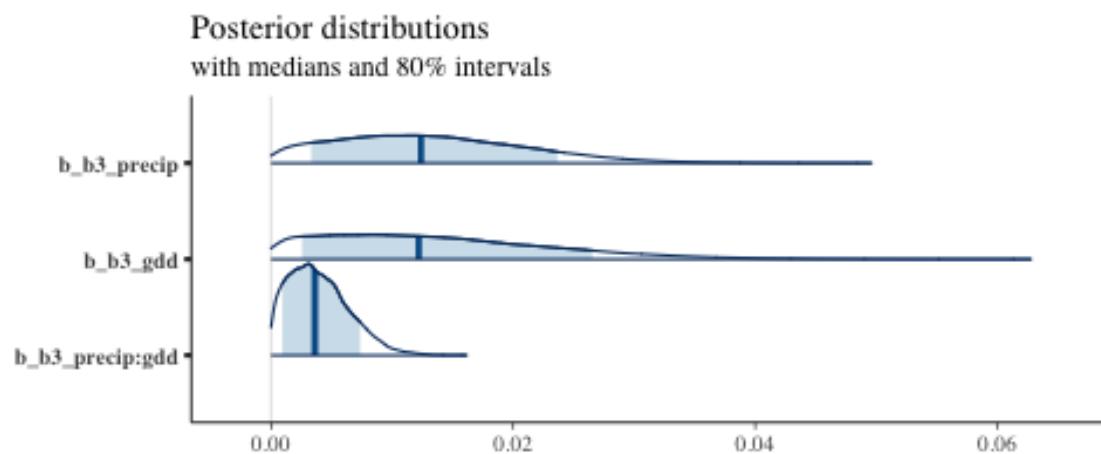
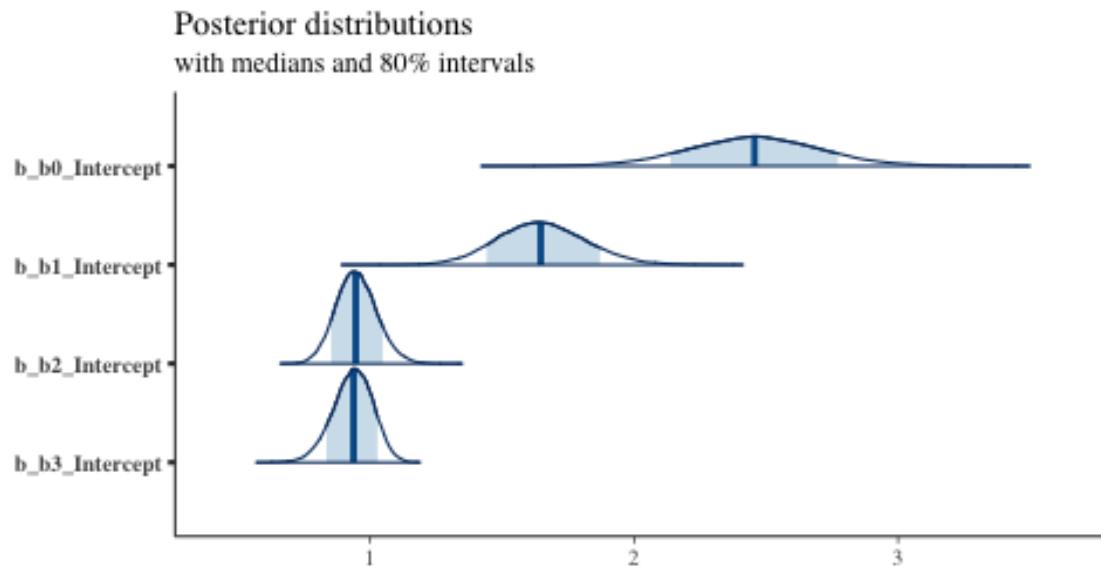
```
pd <- cbind(nd, pred.nd)
```



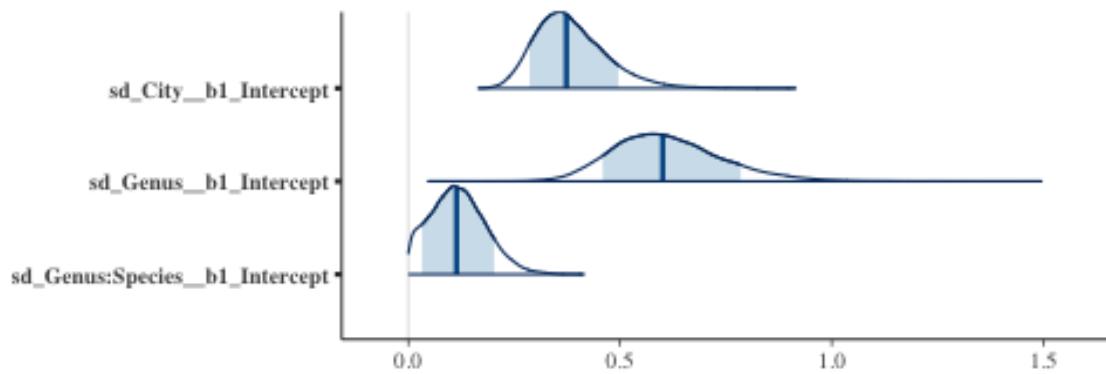
Results

parameter estimates

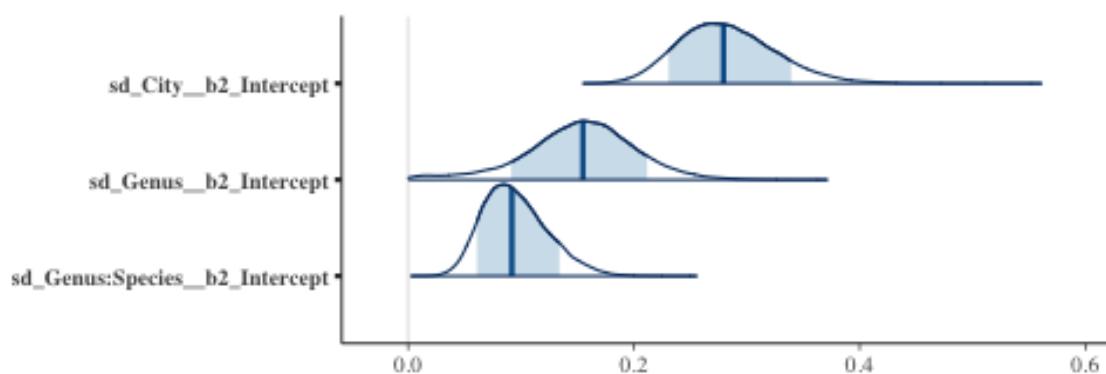
Posterior distributions with 80% interval and median for parameters in



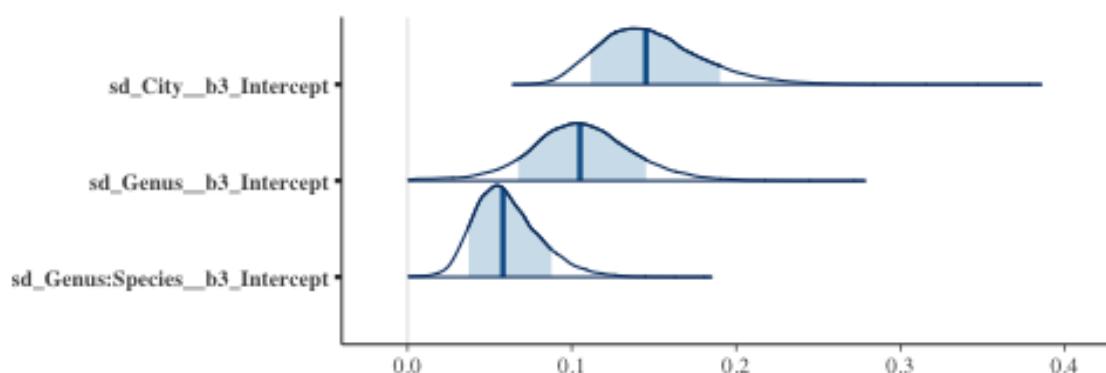
Posterior distributions
with medians and 80% intervals



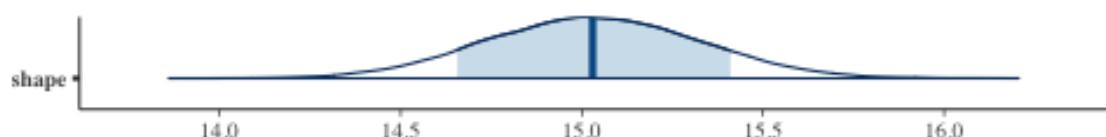
Posterior distributions
with medians and 80% intervals



Posterior distributions
with medians and 80% intervals

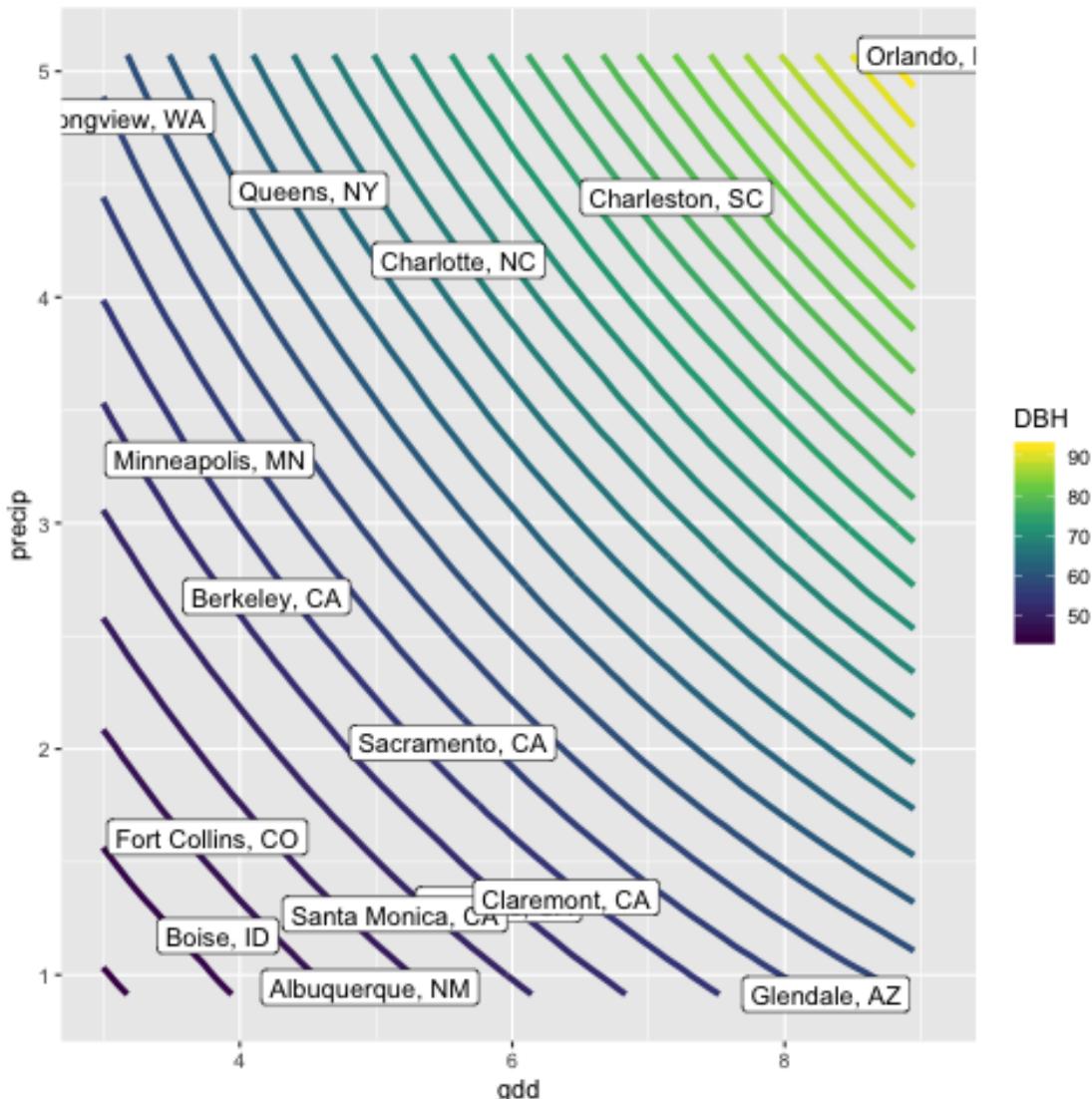


Posterior distributions
with medians and 80% intervals



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.



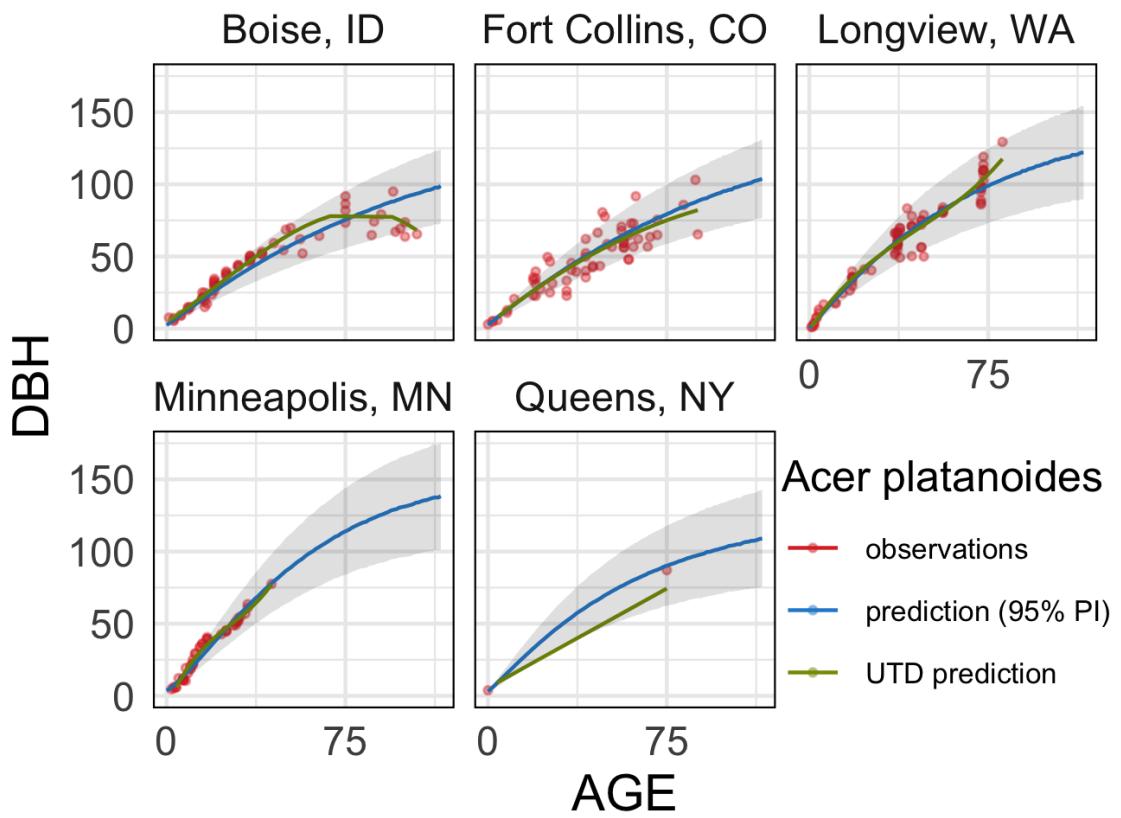


Figure 4: red points are observations, blue lines are predictions of the model and shading indicates 95% prediction interval, green lines show UTD equations

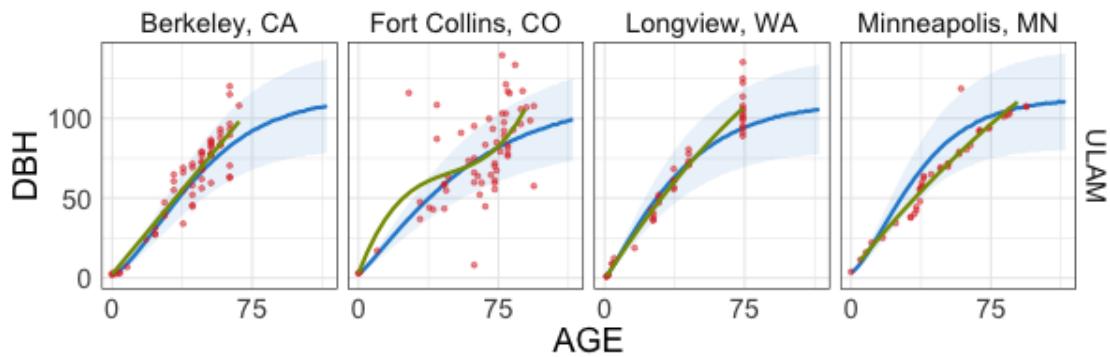
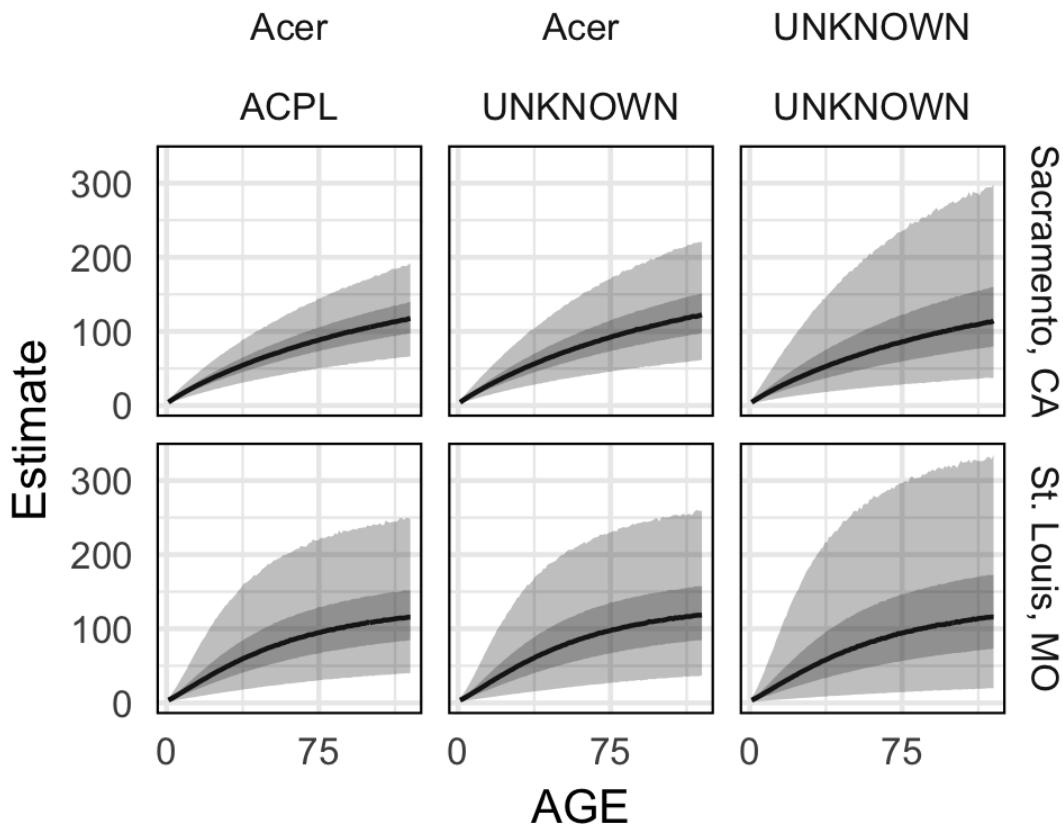


Figure 5: red points are observations, blue lines are predictions of the model and shading indicates 95% prediction interval, green lines show UTD equations

Comparing to existing equations

illustrative comparisons

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



Discussions

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. (2001) - 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. (2001) 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.
- 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?

misc

Meeting with Jun *[2018-08-27 Mon]*

What should the structure of the paper be?

- Premise of paper: showing a new method to fit multilevel non-linear growth models for trees
- Our working example is the urban tree
- Show different types of posterior predictions (e.g. new tree in a observed city of an observed species, versus new tree in an unobserved city of an unobserved species, but known climate and genus)
- Show predictions compared to the existing method.
 - will show better potential to extrapolate, both to unobserved ages and unobserved species/city combinations.
- Model selection
- posterior predictive checks

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.
- 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., McPherson, E. G., and Mori, S. M. (2001). Equations for predicting diameter, height, crown width, and leaf area of san joaquin valley street trees. *Journal of Arboriculture*, 27(6):306–317.
- 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.
- Weiskittel, A. R., Hann, D. W., Kershaw, J. A., and Vanclay, J. K. (2011). *Forest Growth and Yield Modeling*.]. John Wiley & Sons, Ltd.

Yang, Y., Huang, S., Meng, S. X., Trincado, G., and VanderSchaaf, C. L. (2009). A multilevel individual tree basal area increment model for aspen in boreal mixedwood stands. *Canadian Journal of Forest Research*, 39(11):2203–2214.