Ch. 1 - Model comaprisons

Wade VanderWright

2022-02-25

Contents

Model Comparisons

Model 1 - CFAR ~ GSH

1. brms model

```
Model_plain <- brm(Log10CFAR ~ Log10MeanGSH_centered,
                   data = GSH_df_pgls,
                   family = gaussian(),
                   prior = Plain_prior
                   )
```

```
fixef(Model_plain, summary = T)
```

```
##                     Estimate  Est.Error        Q2.5       Q97.5
## Intercept         0.01878724 0.01121401 -0.003048758 0.04068312
## Log10MeanGSH_centered 0.78943177 0.06909079  0.653733169 0.92180824
```
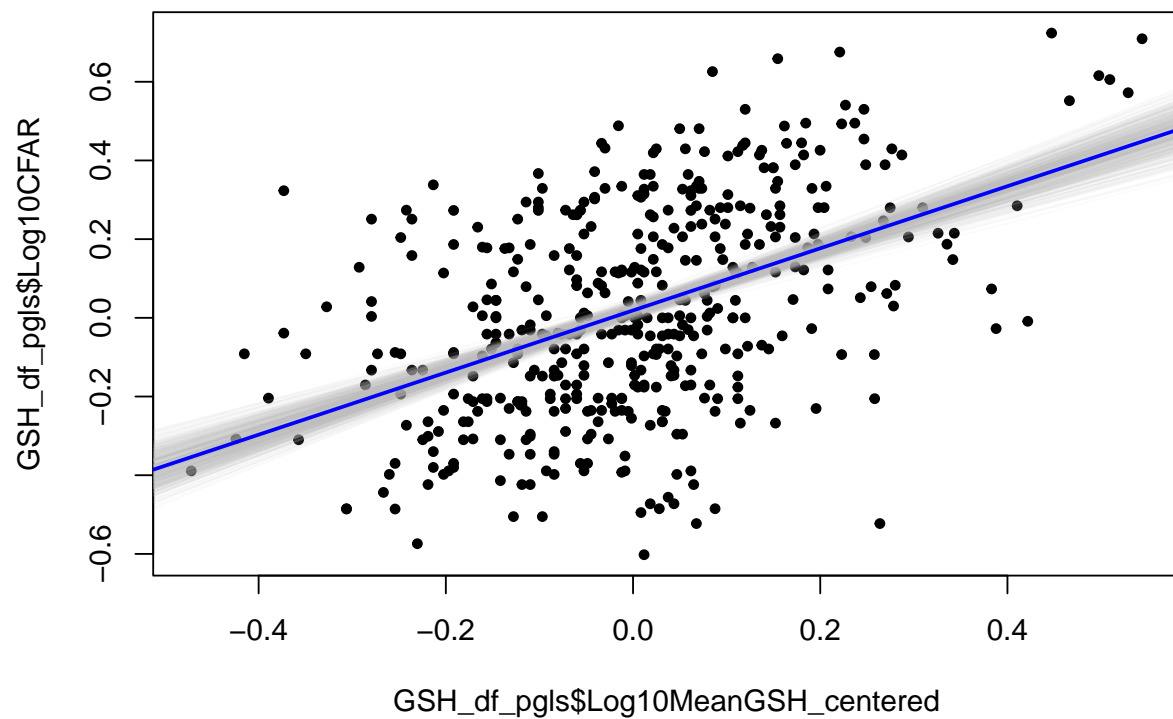
2. STAN model

Data considered

```
data {
  int <lower=1> N;
  vector[N] x;
  vector[N] y;
}
```

The parameters accepted by the model

```
parameters {
  real alpha;
  real beta;
  real<lower=0> sigma;
}
```

Model form:

```
model {

  sigma ~ student_t(3, 0, 10);
```

```
  y ~ normal(alpha + x * beta , sigma);
}
```

```
fit1_summary <- summary(fit1, pars = c("alpha", "beta","sigma"), probs = c(0.05, 0.95))$summary
print(fit1_summary)
```

```
##                 mean       se_mean          sd          5%         95%     n_eff
## alpha 0.01900869 0.0002460092 0.011114187 0.001013608 0.03747443 2041.046
## beta  0.78894370 0.0017257508 0.071948028 0.669736389 0.90516905 1738.128
## sigma 0.23579844 0.0001758531 0.007734538 0.223108150 0.24878671 1934.501
##            Rhat
## alpha 1.0004828
## beta  0.9988377
## sigma 0.9998605
```
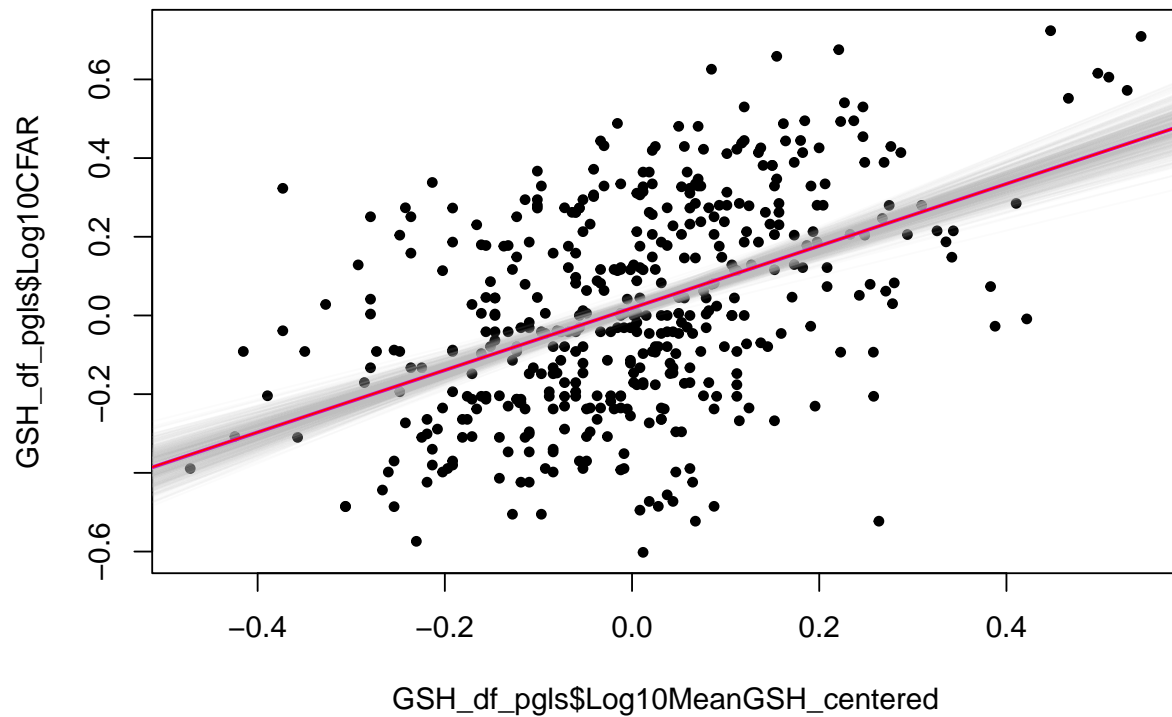
Model 2 - CFAR ~ GSH * Max Size

1. brms model

```
Model_size <- brm(Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered,
                  data = GSH_df_pgls,
                  family = gaussian(),
                  prior = Size_prior
                  )
```

```
fixef(Model_size, summary = T)
```

```
##                                                 Estimate  Est.Error          Q2.5
## Intercept                                     0.01107801 0.01158262 -0.01069365
## Log10MeanGSH_centered                         0.38761087 0.08262037  0.22932179
## Log10MaxSize_centered                         0.30460627 0.04281846  0.22222482
## Log10MeanGSH_centered:Log10MaxSize_centered   0.26628115 0.16406086 -0.05284798
##                                                   Q97.5
## Intercept                                     0.03436049
## Log10MeanGSH_centered                         0.55212619
## Log10MaxSize_centered                         0.38716498
## Log10MeanGSH_centered:Log10MaxSize_centered   0.58368860
```


2. STAN model

4

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // CFAR
}
```

The parameters accepted by the model.

```
parameters {
  real alpha;
  vector[K] beta;
  real<lower=0> sigma;
}
```
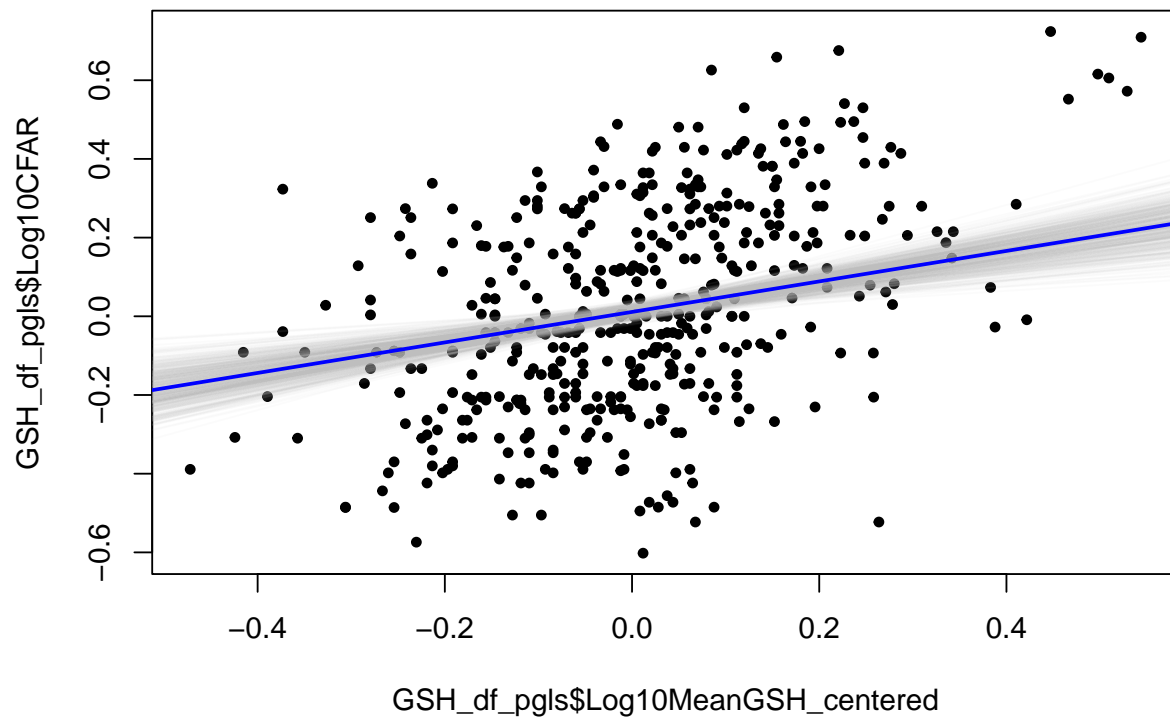
The model to be estimated. We model the output 'y' to be normally distributed with mean 'mu' and standard deviation 'sigma'.

```
model {

  sigma ~ student_t(3, 0, 10);

  y ~ normal(alpha + x * beta , sigma);
}
```

```
fit2_summary <- summary(fit2, pars = c("beta[1]", "beta[2]","beta[3]","beta[4]","sigma"), probs = c(0.0
print(fit2_summary)
```

```
##              mean      se_mean          sd          5%        95%    n_eff
## beta[1] 0.01104707 0.0002537778 0.011373020 -0.007366795 0.02975656 2008.373
## beta[2] 0.38772105 0.0019673695 0.082329451  0.254362486 0.52129402 1751.211
## beta[3] 0.30513071 0.0010166417 0.043336491  0.231788976 0.37746503 1817.070
## beta[4] 0.25558541 0.0037875546 0.158011884 -0.003275980 0.51150818 1740.451
## sigma   0.22114731 0.0001653446 0.007578203  0.209154790 0.23385828 2100.645
##              Rhat
## beta[1] 0.9984008
## beta[2] 0.9999482
## beta[3] 1.0004711
## beta[4] 0.9990472
## sigma   0.9996704
```
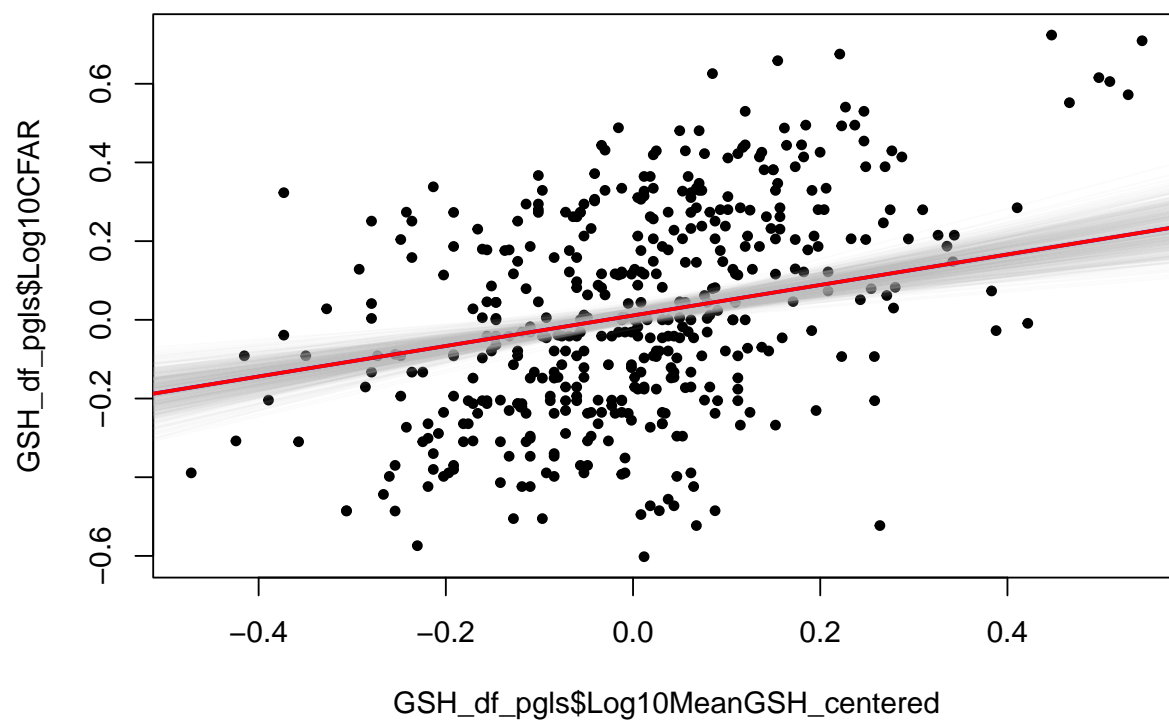
Model 3 - CFAR ~ GSH + Max Size (No interaction)

1. brms model
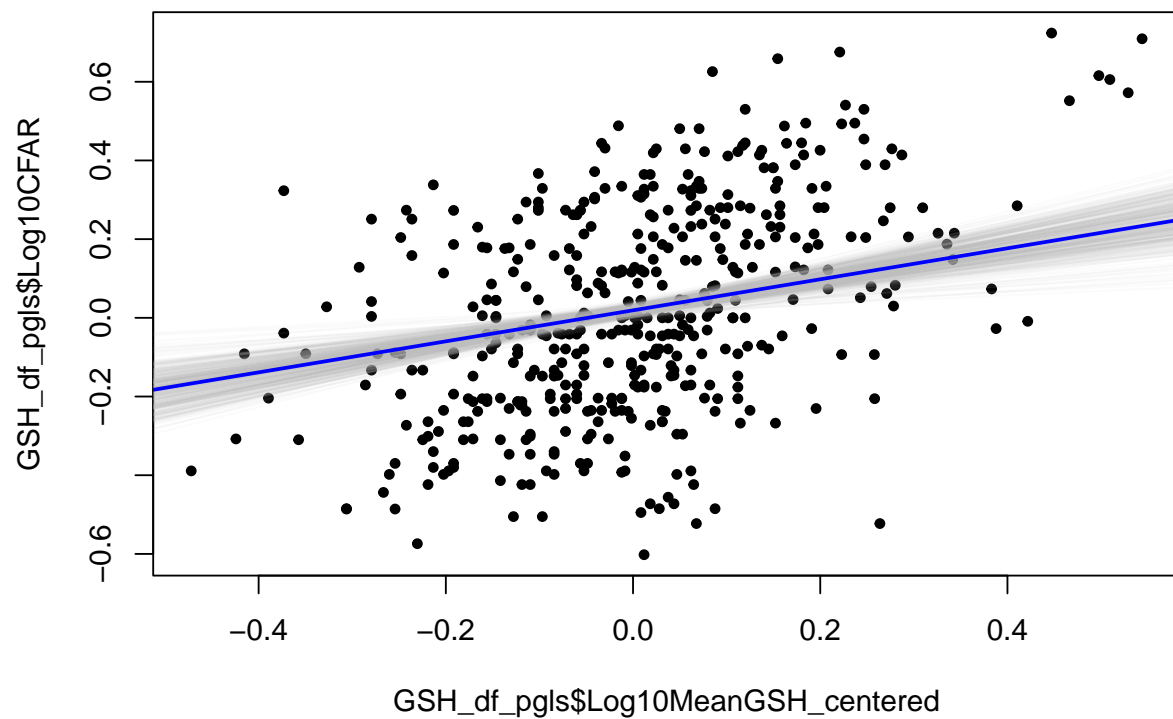
```
Model_size_2 <- brm(Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered,
                    data = GSH_df_pgls,
                    family = gaussian(),
                    prior = Size_prior_2)
```

```
summary(Model_size_2)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered
##    Data: GSH_df_pgls (Number of observations: 456)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Population-Level Effects:
##                        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                  0.02      0.01    -0.00     0.04 1.00     3526
## Log10MeanGSH_centered      0.39      0.08     0.24     0.55 1.00     3567
## Log10MaxSize_centered      0.33      0.04     0.25     0.40 1.00     3528
##                        Tail_ESS
## Intercept                  2556
## Log10MeanGSH_centered      2768
## Log10MaxSize_centered      3043
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.22      0.01     0.21     0.24 1.00     3652     2981
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

2. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // CFAR
}
```

The parameters accepted by the model.

```
parameters {
  real alpha;
  vector[K] beta;
  real<lower=0> sigma;
}
```

The model to be estimated. We model the output 'y' to be normally distributed with mean 'mu' and standard deviation 'sigma'.

```
model {
```

```
  sigma ~ student_t(3, 0, 10);

  y ~ normal(alpha + x * beta , sigma);
}
```

```
fit3_summary <- summary(fit3, pars = c("beta[1]", "beta[2]","sigma"), probs = c(0.05, 0.95))$summary
print(fit3_summary)
```

```
##               mean      se_mean          sd        5%        95%      n_eff
## beta[1] 0.3933554 0.0020241407 0.081480341 0.2564044 0.5235385 1620.408
## beta[2] 0.3242394 0.0010396032 0.040923837 0.2588048 0.3894789 1549.592
## sigma   0.2214438 0.0001766472 0.007474578 0.2094221 0.2342549 1790.440
##              Rhat
## beta[1] 0.9990436
## beta[2] 1.0000786
## sigma   1.0004027
```
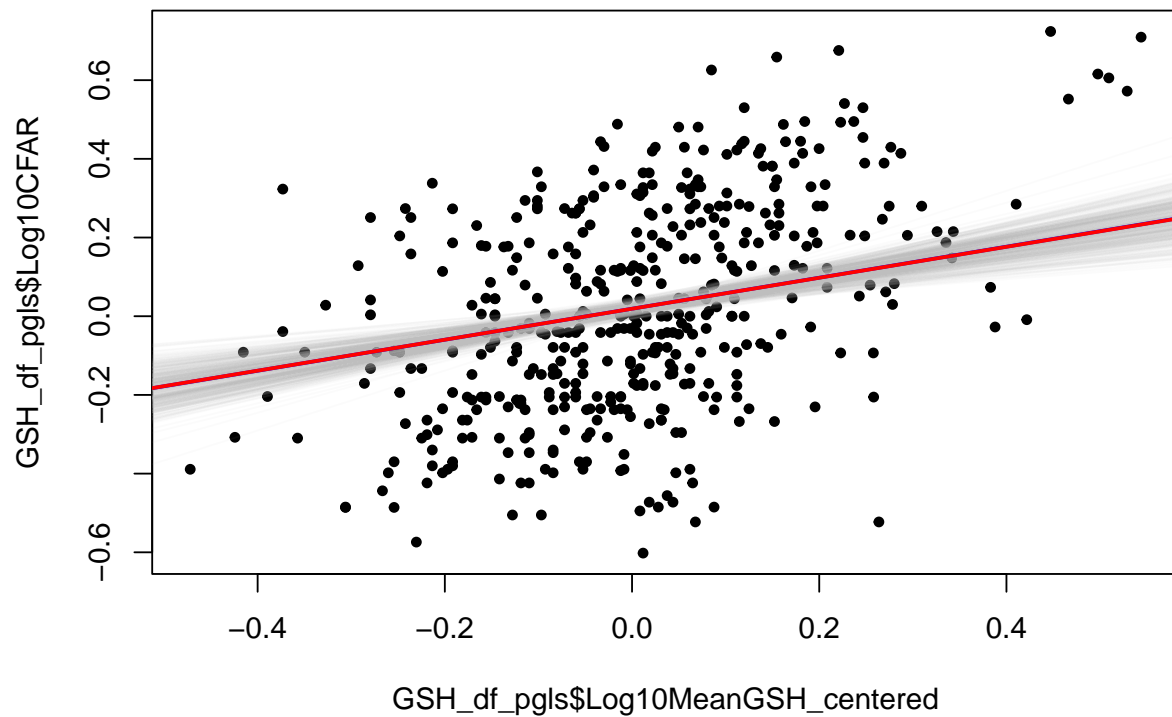
Model 4 - CFAR ~ GSH + PrimaryHabitat (No interaction)

1. brms model

```
habitat_model <- brm(Log10CFAR ~ Log10MeanGSH_centered + PrimaryHabitat,
                     data = GSH_df_pgls,
                     family = gaussian(),
                     prior = habitat_prior,
                     cores = 2)
```

```
summary(habitat_model)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + PrimaryHabitat
##    Data: GSH_df_pgls (Number of observations: 456)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Population-Level Effects:
##                         Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                   0.01      0.02    -0.02     0.04 1.00     4027
## Log10MeanGSH_centered       0.68      0.07     0.54     0.83 1.00     4122
## PrimaryHabitatDeepwater    -0.01      0.02    -0.06     0.04 1.00     3851
## PrimaryHabitatPelagic       0.25      0.05     0.16     0.34 1.00     4558
##                         Tail_ESS
## Intercept                   2871
## Log10MeanGSH_centered       2943
## PrimaryHabitatDeepwater     3151
## PrimaryHabitatPelagic       3159
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.23      0.01     0.21     0.25 1.00     3991     2793
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

2. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // CFAR
}
```

The parameters accepted by the model.

```
parameters {
  real alpha;
  vector[K] beta;
  real<lower=0> sigma;
}
```

The model to be estimated. We model the output 'y' to be normally distributed with mean 'mu' and standard deviation 'sigma'.
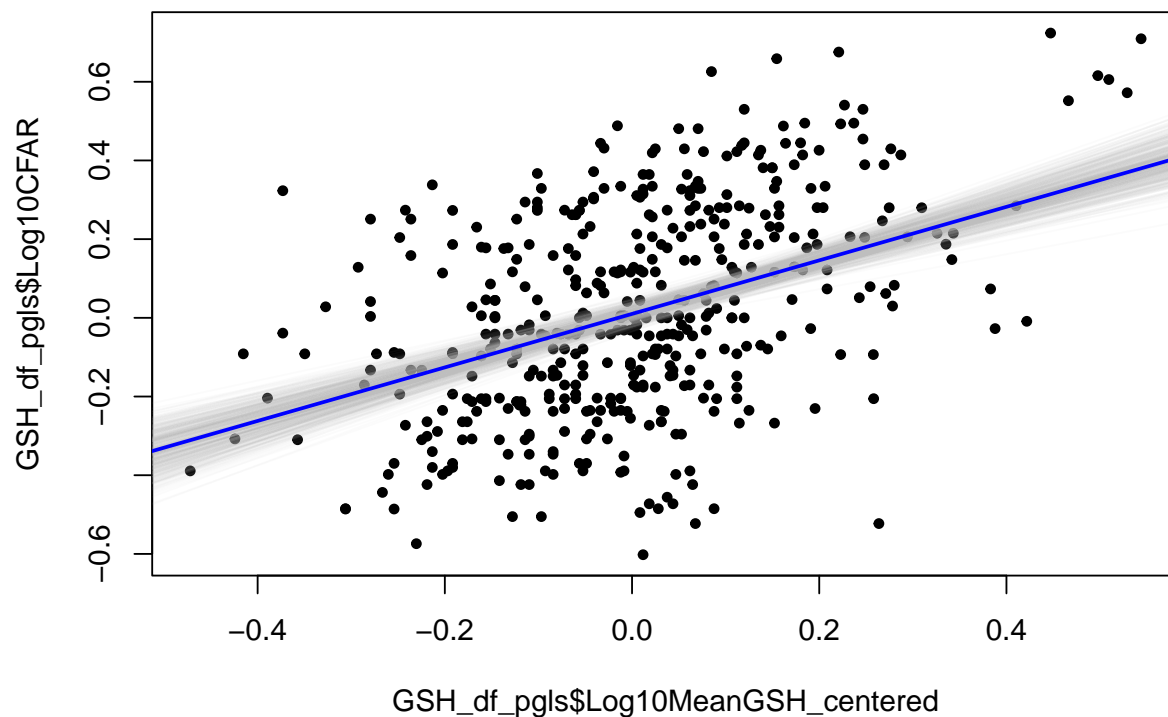
```
model {
```

```
  sigma ~ student_t(3, 0, 10);

  y ~ normal(alpha + x * beta , sigma);
}
```

```
fit6_summary <- summary(fit6, pars = c("beta[1]", "beta[2]","beta[3]","beta[4]", "sigma"), probs = c(0.0
print(fit6_summary)
```

```
##                   mean        se_mean          sd           5%          95%       n_eff
## beta[1]   0.009539473 0.0004751520 0.016255936 -0.01635248 0.03574710 1170.466
## beta[2]   0.682189383 0.0017675003 0.073049165   0.56027854 0.80302487 1708.093
## beta[3]  -0.009420391 0.0006962956 0.023470977 -0.04888522 0.02824612 1136.253
## beta[4]   0.248657184 0.0011071449 0.048845695   0.16598685 0.32799761 1946.452
## sigma     0.229545804 0.0001694504 0.007838642   0.21636947 0.24240175 2139.915
##               Rhat
## beta[1] 0.9998704
## beta[2] 0.9995806
## beta[3] 0.9993113
## beta[4] 0.9989858
## sigma   1.0001887
```
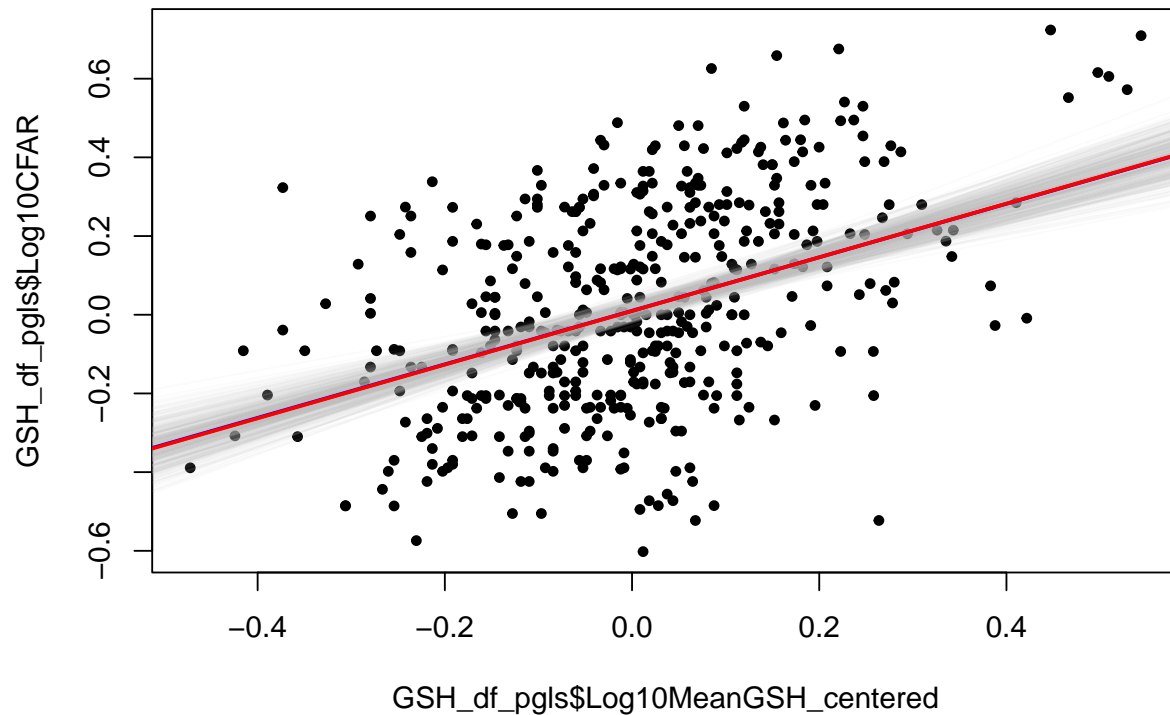
Model 5 - CFAR ~ GSH * PrimaryHabitat

1. brms model

```
habitat_model_2 <- brm(Log10CFAR ~ Log10MeanGSH_centered * PrimaryHabitat,
                       data = GSH_df_pgls,
                       family = gaussian(),
                       prior = habitat_prior2,
                       cores = 2)
```
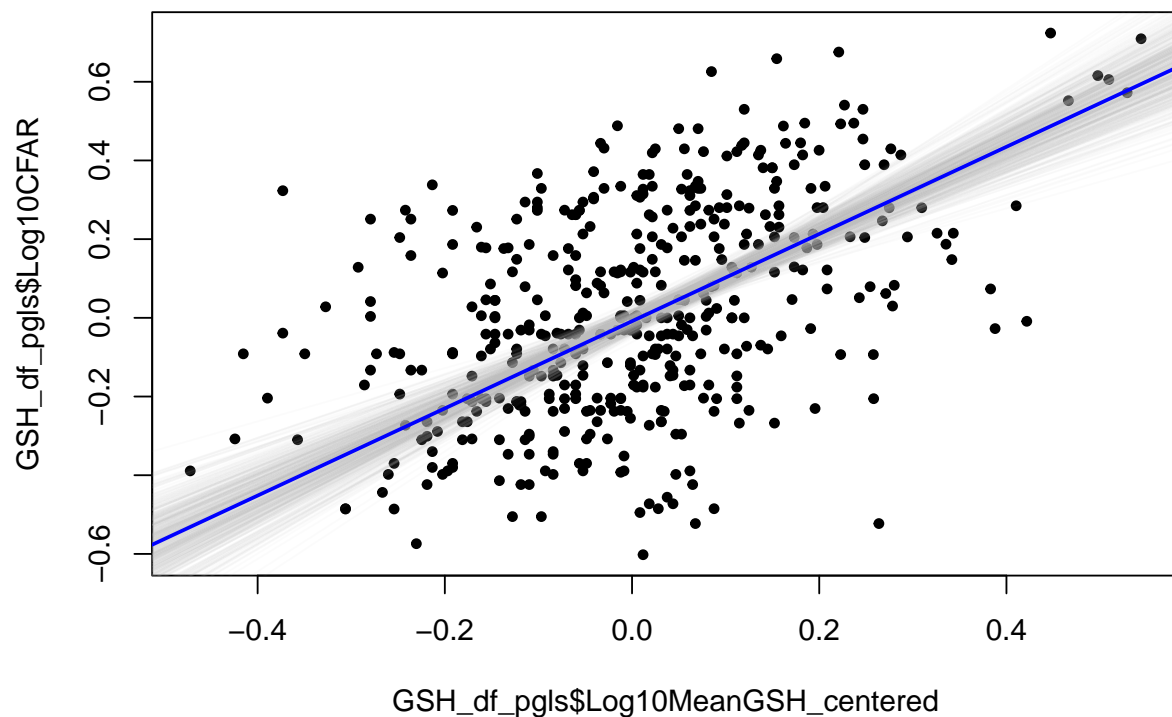
```
summary(habitat_model_2)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered * PrimaryHabitat
##    Data: GSH_df_pgls (Number of observations: 456)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Population-Level Effects:
##                                            Estimate Est.Error l-95% CI
## Intercept                                     -0.01      0.02    -0.04
## Log10MeanGSH_centered                          1.11      0.13     0.86
## PrimaryHabitatDeepwater                       -0.00      0.02    -0.05
## PrimaryHabitatPelagic                          0.29      0.06     0.18
## Log10MeanGSH_centered:PrimaryHabitatDeepwater -0.64      0.16    -0.97
## Log10MeanGSH_centered:PrimaryHabitatPelagic   -0.57      0.22    -1.00
##                                            u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept                                      0.02 1.00     3799     2812
## Log10MeanGSH_centered                          1.37 1.00     2288     2152
## PrimaryHabitatDeepwater                        0.04 1.00     3323     2309
## PrimaryHabitatPelagic                          0.40 1.00     3026     2709
## Log10MeanGSH_centered:PrimaryHabitatDeepwater -0.32 1.00     2347     2400
## Log10MeanGSH_centered:PrimaryHabitatPelagic   -0.13 1.00     2336     2575
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.23      0.01     0.21     0.24 1.00     3870     2628
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

2. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // CFAR
}
```

The parameters accepted by the model.

```
parameters {
  real alpha;
  vector[K] beta;
  real<lower=0> sigma;
}
```

The model to be estimated. We model the output 'y' to be normally distributed with mean 'mu' and standard deviation 'sigma'.

```
model {
```

```
  sigma ~ student_t(3, 0, 10);

  y ~ normal(alpha + x * beta , sigma);
}
```

```
fit7_summary <- summary(fit7, pars = c("beta[1]", "beta[2]","beta[3]","beta[4]", "sigma"), probs = c(0.0
print(fit7_summary)
```

```
##                   mean       se_mean          sd          5%          95%      n_eff
## beta[1]   0.009749888 0.0004713174 0.016827876 -0.01878946 0.03663943 1274.770
## beta[2]   0.682363533 0.0017507064 0.073871816  0.55984160 0.80141279 1780.455
## beta[3]  -0.009820551 0.0006737056 0.023797202 -0.04919637 0.02937670 1247.704
## beta[4]   0.249602076 0.0010174195 0.046573051  0.17226517 0.32740428 2095.411
## sigma     0.229530477 0.0001583226 0.007537908  0.21747364 0.24195594 2266.815
##                Rhat
## beta[1] 0.9998695
## beta[2] 0.9989966
## beta[3] 0.9988179
## beta[4] 0.9987428
## sigma   0.9995975
```
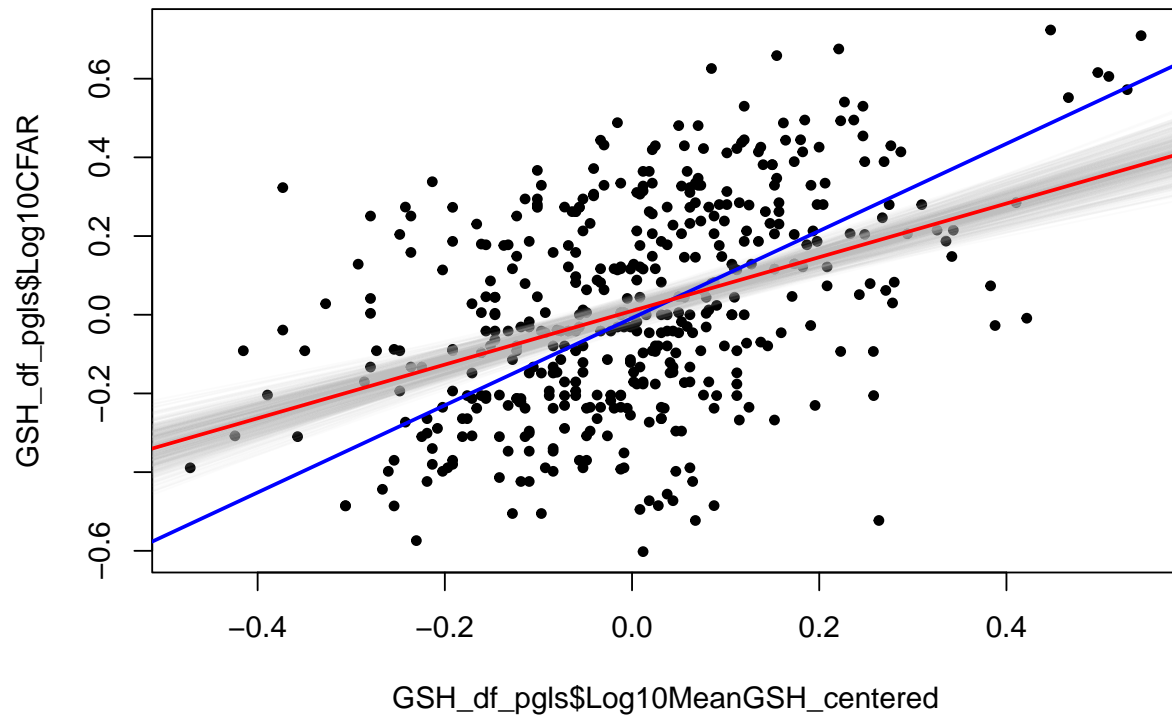
Model 6 - CFAR ~ GSH + Phylogeny

1. PGLS model

```
pglsMod1 <- gls(Log10CFAR ~ Log10MeanGSH_centered, correlation = corPagel(phy = SampleTree, value = 0.5
```

```
summary(pglsMod1)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: Log10CFAR ~ Log10MeanGSH_centered
##   Data: GSH_df_pgls
##          AIC       BIC    logLik
##    -427.4206 -410.9306 217.7103
##
## Correlation Structure: corPagel
##  Formula: ~Binomial
##  Parameter estimate(s):
##     lambda
## 0.7030228
##
## Coefficients:
##                           Value  Std.Error   t-value p-value
## (Intercept)          -0.0224754 0.06282858 -0.357726  0.7207
## Log10MeanGSH_centered  0.3728837 0.05950418  6.266513  0.0000
##
##  Correlation:
##                       (Intr)
## Log10MeanGSH_centered -0.065
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.7884571 -0.6933649  0.1151372  1.0885641  2.9034051
##
## Residual standard error: 0.2147249
## Degrees of freedom: 456 total; 454 residual
```
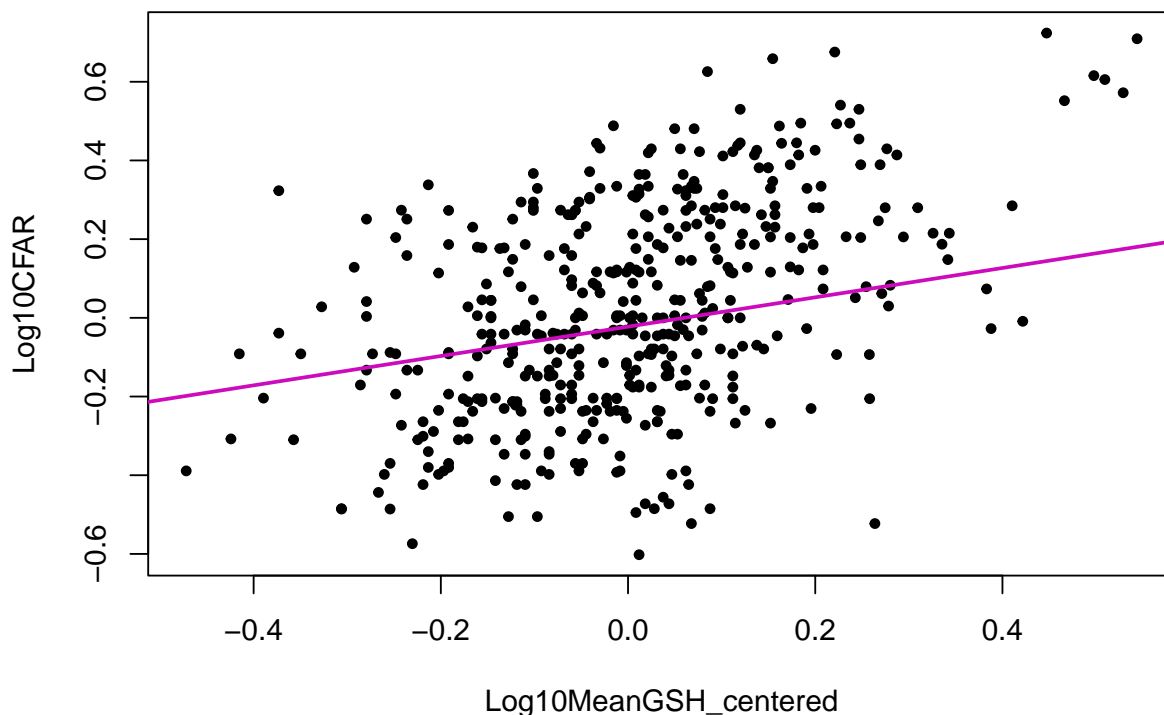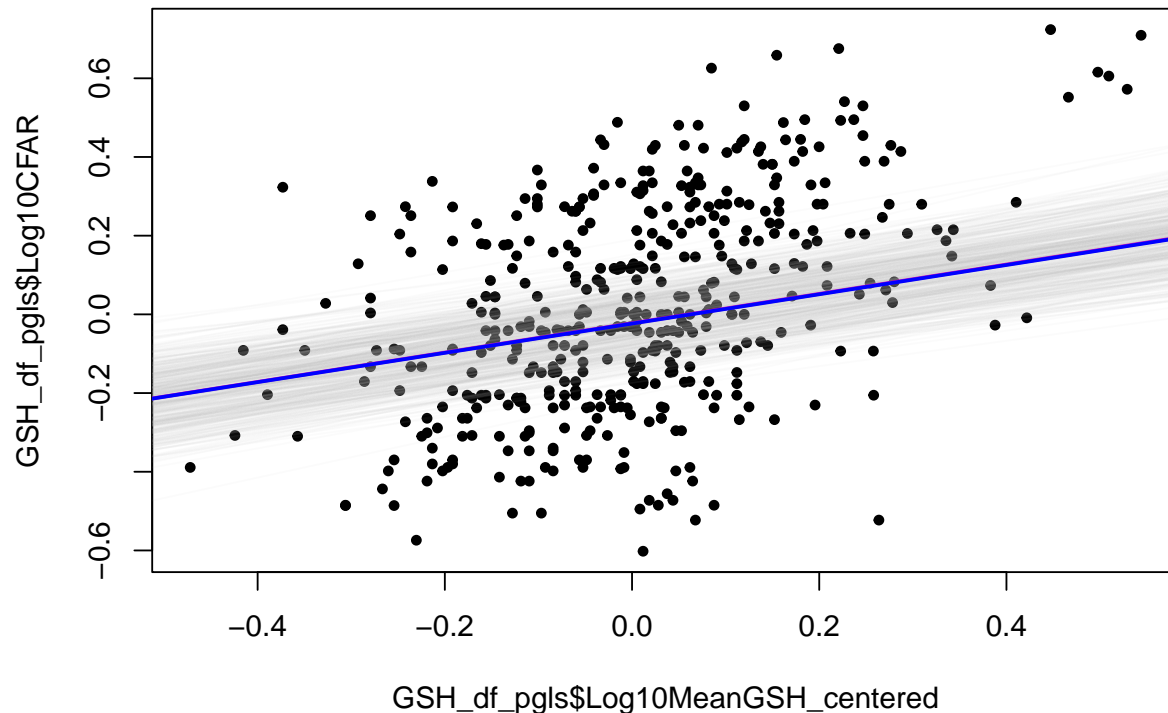
2. brms model

```
Model_simple <- brm(Log10CFAR ~ Log10MeanGSH_centered + (1|gr(Binomial, cov = A)),
                    data = GSH_df_pgls,
                    family = gaussian(),
                    data2 = list(A = A),
                    prior = Simple_prior,
                    cores = 2)
```

```
summary(Model_simple)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + (1 | gr(Binomial, cov = A))
##    Data: GSH_df_pgls (Number of observations: 456)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Group-Level Effects:
## ~Binomial (Number of levels: 456)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.01      0.00     0.01     0.01 1.00     1200     1938
##
## Population-Level Effects:
```

```
##                         Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                  -0.02      0.06    -0.15     0.10 1.00     1257
## Log10MeanGSH_centered       0.37      0.06     0.25     0.49 1.00     4238
##                         Tail_ESS
## Intercept                   2076
## Log10MeanGSH_centered       3483
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.12      0.01     0.11     0.13 1.00     1717     2848
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```


3. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // outcome (CFAR)
```

18

```
    matrix[N, N] d_mat; // sigma matrix
    matrix[N, N] A; // vcov matrix
        }
```

The parameters accepted by the model.

```
 parameters {
        real alpha;
        vector[K] beta; // coefficients
        real<lower=0> sigma; // error
        real<lower=0,upper=1> lambda; // phylogenetic signal
        }


        transformed parameters {

          matrix[N, N] sigma_mat;
          matrix[N, N] sigma_total;

          vector[N] mu_y;


          sigma_mat = (1-lambda)*d_mat + lambda*A;
          sigma_total = sigma*sigma_mat;



        }
```

The model to be estimated. We model the output 'y' to be normally distributed with mean 'mu' and standard deviation 'sigma'.

```
model {

  beta ~ student_t(3, 0, 10);
  lambda ~ uniform(0,1);
  sigma ~ student_t(3, 0, 2.5);

  y ~ multi_normal(alpha + x * beta, sigma_total);
}
```

```
fit4
```

```
## Inference for Stan model: StanModel3.
## 4 chains, each with iter=1000; warmup=500; thin=10;
## post-warmup draws per chain=50, total post-warmup draws=200.
##
##            mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
## alpha     -0.02     0.0 0.06  -0.15  -0.06  -0.03   0.02   0.09   219 0.99
## beta[1]    0.37     0.0 0.06   0.25   0.33   0.36   0.41   0.47   216 1.01
## lp__     625.81     0.1 1.36 622.82 625.03 626.10 626.84 627.47   193 1.02
##
```

19

Samples were drawn using NUTS(diag_e) at Thu Mar 03 14:30:50 2022.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
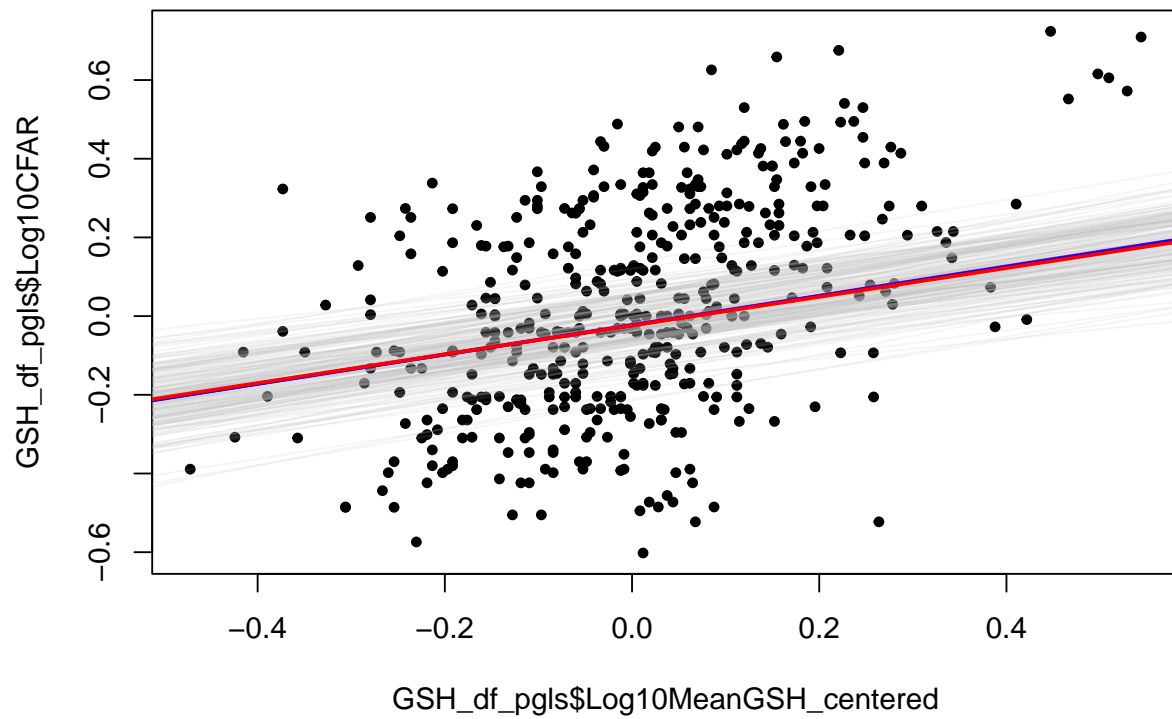
Model 7 - CFAR ~ GSH * Max Size + Phylogeny

1. PGLS model

```
pglsMod2 <- gls(Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered, correlation = corPagel(phy =
```

```
summary(pglsMod2)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered
##   Data: GSH_df_pgls
##         AIC       BIC    logLik
##   -457.6884 -432.9534 234.8442
##
## Correlation Structure: corPagel
##  Formula: ~Binomial
##  Parameter estimate(s):
##     lambda
## 0.6671208
##
## Coefficients:
##                                               Value  Std.Error    t-value
## (Intercept)                              -0.0575841 0.05765480 -0.998773
## Log10MeanGSH_centered                     0.3102797 0.05966101  5.200711
## Log10MaxSize_centered                     0.1272174 0.03524896  3.609111
## Log10MeanGSH_centered:Log10MaxSize_centered 0.4334147 0.12326258  3.516191
##                                           p-value
## (Intercept)                                0.3184
## Log10MeanGSH_centered                      0.0000
## Log10MaxSize_centered                      0.0003
## Log10MeanGSH_centered:Log10MaxSize_centered  0.0005
##
##  Correlation:
##                                           (Intr) Lg10MGSH_ L10MS_
## Log10MeanGSH_centered                     -0.053
## Log10MaxSize_centered                     -0.025 -0.268
## Log10MeanGSH_centered:Log10MaxSize_centered -0.095  0.030     -0.293
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.7953253 -0.5679176  0.2824644  1.2747739  2.9958494
##
## Residual standard error: 0.1999719
## Degrees of freedom: 456 total; 452 residual
```
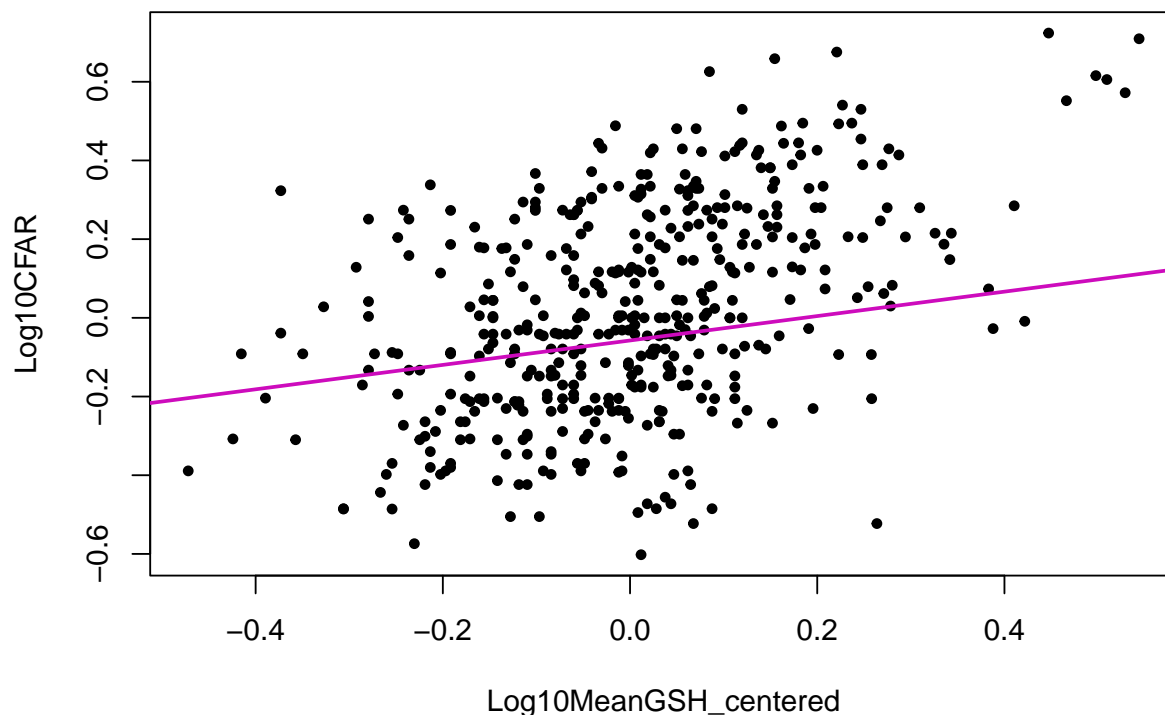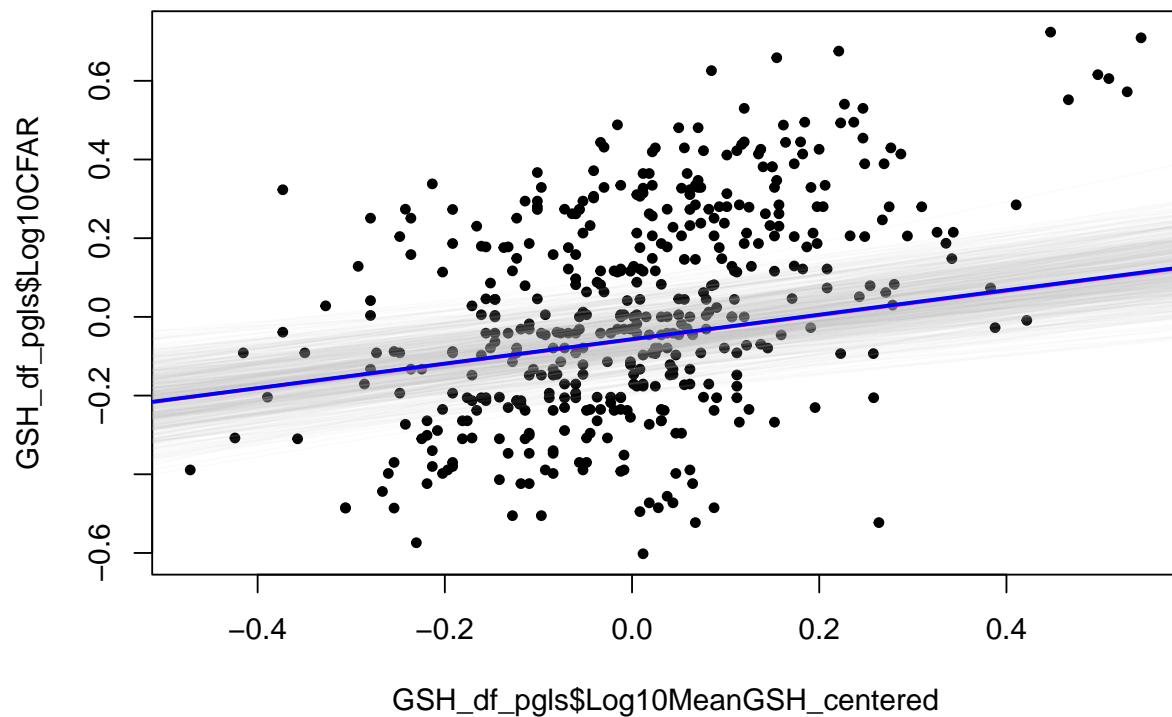
2. brms model

```
Model_BS <- brm(Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered + (1|gr(Binomial, cov = A)),
                data = GSH_df_pgls,
                family = gaussian(),
                data2 = list(A = A),
                prior = BS_prior,
            sample_prior = TRUE, chains = 4, cores = 2)
```

```
summary(Model_BS)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered * Log10MaxSize_centered + (1 | gr(Binomial, cov = A))
##    Data: GSH_df_pgls (Number of observations: 456)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Group-Level Effects:
## ~Binomial (Number of levels: 456)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.01      0.00     0.01     0.01 1.00     1303     2376
##
## Population-Level Effects:
```

```
##                                              Estimate Est.Error l-95% CI
## Intercept                                       -0.06      0.06    -0.17
## Log10MeanGSH_centered                            0.31      0.06     0.19
## Log10MaxSize_centered                            0.13      0.04     0.06
## Log10MeanGSH_centered:Log10MaxSize_centered      0.43      0.12     0.20
##                                              u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept                                        0.05 1.00     1824     2402
## Log10MeanGSH_centered                            0.43 1.00     6595     3478
## Log10MaxSize_centered                            0.20 1.00     4866     3342
## Log10MeanGSH_centered:Log10MaxSize_centered      0.67 1.00     5734     3590
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.12      0.01     0.11     0.13 1.00     2164     3283
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```


3. STAN model

Data considered

```
data {
```

```
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
  matrix[N,K] x; // predictor matrix
  vector[N] y; // outcome (CFAR)
  matrix[N, N] d_mat; // sigma matrix
  matrix[N, N] A; // vcov matrix
      }
```

The parameters accepted by the model.

```
 parameters {
      real alpha;
      vector[K] beta; // coefficients
      real<lower=0> sigma; // error
      real<lower=0,upper=1> lambda; // phylogenetic signal
      }


      transformed parameters {

        matrix[N, N] sigma_mat;
        matrix[N, N] sigma_total;

        vector[N] mu_y;


        sigma_mat = (1-lambda)*d_mat + lambda*A;
        sigma_total = sigma*sigma_mat;



      }
```

The model to be estimated. We model the output 'y' to be normally distributed with mean 'mu' and standard deviation 'sigma'.

```
model {

  beta ~ student_t(3, 0, 10);
  lambda ~ uniform(0,1);
  sigma ~ student_t(3, 0, 2.5);

  y ~ multi_normal(alpha + x * beta, sigma_total);
}
```

```
fit5
```

```
## Inference for Stan model: StanModel3.
## 4 chains, each with iter=1000; warmup=500; thin=10;
## post-warmup draws per chain=50, total post-warmup draws=200.
##
##          mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
```

```
## alpha     -0.06     0.00 0.06  -0.17  -0.10  -0.06  -0.02    0.05    154 1.01
## beta[1]    0.31     0.00 0.06   0.20   0.26   0.32   0.35    0.44    272 0.99
## beta[2]    0.12     0.00 0.04   0.06   0.10   0.12   0.15    0.20    194 0.99
## beta[3]    0.43     0.01 0.12   0.22   0.35   0.44   0.51    0.65    210 0.99
## lp__     641.68     0.12 1.76 637.09 640.81 642.00 643.00  643.99    199 1.00
##
## Samples were drawn using NUTS(diag_e) at Mon Mar 07 12:09:07 2022.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```
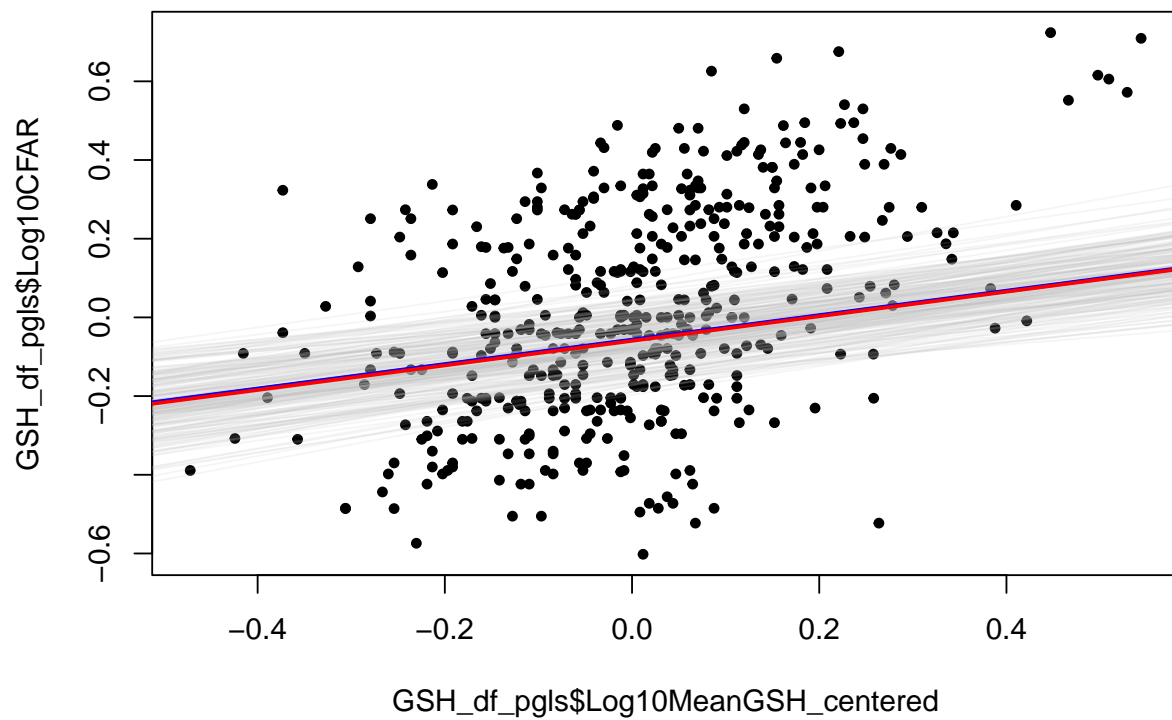
Model 8 - CFAR ~ GSH + Max Size + Phylogeny

1. PGLS model

```
pglsMod3 <- gls(Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered, correlation = corPagel(phy =
```

```
summary(pglsMod3)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered
##   Data: GSH_df_pgls
##        AIC       BIC    logLik
##   -447.4392 -426.8268 228.7196
##
## Correlation Structure: corPagel
##  Formula: ~Binomial
##  Parameter estimate(s):
##    lambda
## 0.6786106
##
## Coefficients:
##                            Value  Std.Error    t-value p-value
## (Intercept)           -0.03827951 0.05915461 -0.647110  0.5179
## Log10MeanGSH_centered  0.30222706 0.06034147  5.008613  0.0000
## Log10MaxSize_centered  0.16260955 0.03412415  4.765233  0.0000
##
##  Correlation:
##                       (Intr) L10MGS
## Log10MeanGSH_centered -0.050
## Log10MaxSize_centered -0.056 -0.268
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.8353741 -0.5472167  0.2440250  1.1919410  2.8367324
##
## Residual standard error: 0.2047976
## Degrees of freedom: 456 total; 453 residual
```
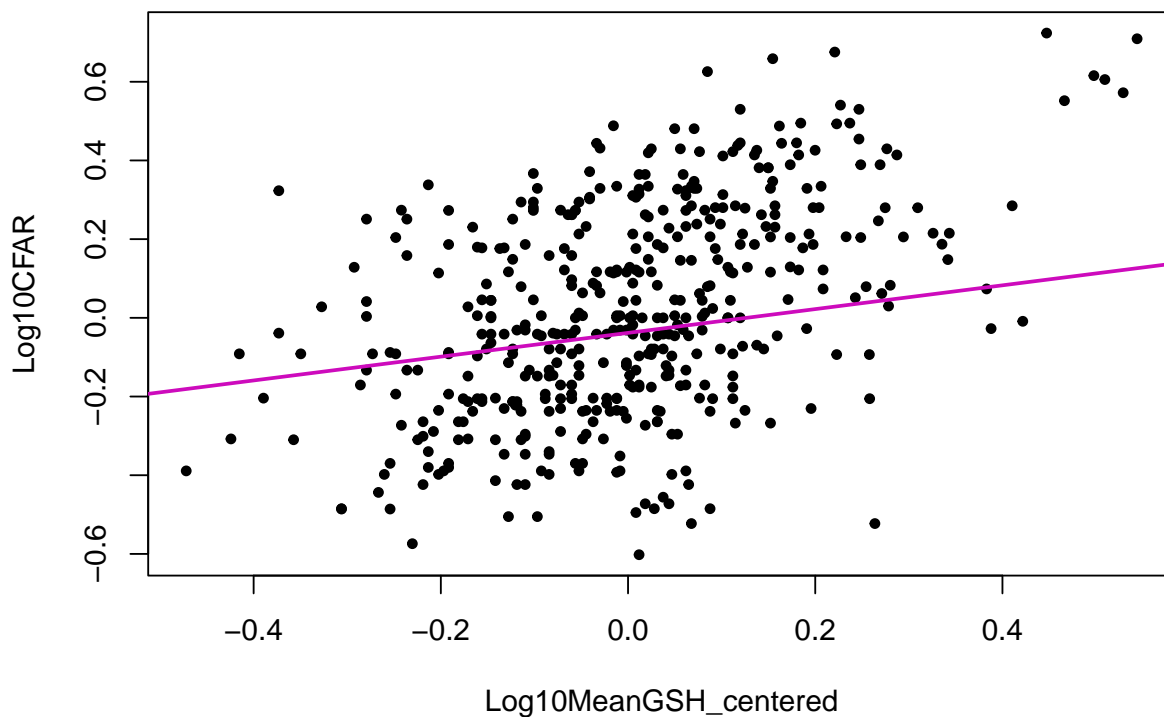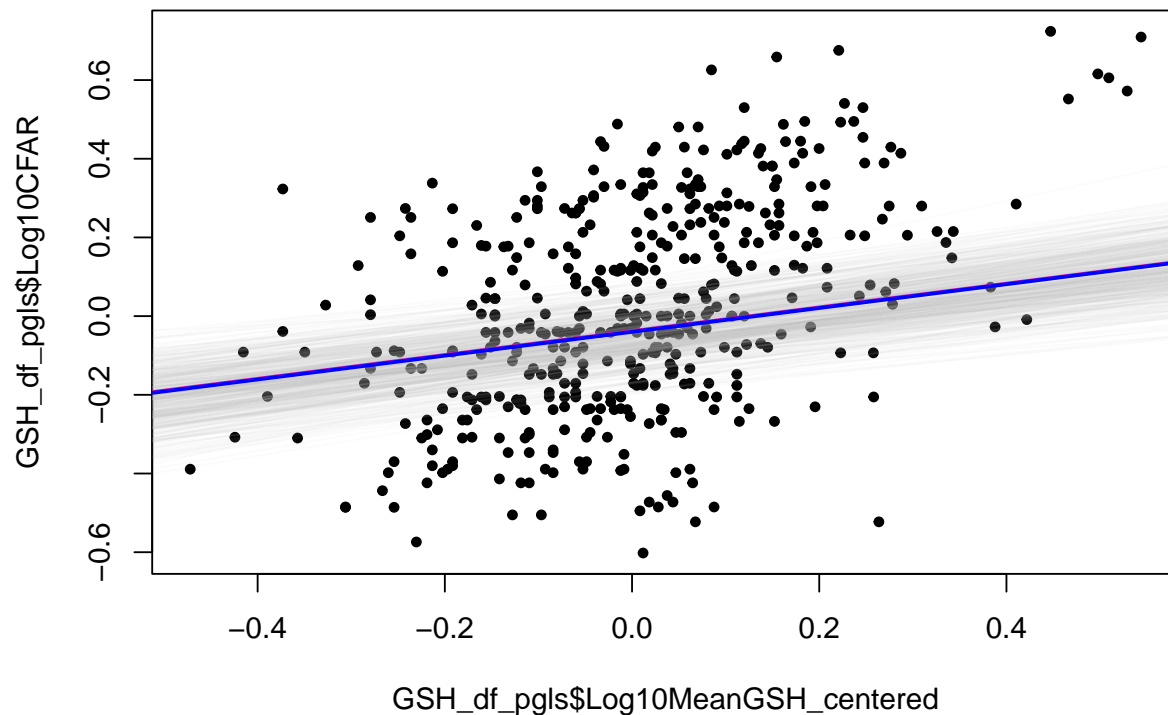
2. brms model

```
Model_BS <- brm(Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered + (1|gr(Binomial, cov = A)),
                data = GSH_df_pgls,
                family = gaussian(),
                data2 = list(A = A),
                prior = BS_prior,
            sample_prior = TRUE, chains = 4, cores = 2)
```

```
summary(Model_BS2)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: Log10CFAR ~ Log10MeanGSH_centered + Log10MaxSize_centered + (1 | gr(Binomial, cov = A))
##    Data: GSH_df_pgls (Number of observations: 456)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Group-Level Effects:
## ~Binomial (Number of levels: 456)
##                Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.01      0.00     0.01     0.01 1.00     1186     1678
##
## Population-Level Effects:
```

```
##                         Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                  -0.04      0.06    -0.16     0.08 1.00     1283
## Log10MeanGSH_centered       0.30      0.06     0.19     0.42 1.00     3622
## Log10MaxSize_centered       0.16      0.03     0.10     0.23 1.00     3812
##                         Tail_ESS
## Intercept                   2054
## Log10MeanGSH_centered       3227
## Log10MaxSize_centered       3414
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.12      0.01     0.11     0.13 1.00     1686     2672
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```


3. STAN model

Data considered

```
data {
  int <lower=0> N; // number of data points
  int <lower=0> K; // number of predictors
```

28 of 31

```
   matrix[N,K] x; // predictor matrix
   vector[N] y; // outcome (CFAR)
   matrix[N, N] d_mat; // sigma matrix
   matrix[N, N] A; // vcov matrix
       }
```

The parameters accepted by the model.

```
 parameters {
       real alpha;
       vector[K] beta; // coefficients
       real<lower=0> sigma; // error
       real<lower=0,upper=1> lambda; // phylogenetic signal
       }


       transformed parameters {

         matrix[N, N] sigma_mat;
         matrix[N, N] sigma_total;

         vector[N] mu_y;


         sigma_mat = (1-lambda)*d_mat + lambda*A;
         sigma_total = sigma*sigma_mat;



       }
```

The model to be estimated. We model the output 'y' to be normally distributed with mean 'mu' and standard deviation 'sigma'.

```
model {

  beta ~ student_t(3, 0, 10);
  lambda ~ uniform(0,1);
  sigma ~ student_t(3, 0, 2.5);

  y ~ multi_normal(alpha + x * beta, sigma_total);
}
```

```
fit3
```

```
## Inference for Stan model: StanModel2.
## 4 chains, each with iter=1000; warmup=500; thin=1;
## post-warmup draws per chain=500, total post-warmup draws=2000.
##
##            mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
## alpha      0.02    0.00 0.01   0.00   0.01   0.02   0.03   0.04  2075    1
## beta[1]    0.39    0.00 0.08   0.23   0.34   0.39   0.45   0.55  1620    1
```

```
## beta[2]    0.32    0.00 0.04   0.24   0.30   0.32   0.35   0.40 1550     1
## sigma      0.22    0.00 0.01   0.21   0.22   0.22   0.23   0.24 1790     1
## lp__     458.67    0.05 1.41 455.06 458.03 458.99 459.70 460.42  948     1
##
## Samples were drawn using NUTS(diag_e) at Wed Mar 02 11:22:15 2022.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```
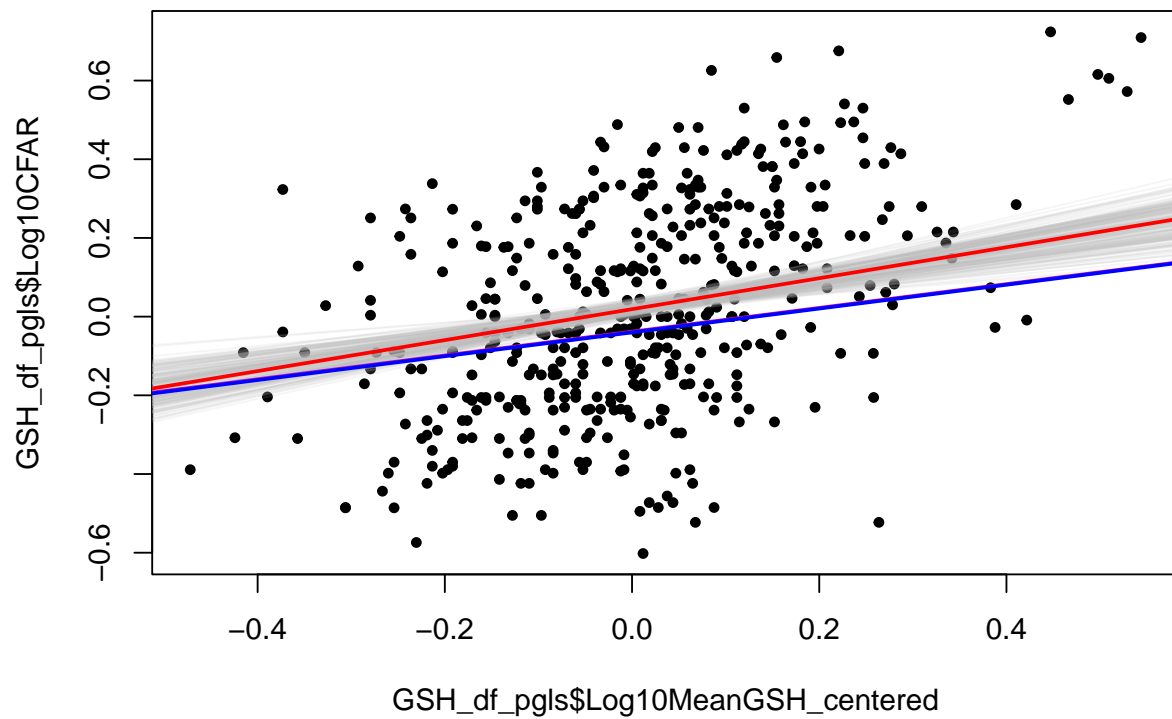
Model 9 - CFAR ~ GSH + PrimaryHabitat + Phylogeny

Model 10 - CFAR ~ GSH * PrimaryHabitat + Phylogeny

Model 11 - CFAR ~ GSH * MaxSize + PrimaryHabitat + Phylogeny