LegumeLegacy

SS8 The statistical analysis of multi-site data (part 2)

From fixed-effects models to random-coefficient mixed models

Emery Wang
Supervisors: John Finn and Caroline Brophy

Funded by the European Union

Outline

- Simulated data introduction
- Raw data exploration
- Fixed-effects DI model
- Mixed-effects DI model
- Model diagnostics
- Ecological interpretation and visualization

Simulated data

Data was simulated under the following LegacyNet experiments:

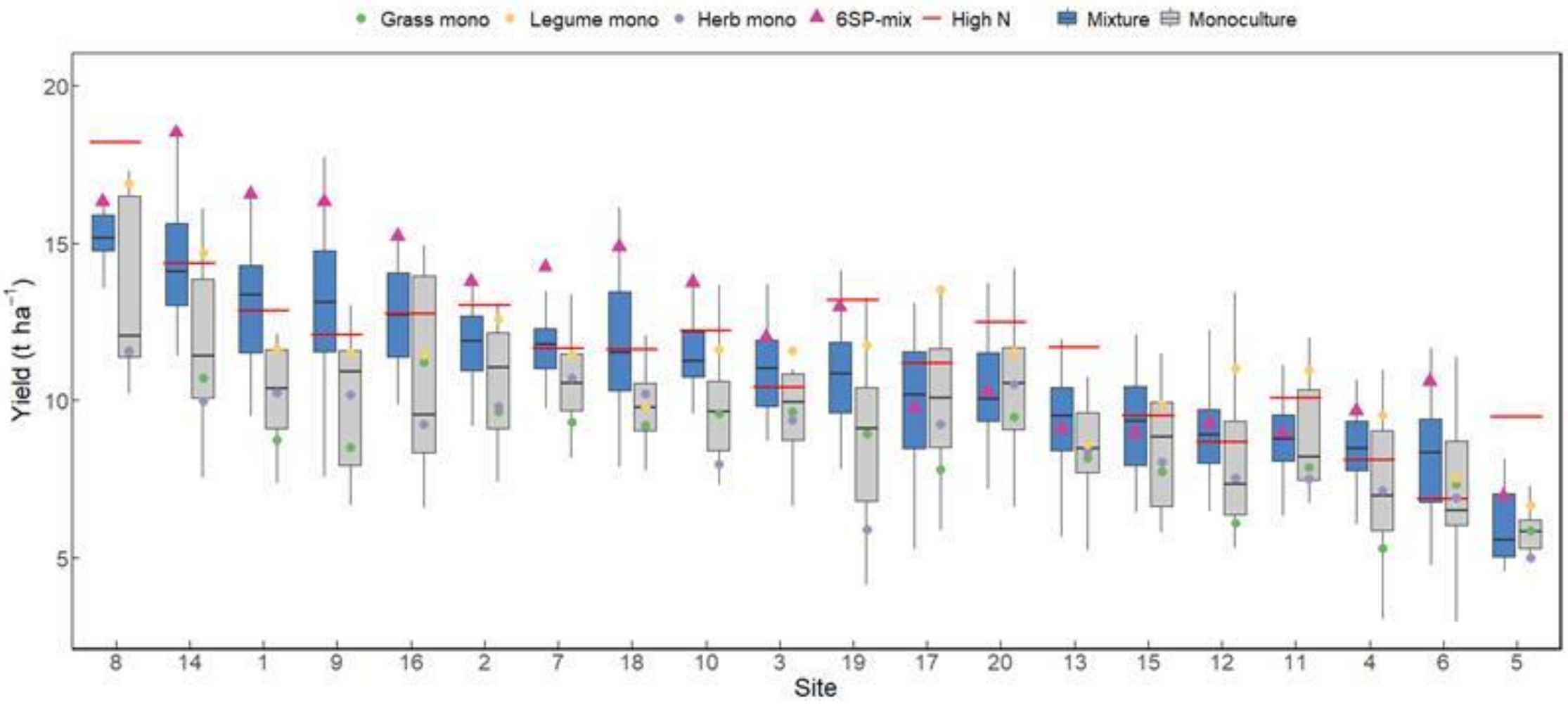
At each site, 6 species were selected from 3 functional groups (2 grasses, 2 legumes, and 2 herbs; denoted G1, G2, L1, L2, H1, H2) for use in a simplex design.

It comprises 33 unique communities of systematically varying sown proportions of each of 6 species. There were monocultures of each of the 6 species, and mixture plots with 2, 3, 4, or 6 species at a moderate N level. Additionally, 5 replicated plots of the G1 monoculture were managed at a high N level. This design comprised 52 plot at each site (O'Malley et al., 2026).

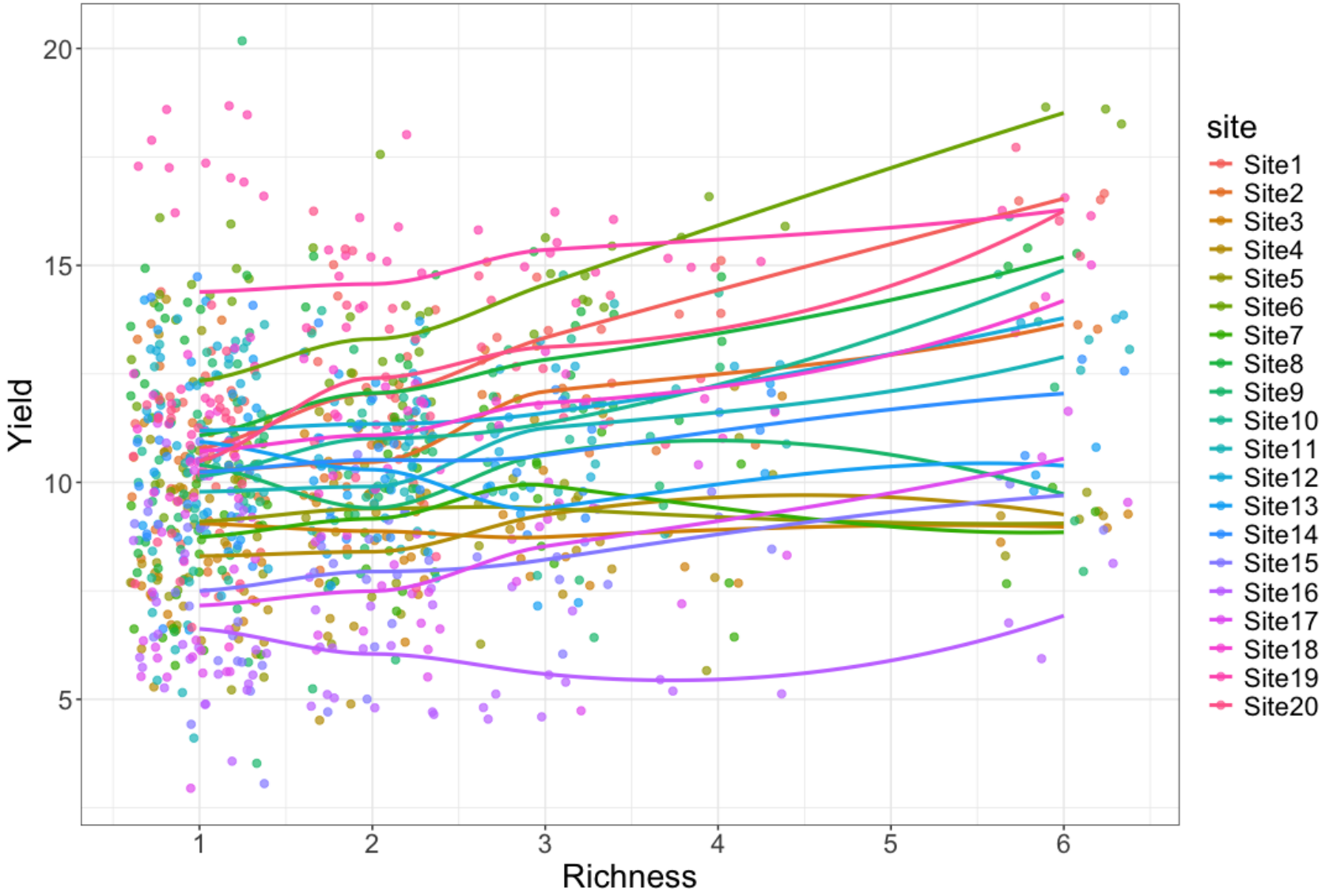
This data was simulated for 20 sites, Yield presents annual dry matter yield for each plot (no repeat measurements).

```
> head(data)
   site Plot Nfert RichnessSP RichnessFG G1 G2 L1 L2 H1 H2 Indicator1     Yield
1 Site1    1   Low          1          1  1  0  0  0  0  0       Core  8.827067
2 Site1    2   Low          1          1  1  0  0  0  0  0       Core  8.580665
3 Site1    3   Low          1          1  1  0  0  0  0  0       Core  7.361423
4 Site1    4   Low          1          1  0  1  0  0  0  0       Core  7.678456
5 Site1    5   Low          1          1  0  1  0  0  0  0       Core  9.897026
6 Site1    6   Low          1          1  0  1  0  0  0  0       Core 10.086476
```

Raw data exploration

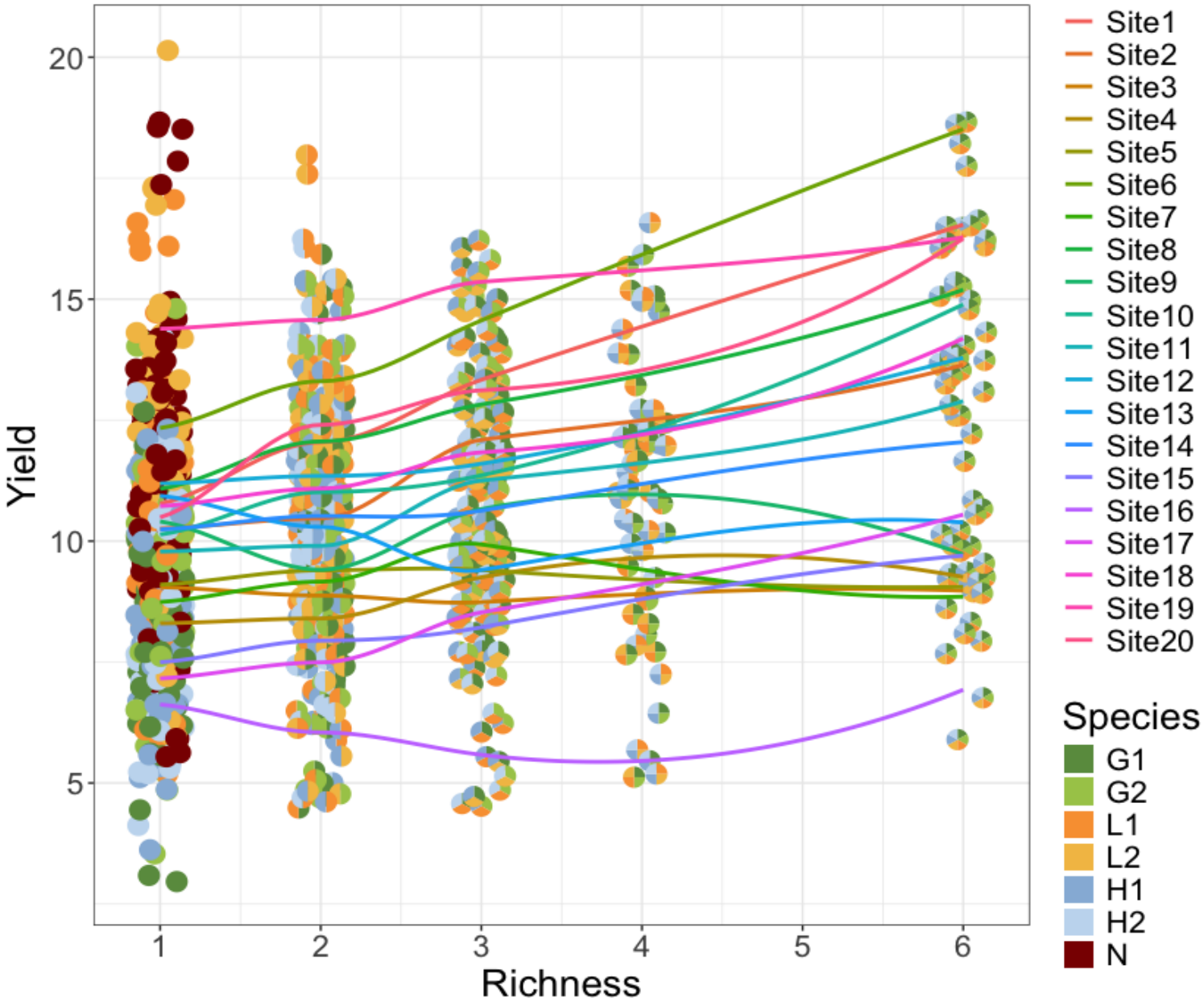


Raw data exploration



What are the trends among sites?
Any extreme values?

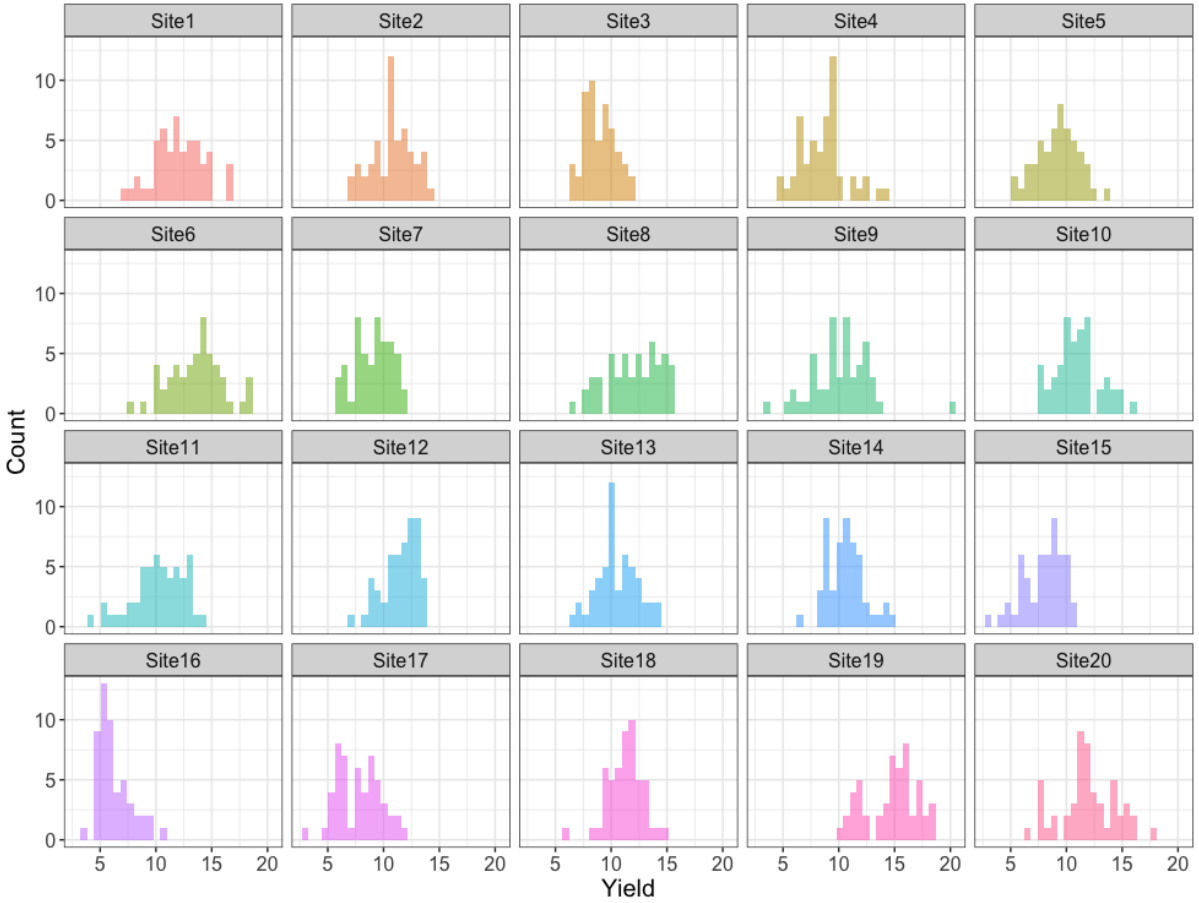
Raw data exploration

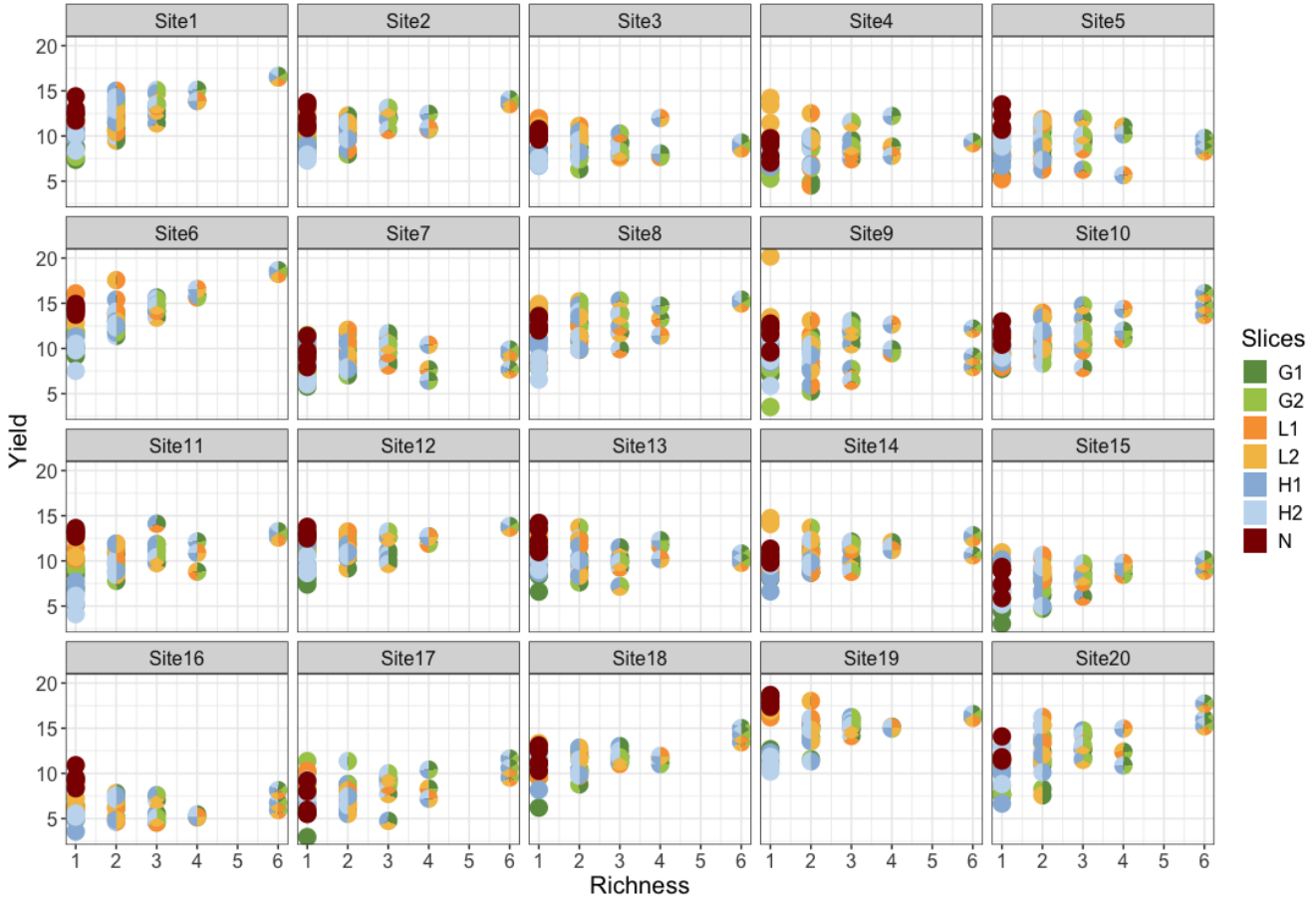


Strong between-site variability is visible.

Yield increases with richness at most sites.

Raw data exploration at site level





The key question is whether diversity effects are consistent across all sites.

Raw data exploration

Follow the steps in SS8. Lab sheet 0

Fixed-effects DI model

- Fit DI models using site as a blocking factor (γ_k).
- Which DI structure gives the best fit?
- Does theta estimation improve model fit?
- Use AICc to compare model.

Example (AV model with theta estimation):

```
DI_AV<- DI(
    y = "Yield",
    prop =
c("G1","G2","L1","L2","H1","H2", "N"),
    DImodel = "AV",
    block = "site",
    estimate_theta = TRUE,
    data = data)
```

$$y_{kl} = \sum_{i=1}^{6} \beta_i P_{ikl} + \sum_{i<j} \delta_{ij} (P_{ikl} P_{jkl})^{\theta} + \alpha X_{kl} + \gamma_k Site_{kl} + \epsilon_{kl}$$

y_{kl}: response in plot *l* at site *k (k = 1,…., 20),*
P_{ikl} and P_{jkl}: sown proportion of species *i* and *j* in plot *l* at site *k,*
X_{kl}: dummy variable for high-N grass plots,
$Site_{kl}$: factors for sites, and $\epsilon_{kl} \sim N(0, \sigma^2)$.

Fixed-effects DI model

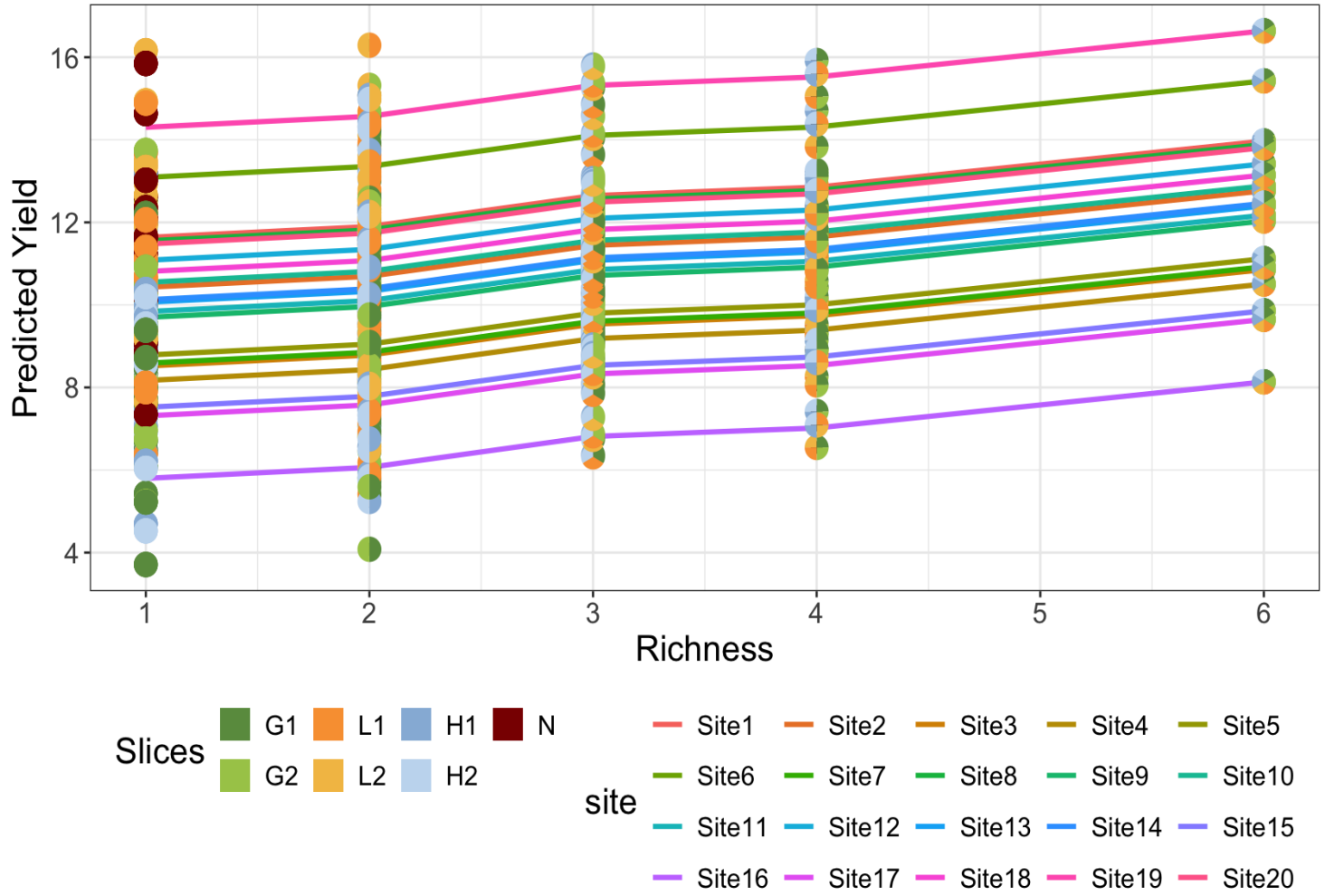
Follow the steps in SS8. Lab sheet 1(a)

Example (AV model with theta estimation):

```
DI_AV<- DI(
    y = "Yield",
    prop =
c("G1","G2","L1","L2","H1","H2", "N"),
    DImodel = "AV",
    block = "site",
    estimate_theta = TRUE,
    data = data)
```

Fixed-effects DI model

```
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
G1_ID       9.30671    0.28828  32.284  < 2e-16 ***
G2_ID      10.83385    0.28828  37.581  < 2e-16 ***
L1_ID      11.73003    0.28828  40.690  < 2e-16 ***
L2_ID      12.99949    0.28828  45.094  < 2e-16 ***
H1_ID      10.45012    0.28828  36.250  < 2e-16 ***
H2_ID      10.27917    0.28828  35.657  < 2e-16 ***
N_ID       13.02081    0.27729  46.958  < 2e-16 ***
AV          4.60035    0.34775  13.229  < 2e-16 ***
siteSite1   0.15876    0.32344   0.491  0.62363
siteSite2  -1.04662    0.32344  -3.236  0.00125 **
siteSite3  -2.95500    0.32344  -9.136  < 2e-16 ***
siteSite4  -3.30431    0.32344 -10.216  < 2e-16 ***
siteSite5  -2.68971    0.32344  -8.316 2.91e-16 ***
siteSite6   1.61554    0.32344   4.995 6.93e-07 ***
siteSite7  -2.88601    0.32344  -8.923  < 2e-16 ***
siteSite8   0.07646    0.32344   0.236  0.81317
siteSite9  -1.77689    0.32344  -5.494 4.98e-08 ***
siteSite10 -0.92723    0.32344  -2.867  0.00423 **
siteSite11 -1.63575    0.32344  -5.057 5.04e-07 ***
siteSite12 -0.39315    0.32344  -1.216  0.22444
siteSite13 -1.41056    0.32344  -4.361 1.43e-05 ***
siteSite14 -1.34937    0.32344  -4.172 3.28e-05 ***
siteSite15 -3.95375    0.32344 -12.224  < 2e-16 ***
siteSite16 -5.67533    0.32344 -17.547  < 2e-16 ***
siteSite17 -4.16164    0.32344 -12.867  < 2e-16 ***
siteSite18 -0.66454    0.32344  -2.055  0.04017 *
siteSite19  2.82863    0.32344   8.746  < 2e-16 ***
```



- This model assumes all sites share the same trend but different additive effects by sites.
- Is it realistic from an ecological perspective?

Fixed-effects DI model

- What if diversity effects themselves vary among sites?

- Use "DI" or "lm" function to allow each effect varying in each site.
- Which DI structure gives the best fit?
- Do interaction terms improve model fit?
- Use AICc to compare model.

Example (AV model):

```
DI_AV_inter<- DI(
    y = "Yield",
    prop =
c("G1","G2","L1","L2","H1","H2", "N"),
    custom_formula = Yield ~ 0 +
(G1+G2+L1+L2+H1+H2+N+AV):site,
    data = data)
```

Or

```
AV_inter <- lm(Yield ~ 0 +
(G1+G2+L1+L2+H1+H2+N+AV):site,
data = data)
```

$$y_{kl} = \left(\sum_{i=1}^{6} \beta_i P_{ikl} + \sum_{i<j} \delta_{ij} (P_{ikl} P_{jkl})^{\theta} + \alpha X_{kl} \right) \times \gamma_k Site_{kl} + \epsilon_{kl}$$

y_{kl}: response in plot *l* at site *k (k = 1,…., 20),*
P_{ikl} and P_{jkl}: sown proportion of species *i* and *j* in plot *l* at site *k,*
X_{kl}: dummy variable for high-N grass plots,
$Site_{kl}$: factors for sites, and $\epsilon_{kl} \sim N(0, \sigma^2)$.

Fixed-effects DI model

Follow the steps in SS8. Lab sheet 1(b)

Example (AV model):

```
DI_AV_inter<- DI(
    y = "Yield",
    prop =
c("G1","G2","L1","L2","H1","H2", "N"),
    custom_formula = Yield ~ 0 +
(G1+G2+L1+L2+H1+H2+N+AV):site,
    data = data)
```

Or

```
AV_inter <- lm(Yield ~ 0 +
(G1+G2+L1+L2+H1+H2+N+AV):site,
data = data)
```

Note: Use “DI_data” function to create interaction terms with estimated theta and include them in the dataset.

```
newlist <- DI_data(prop = c("G1", "G2", "L1", "L2", "H1", "H2"),
                   FG = c("G", "G", "L", "L", "H", "H"),
                   data = data, theta = estimated theta,
                   what = c("AV","FG"))
data <-  data.frame(data, "AV" = newlist$AV, newlist$FG)
```

Results

Every site can now have:
different species identity effects,
different interaction effects,
different high N plot effects.

For example:
FG model across sites will have: (6 identity + 6 FG interactions + 1 high N) × 20 sites = 260 parameters.

This greatly increases model complexity and parameter count.

Coefficients:

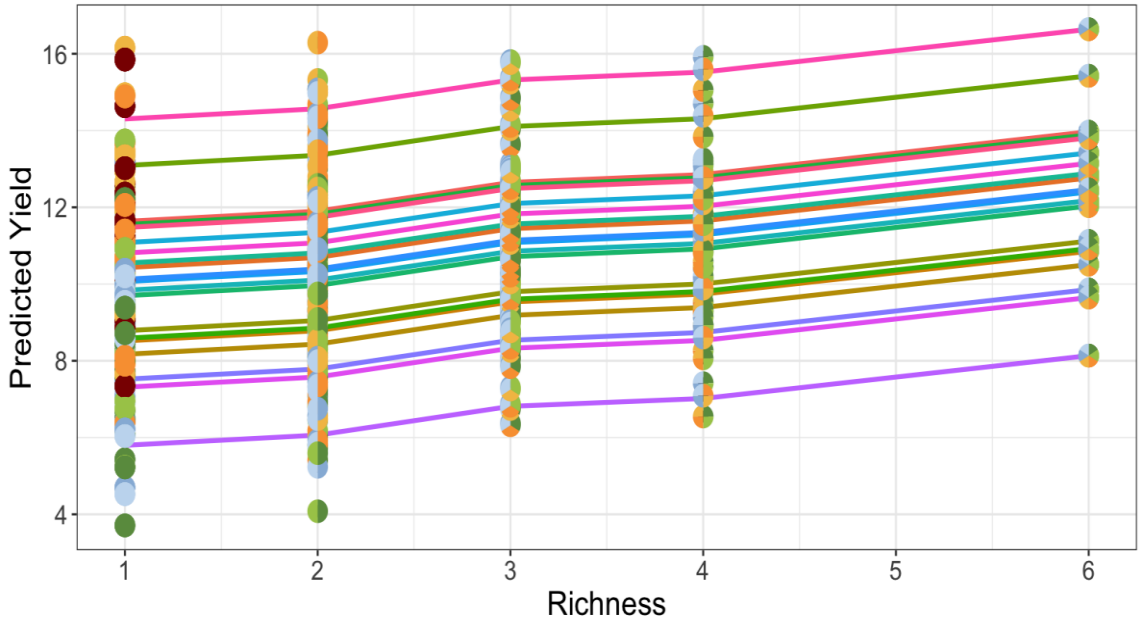
	Estimate	Std. Error	t value	Pr(>|t|)	
G1:siteSite1	7.630777	0.577545	13.212	< 2e-16	***
G1:siteSite2	9.246478	0.577545	16.010	< 2e-16	***
G1:siteSite3	7.724023	0.577545	13.374	< 2e-16	***
G1:siteSite4	6.038133	0.577545	10.455	< 2e-16	***
G1:siteSite5	7.701112	0.577545	13.334	< 2e-16	***
G1:siteSite6	9.796981	0.577545	16.963	< 2e-16	***
G1:siteSite7	6.576818	0.577545	11.388	< 2e-16	***
G1:siteSite8	8.273304	0.577545	14.325	< 2e-16	***
G1:siteSite9	8.650270	0.577545	14.978	< 2e-16	***
G1:siteSite10	8.881625	0.577545	15.378	< 2e-16	***
G1:siteSite11	8.509208	0.577545	14.733	< 2e-16	***
G1:siteSite12	7.980230	0.577545	13.817	< 2e-16	***
G1:siteSite13	7.999209	0.577545	13.850	< 2e-16	***
G1:siteSite14	8.476395	0.577545	14.677	< 2e-16	***
G1:siteSite15	4.629448	0.577545	8.016	3.99e-15	***
G1:siteSite16	6.318707	0.577545	10.941	< 2e-16	***
G1:siteSite17	4.859487	0.577545	8.414	< 2e-16	***
G1:siteSite18	8.803678	0.577545	15.243	< 2e-16	***
G1:siteSite19	12.086071	0.577545	20.927	< 2e-16	***
G1:siteSite20	7.385215	0.577545	12.787	< 2e-16	***

...

bfg_H_L:siteSite1	3.378898	0.981315	3.443	0.000605	***
bfg_H_L:siteSite2	1.589859	0.981315	1.620	0.105608	
bfg_H_L:siteSite3	1.897920	0.981315	1.934	0.053467	.
bfg_H_L:siteSite4	-1.708552	0.981315	-1.741	0.082063	.
bfg_H_L:siteSite5	-1.354445	0.981315	-1.380	0.167910	
bfg_H_L:siteSite6	3.442073	0.981315	3.508	0.000478	***
bfg_H_L:siteSite7	0.957223	0.981315	0.975	0.329640	
bfg_H_L:siteSite8	1.670911	0.981315	1.703	0.089018	.
bfg_H_L:siteSite9	-3.065613	0.981315	-3.124	0.001850	**
bfg_H_L:siteSite10	3.465511	0.981315	3.531	0.000437	***
bfg_H_L:siteSite11	3.406342	0.981315	3.471	0.000547	***
bfg_H_L:siteSite12	0.383108	0.981315	0.390	0.696345	
bfg_H_L:siteSite13	-3.371169	0.981315	-3.435	0.000623	***
bfg_H_L:siteSite14	-0.496410	0.981315	-0.506	0.613096	
bfg_H_L:siteSite15	1.411904	0.981315	1.439	0.150612	
bfg_H_L:siteSite16	-1.089359	0.981315	-1.110	0.267298	
bfg_H_L:siteSite17	-0.965348	0.981315	-0.984	0.325554	
bfg_H_L:siteSite18	1.775965	0.981315	1.810	0.070715	.
bfg_H_L:siteSite19	0.840461	0.981315	0.856	0.392004	
bfg_H_L:siteSite20	5.740473	0.981315	5.850	7.24e-09	***

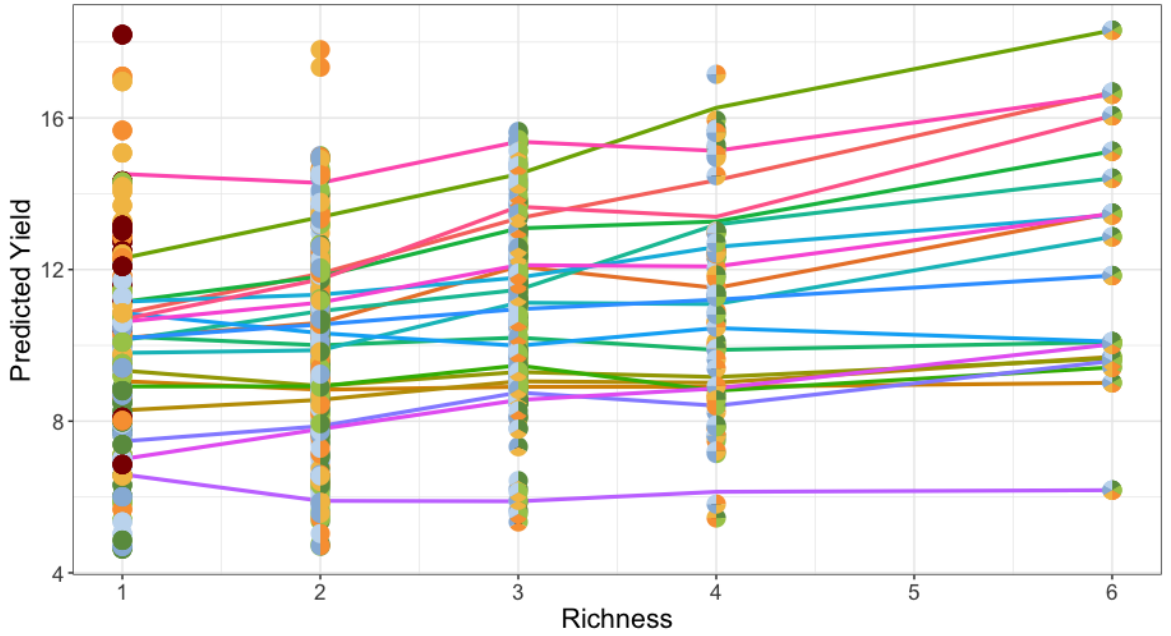
Predictions

(A) Sites are block effects





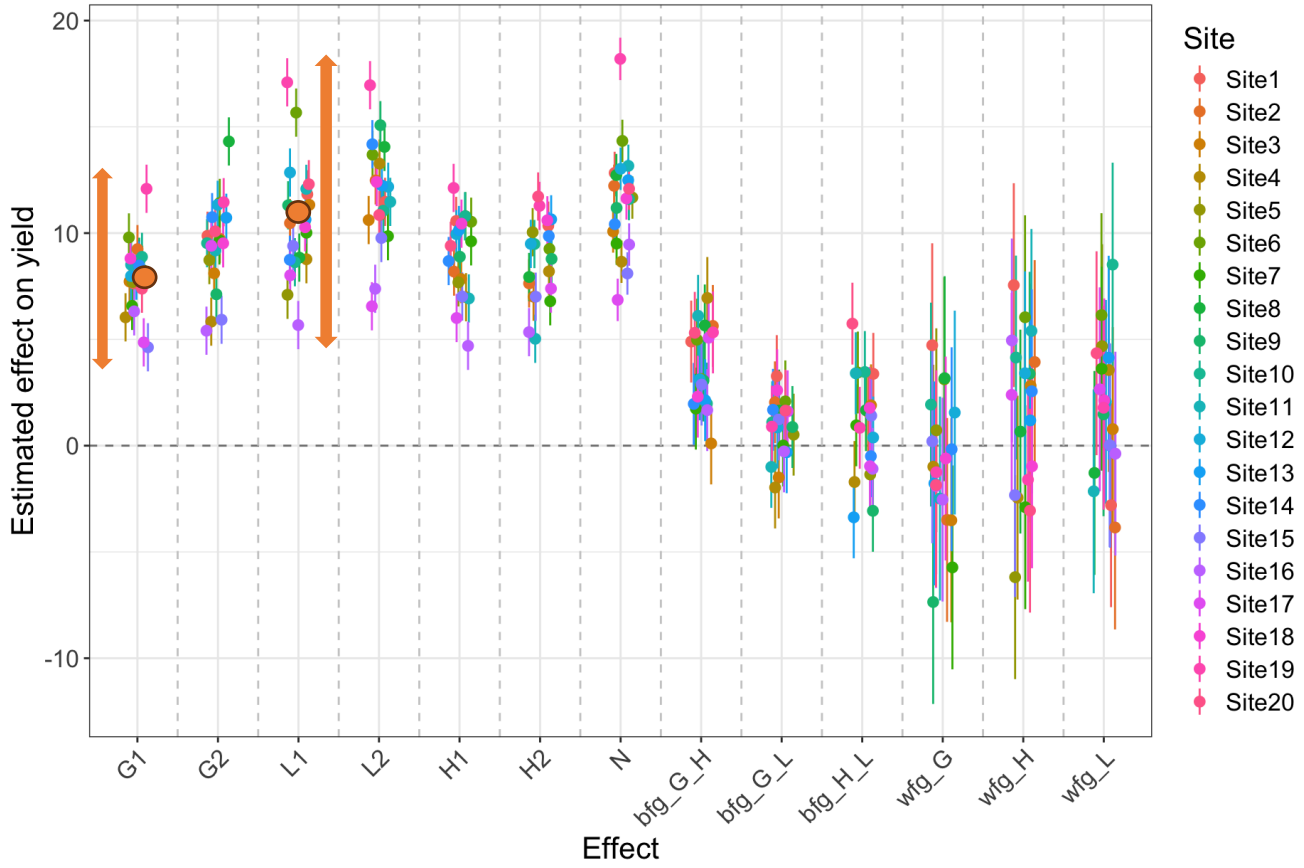
(B) Sites interact with all effects




```
> MuMIn::AICc(lm_FG.A, lm_FG.B)
         df     AICc
lm_FG.A  33 3939.414
lm_FG.B 261 3615.251
```

This improves flexibility, but may overfit the data.

From site-specific fixed effects to random coefficients



Fixed effects: **260 parameters**

Site-specific fixed-effects models become parameter-heavy and difficult to generalise conclusions.

Use random coefficients

1. Overall mean effect plus site variation supports broader multi-site inference with fewer parameter required.
2. Experimental design, assuming sites selected random from all possible sites.
3. Provide generalised inference for new sites.

DI model with random coefficients equation

Identity effects | Interaction effects | Additional treatments

$$y_{kl} = \sum_{i=1}^{6}(\beta_i + b_{ik})P_{ikl} + \sum_{i<j}(\delta_{ij}+d_{ijk})(P_{ikl}P_{jkl})^{\theta} + (\alpha + a_k)X_{kl} + \epsilon_{kl}$$

$(b_{1k} \cdots b_{6k}\ d_{12k} \cdots d_{56k}\ \alpha_k)^T \sim MVN(\mathbf{0}, \mathbf{G})$, and $\epsilon_{kl} \sim N(0, \sigma_k^2)$

$$(b_{1k} \cdots b_{6k}\ d_{12k} \cdots d_{56k}\ \alpha_k)^T = \begin{bmatrix} b_{1k} \\ \vdots \\ b_{6k} \\ d_{12k} \\ \vdots \\ d_{56k} \\ \alpha_k \end{bmatrix}_{22\times 1}$$

$$\mathbf{G}_{22\times 22} = \begin{bmatrix} \sigma_{b1}^2 & \cdots & \sigma_{b1b6} & \sigma_{b1d12} & \cdots & \sigma_{b1d56} & \sigma_{b1\alpha} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{b1b6} & \cdots & \sigma_{b6}^2 & \sigma_{b6d12} & \vdots & \sigma_{b6d56} & \sigma_{b6\alpha} \\ \sigma_{b1d12} & \cdots & \sigma_{b6d12} & \sigma_{d12}^2 & \vdots & \sigma_{d12d56} & \sigma_{d12\alpha} \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{b1d56} & \cdots & \sigma_{b6d56} & \sigma_{d12d56} & \vdots & \sigma_{d56}^2 & \sigma_{d56\alpha} \\ \sigma_{b1\alpha} & \cdots & \sigma_{b6\alpha} & \sigma_{d12\alpha} & \cdots & \sigma_{d56\alpha} & \sigma_{\alpha}^2 \end{bmatrix}$$

• DI model with random coefficients

- Fixed effects = overall biodiversity pattern
- Random effects = site-to-site deviations
- Residual structure = within-site variability

Alternative random coefficients include:

1. Variance-covariance components and structures (Diagonal, Unstructured, etc.).

2. Residual variances can be assumed to be consistent or vary by sites.

Example (AV model with diagonal var-cov random-effects structure, allowing sites have different residual variances):

```
lme(
Yield ~ 0 + G1 + G2 + L1 + L2 + H1 + H2 + N + AV,
random = list(site = pdDiag(~ 0 + G1 + G2 + L1 +
L2 + H1 + H2 + N + AV)),
weight = varIdent(form = ~ 1 | site),
data = data,
method = "REML",
control = lmeControl(opt = "optim"))
```

Variance-covariance matrix

Diagonal matrix:
(pdDiag)

$$\mathbf{G} = \begin{bmatrix} \sigma_{b1}^2 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \cdots & \sigma_{b6}^2 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \sigma_{d12}^2 & \vdots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \sigma_{d56}^2 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \sigma_{\alpha}^2 \end{bmatrix}$$

Compound symmetry matrix:
(pdCompSymm)

$$\mathbf{G} = \begin{bmatrix} \sigma_b^2 & \sigma_d & \cdots & \cdots & \cdots & \cdots & \sigma_d \\ \sigma_d & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \cdots & \sigma_b^2 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \sigma_b^2 & \vdots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \sigma_b^2 & \sigma_d \\ \sigma_d & \cdots & \cdots & \cdots & \cdots & \sigma_d & \sigma_b^2 \end{bmatrix}$$

Unstructured:
(pdSymm)

$$\mathbf{G} = \begin{bmatrix} \sigma_{b1}^2 & \cdots & \sigma_{b1b6} & \sigma_{b1d12} & \cdots & \sigma_{b1d56} & \sigma_{b1\alpha} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{b1b6} & \cdots & \sigma_{b6}^2 & \sigma_{b6d12} & \vdots & \sigma_{b6d56} & \sigma_{b6\alpha} \\ \sigma_{b1d12} & \cdots & \sigma_{b6d12} & \sigma_{d12}^2 & \vdots & \sigma_{d12d56} & \sigma_{d12\alpha} \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{b1d56} & \cdots & \sigma_{b6d56} & \sigma_{d12d56} & \vdots & \sigma_{d56}^2 & \sigma_{d56\alpha} \\ \sigma_{b1\alpha} & \cdots & \sigma_{b6\alpha} & \sigma_{d12\alpha} & \cdots & \sigma_{d56\alpha} & \sigma_{\alpha}^2 \end{bmatrix}$$

DI model with random coefficients

Follow the steps in SS8. Lab sheet 2(a)

Example (AV model with diagoal var-cov random effect structure, allowing sites have different residual variances):

```
lme(
Yield ~ 0 + G1 + G2 + L1 + L2 + H1 + H2 + N + AV,
random = list(site = pdDiag(~ 0 + G1 + G2 + L1 +
L2 + H1 + H2 + N + AV)),
weight = varIdent(form = ~ 1 | site),
data = data,
method = "REML",
control = lmeControl(opt = "optim"))
```

Note: Fit mixed models with the consistent fixed effects using REML and compare alternative random-effects structures using AICc.

Variance-covariance matrix:

1. Diagonal matrix (`pdDiag`)
2. Compound symmetry (`pdCompSymm`)
3. Unstructured matrix (`pdSymm`)

⚠ Warning: Unstructured matrix might take a long time to run

DI model with random coefficients

Re-check fixed effects in the random-coefficients model

The preferred DI structure may change once between-site random coefficients are modelled explicitly.

Example (AV model with diagonal var-cov random effect structure, allowing sites have different residual variances):

```
lme(
Yield ~ 0 + G1 + G2 + L1 + L2 + H1 + H2 + N + AV,
random = list(site = pdDiag(~ 0 + G1 + G2 + L1 +
L2 + H1 + H2 + N + AV)),
weight = varIdent(form = ~ 1 | site),
data = data,
method = "ML",
control = lmeControl(opt = "optim"))
```

DI model with random coefficients

Follow the steps in SS8. Lab sheet 2(b)

Example (AV model with diagonal var-cov random effect structure, allowing sites have different residual variances):

```
lme(
Yield ~ 0 + G1 + G2 + L1 + L2 + H1 + H2 + N + AV,
random = list(site = pdDiag(~ 0 + G1 + G2 + L1 + L2
+ H1 + H2 + N + AV)),
weight = varIdent(form = ~ 1 | site),
data = data,
method = "ML",
control = lmeControl(opt = "optim"))
```

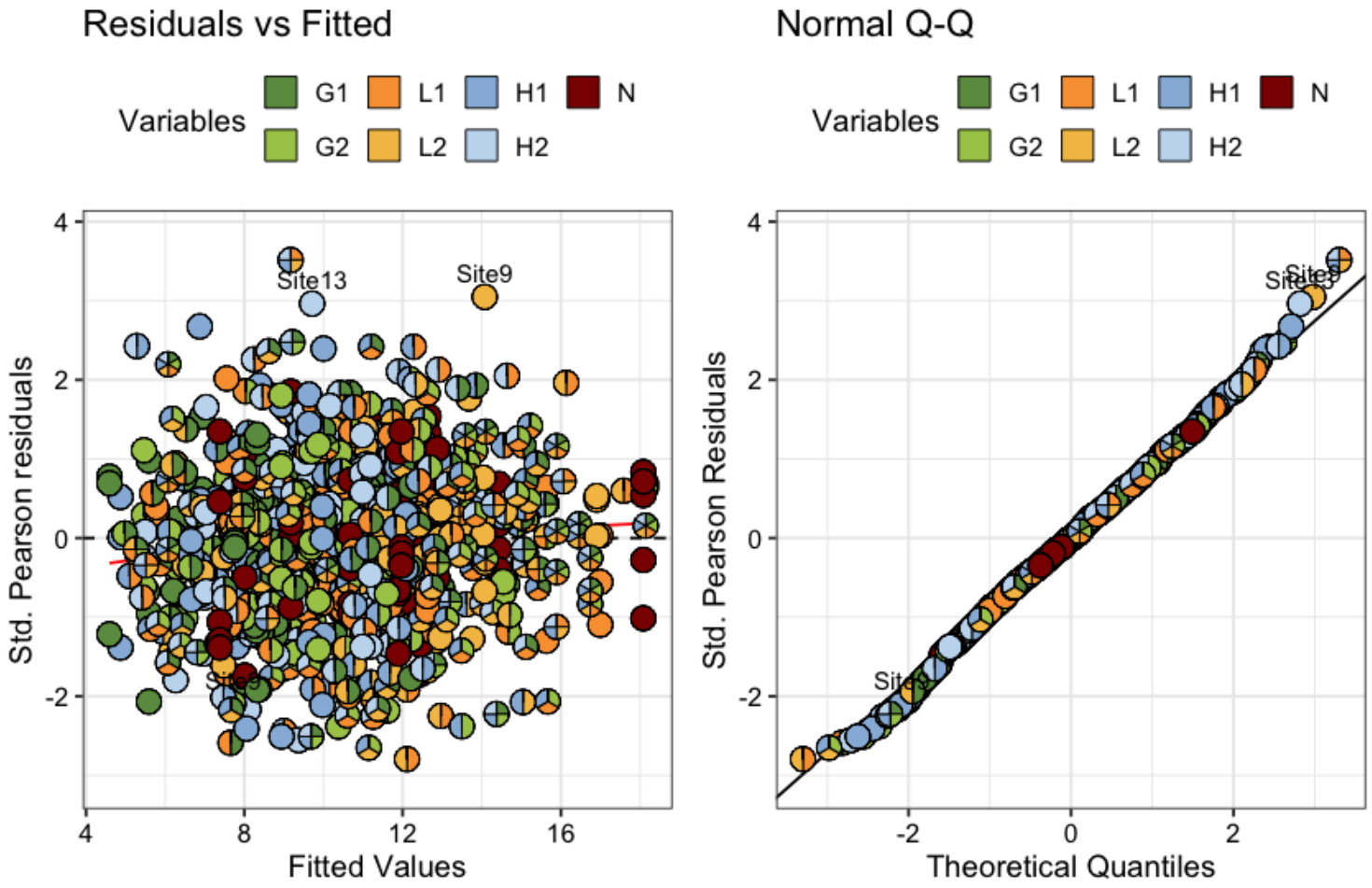
Note: Fit mixed models with the consistent random effects using ML and compare alternative fixed effects using AICc.

Model diagnostics

The final selected model should be refit by REML before diagnostics and interpretation.
Diagnostics are part of modelling, not optional.

Selected model example (FG model with unstructured random effect structure, allowing sites have different residual variances):

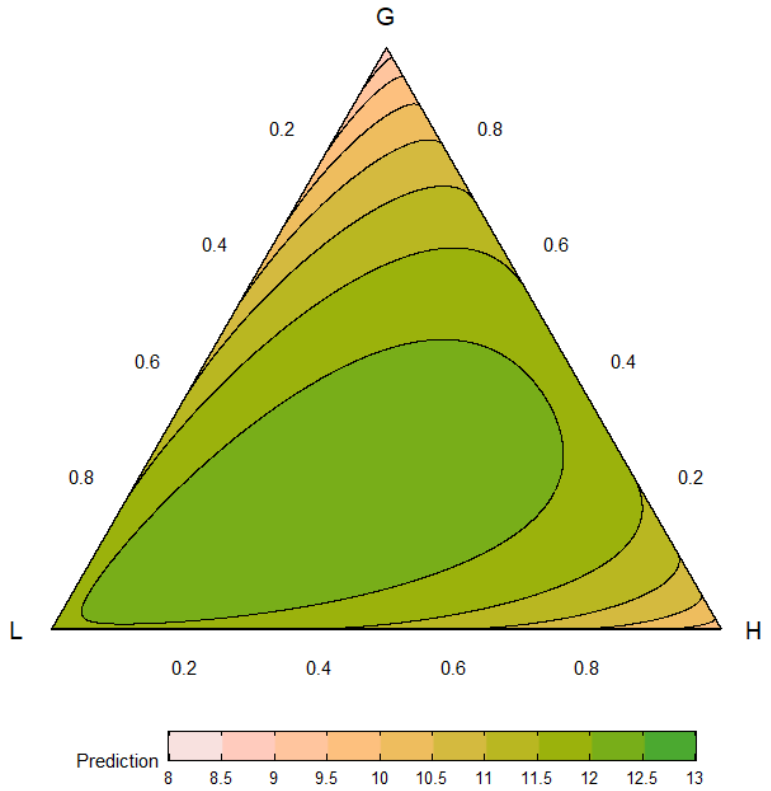
```
lme(Yield ~ 0 + G1 + G2 + L1 + L2 +
H1 + H2 + N + bfg_G_H + bfg_G_L +
bfg_H_L + wfg_G + wfg_H + wfg_L,
random = list(site = pdSymm(~ 0 + G1
+ G2 + L1 + L2 + H1 + H2 + N + AV)),
weight = varIdent(form = ~ 1 | site),
data = data,
method = "REML",
control = lmeControl(opt = "optim"))
```

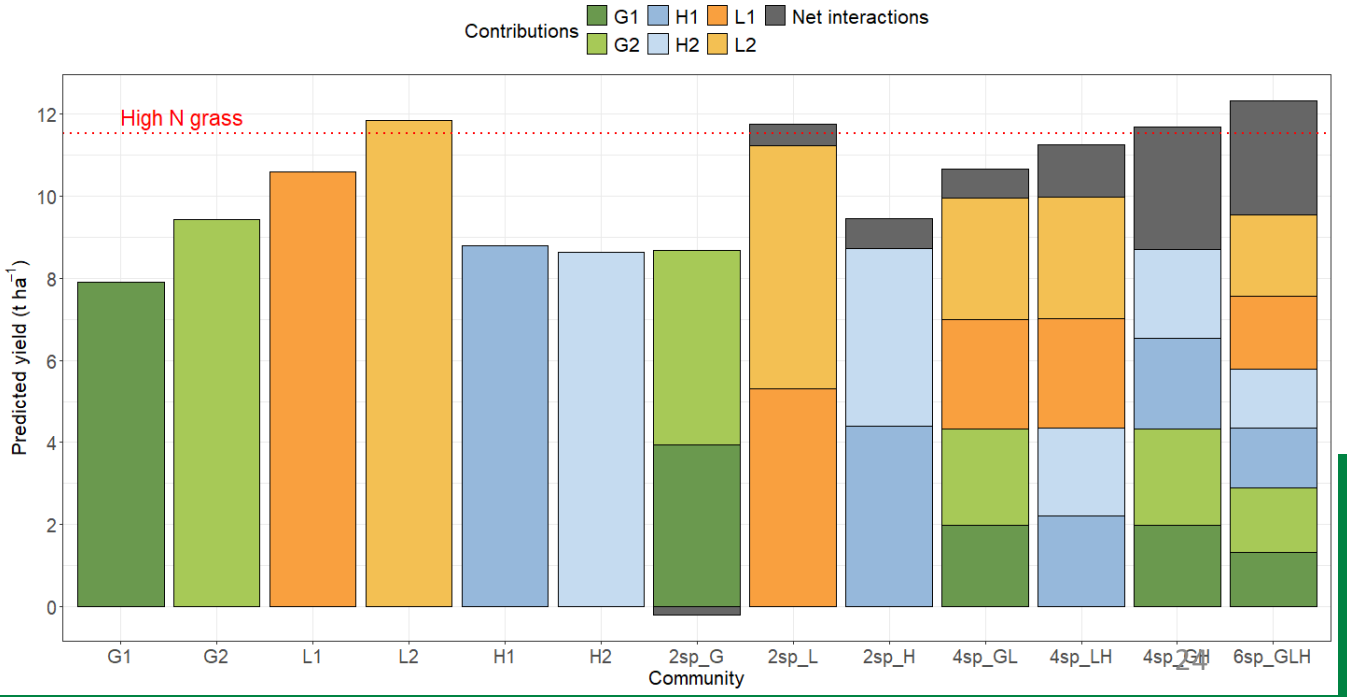


Model predictions

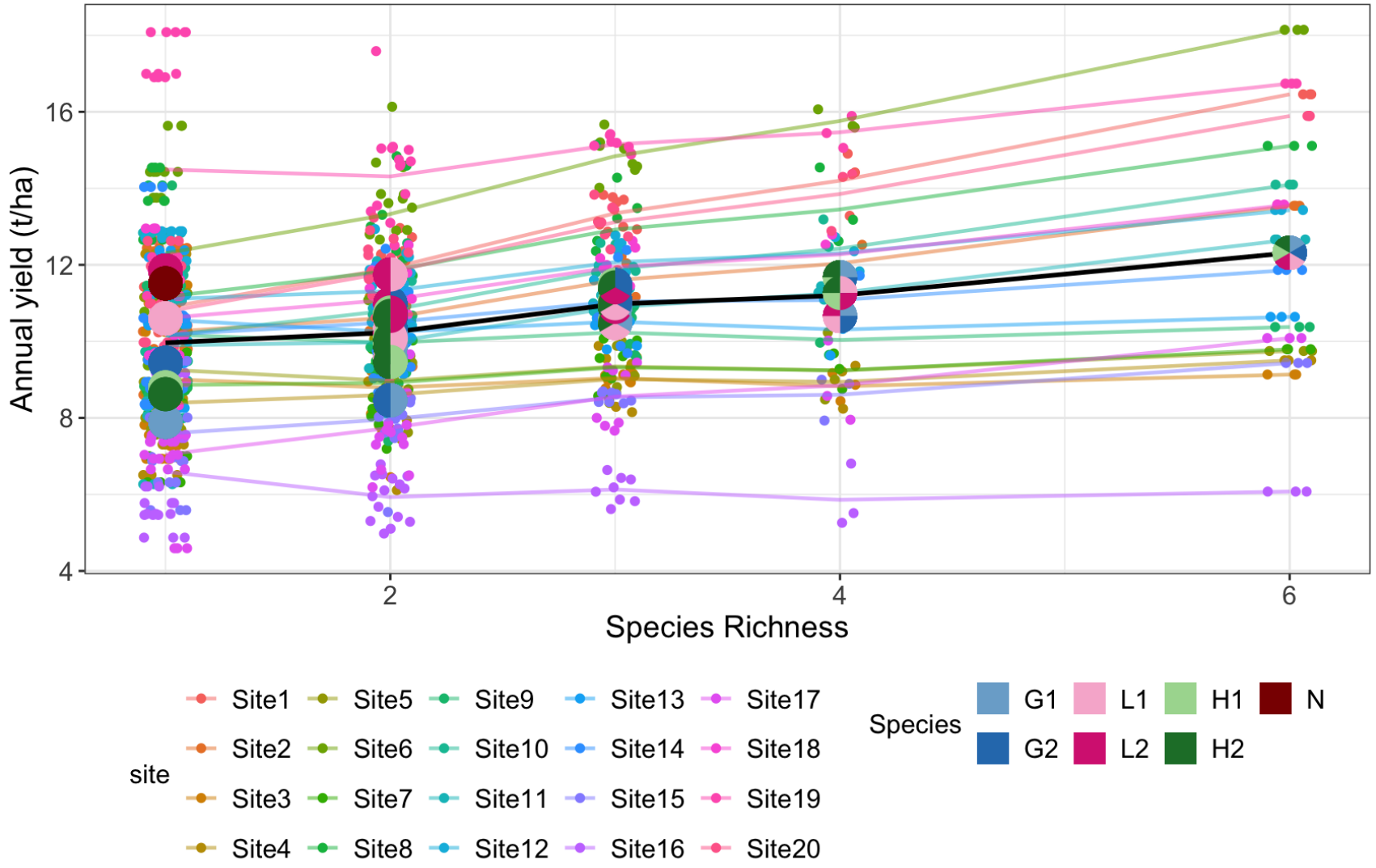
Predictions represent the average responses across all sites after accounting for site-level variation.

	Value	Std.Error	DF	t-value	p-value
G1	7.895910	0.3801676	1008	20.769547	0.0000
G2	9.436382	0.4954771	1008	19.045044	0.0000
L1	10.598213	0.6186187	1008	17.132061	0.0000
L2	11.849226	0.5326693	1008	22.244995	0.0000
H1	8.797856	0.4015156	1008	21.911616	0.0000
H2	8.621171	0.3918812	1008	21.999454	0.0000
N	11.517487	0.5534083	1008	20.811917	0.0000
bfg_G_H	3.670999	0.3276398	1008	11.204375	0.0000
bfg_G_L	0.776919	0.3276398	1008	2.371260	0.0179
bfg_H_L	0.995685	0.3276398	1008	3.038963	0.0024
wfg_G	-0.477819	0.5544516	1008	-0.861787	0.3890
wfg_H	1.743905	0.5544516	1008	3.145279	0.0017
wfg_L	1.205367	0.5544516	1008	2.173981	0.0299





Model predictions



By modelling site-level variability explicitly, mixed-effects models improve generalisability and reduce overfitting compared with site-specific fixed-effects models.

Interpreting and reporting the final model

Random standard deviation between sites:

```
Random effects:
 Formula: ~0 + G1 + G2 + L1 + L2 + H1 + H2 + N + AV | site
 Structure: General positive-definite
          StdDev   Corr
G1        1.611748 G1    G2    L1    L2    H1    H2    N
G2        2.138205 0.577
L1        2.703515 0.680 0.405
L2        2.309941 0.883 0.547 0.617
H1        1.709974 0.766 0.813 0.754 0.782
H2        1.653771 0.602 0.491 0.537 0.642 0.766
N         2.421736 0.949 0.632 0.773 0.747 0.817 0.585
AV        1.161684 0.309 0.525 0.447 0.181 0.557 0.389 0.435
Residual  1.132112
```

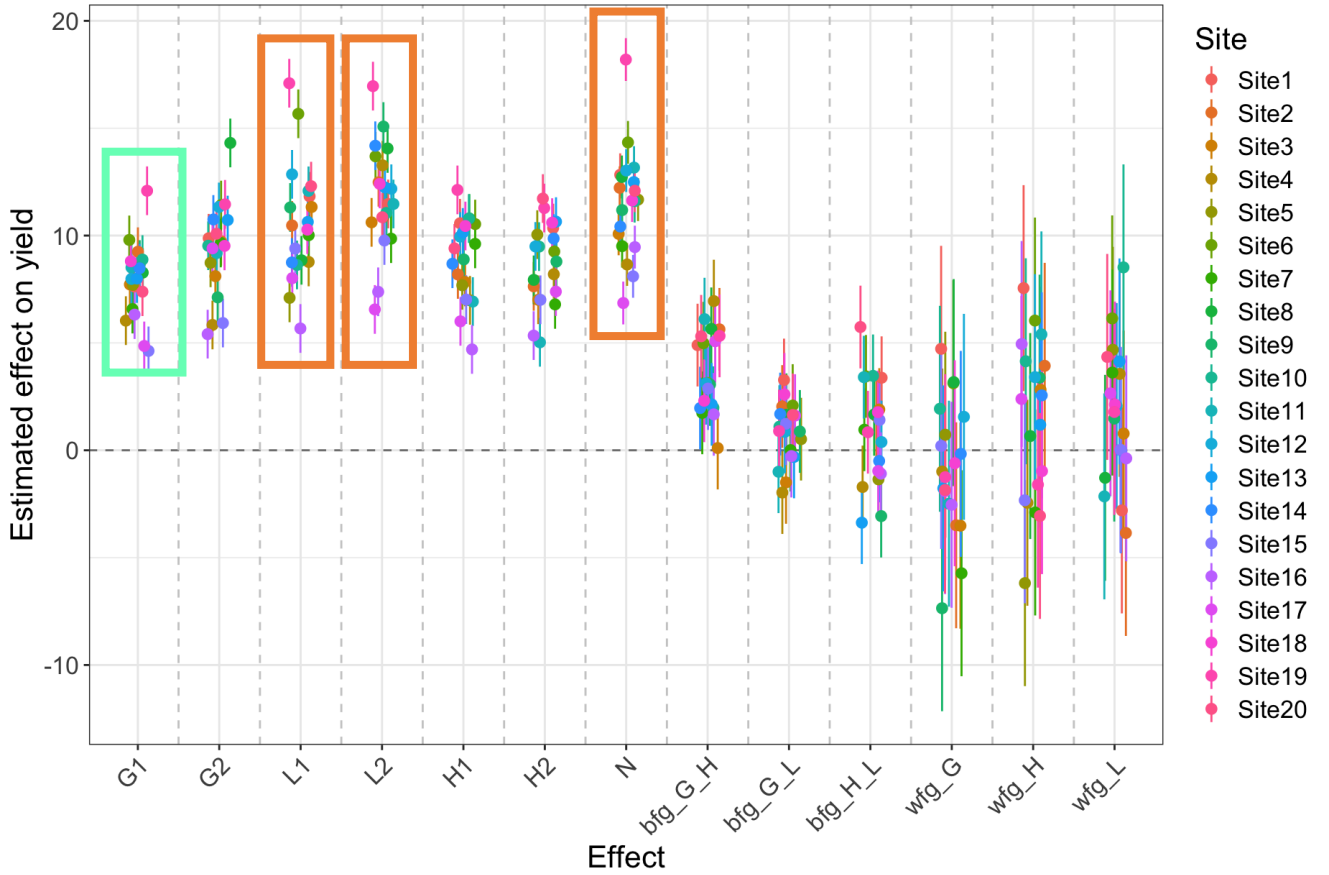
Residual variance within each site:

```
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | site
 Parameter estimates:
    Site1     Site2     Site3     Site4     Site5     Site6     Site7     Site8     Site9    Site10
1.0000000 0.7985530 0.7076971 1.0692834 1.2084447 0.6416359 1.1438282 0.6298150 1.7671241 1.3301604
   Site11    Site12    Site13    Site14    Site15    Site16    Site17    Site18    Site19    Site20
1.0620777 0.6964363 1.3295865 0.8151904 1.0806133 0.8273090 1.1910730 0.9764695 0.6308598 1.3876942
```

Interpreting and reporting the final model

```
> getVarCov(lme_FG.3.Un.AV.heter, type = "random.effects")
Random effects variance covariance matrix
       G1     G2     L1      L2     H1     H2     N      AV
G1 2.5977 1.9888 2.9627 3.28600 2.1110 1.6042 3.7030 0.57870
G2 1.9888 4.5719 2.3426 2.70130 2.9711 1.7358 3.2739 1.30420
L1 2.9627 2.3426 7.3090 3.85570 3.4851 2.4016 5.0585 1.40450
L2 3.2860 2.7013 3.8557 5.33580 3.0900 2.4531 4.1766 0.48571
H1 2.1110 2.9711 3.4851 3.09000 2.9240 2.1666 3.3850 1.10600
H2 1.6042 1.7358 2.4016 2.45310 2.1666 2.7350 2.3420 0.74760
N  3.7030 3.2739 5.0585 4.17660 3.3850 2.3420 5.8648 1.22470
AV 0.5787 1.3042 1.4045 0.48571 1.1060 0.7476 1.2247 1.34950
```

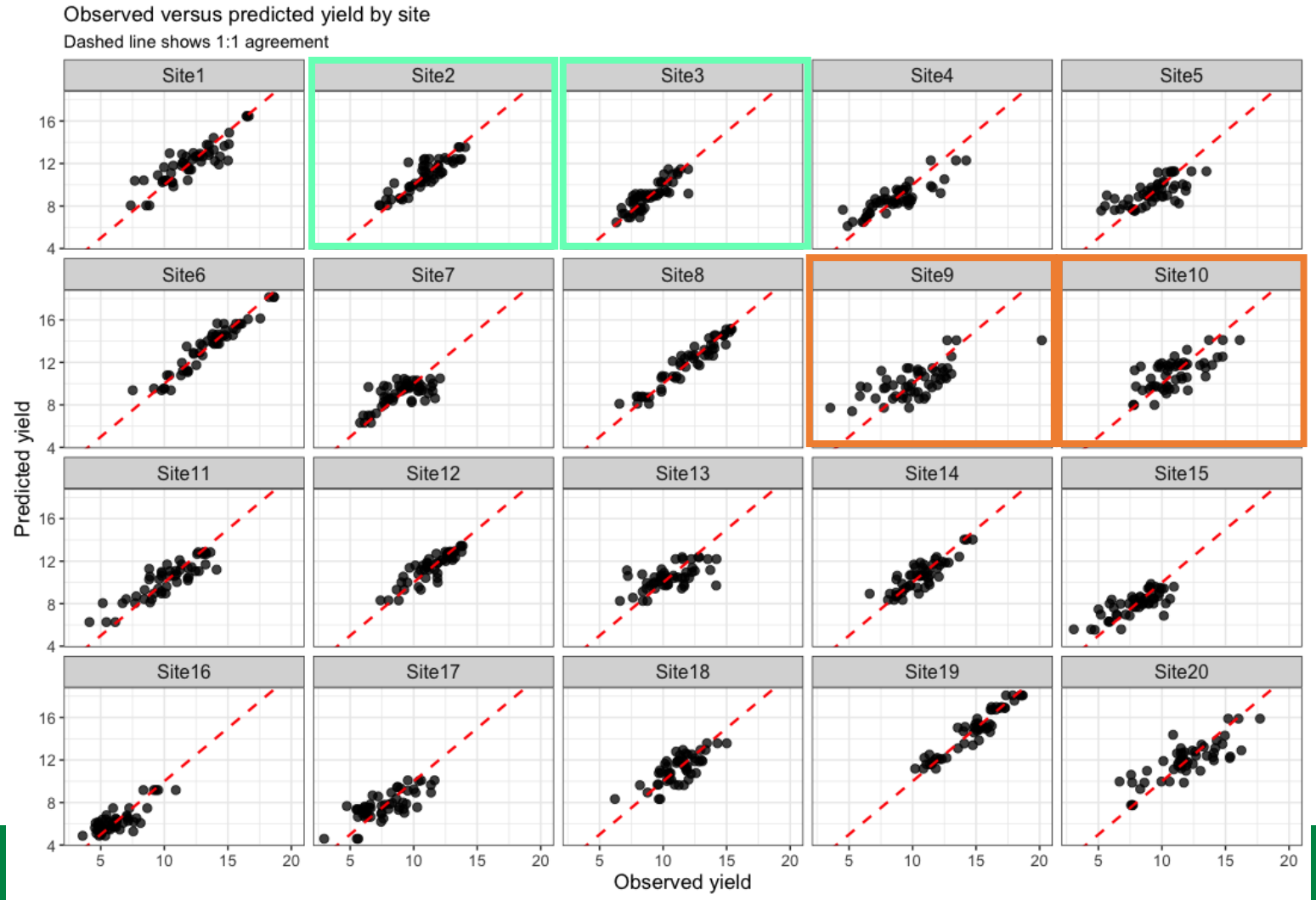
Larger variances correspond to wider effects variation across sites in fixed-effects models. E.g. L1, L2, and N.



Interpreting and reporting the final model

Residual variance within each site:

site	resid_sd_ratio	resid_variance
Site1	1.0000000	1.2816774
Site2	0.7985530	0.8173088
Site3	0.7076971	0.6419091
Site4	1.0692834	1.4654276
Site5	1.2084447	1.8716829
Site6	0.6416359	0.5276622
Site7	1.1438282	1.6768735
Site8	0.6298150	0.5083991
Site9	1.7671241	4.0023291
Site10	1.3301604	2.2677062
Site11	1.0620777	1.4457436
Site12	0.6964363	0.6216436
Site13	1.3295865	2.2657495
Site14	0.8151904	0.8517199
Site15	1.0806133	1.4966468
Site16	0.8273090	0.8772314
Site17	1.1910730	1.8182580
Site18	0.9764695	1.2220700
Site19	0.6308598	0.5100872
Site20	1.3876942	2.4681200



Sites with points tightly clustered around the red line should have **small residual variance.**

Sites with points widely scattered around the red line should have **large residual variance.**

SS8. Lab sheet

2. Identify the best mixed-effects model

(a) Compare random-effects and residual variance structures
(b) Refit and compare fixed effects in mixed models
(c) Model diagnostics
(d) Interpret results and visualisation