Last week

- Why do we need mixed models?
- What are fixed vs. random effects?
- Think about how fixed/random effects relate to experimental design
- Practice identifying fixed vs. random effects
- Understand the architecture of multilevel/mixed models

This week

- How to implement linear mixed models in R
  - Different packages
  - How to specify the models and random effects structure (syntax)
- Hypothesis testing in the mixed model framework
- Midterm survey
- Note about readings and additional resources this week

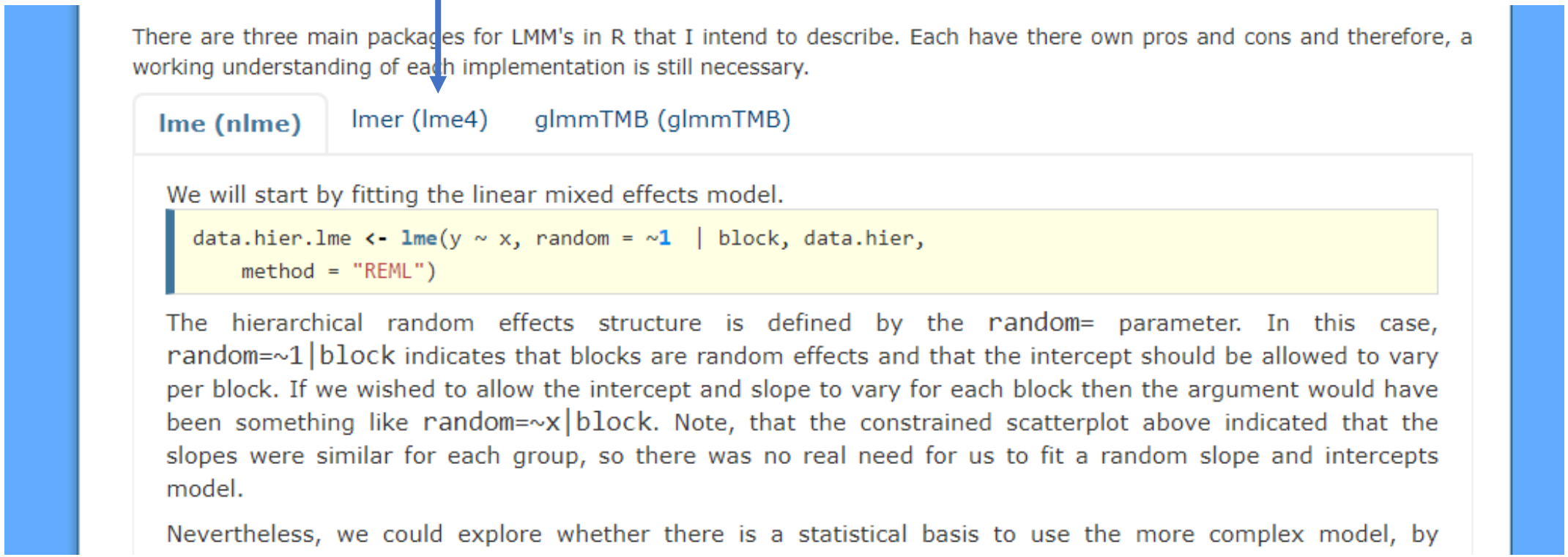
Two most common packages for implementing mixed models in R

	nlme	**lme4**
REML	Y	Y
Optimizer	nlminb (default) or optim (BFGS or L-BFGS-B)	bobyqa (default) or Nelder_Mead
Spatial/temporal autocorrelation	Y	N
Variance structures	Y	N
Crossed random effects	N	Y
Frequentist hypothesis testing	Built-in	Need to use in conjunction with lmerTest package
Speed	Slow	Fast

Adapted from Murray

We'll focus on using lmer() from the lme4 package.

The Murray tutorial linked this week has convenient tabs that demonstrate how the same things work using three different packages in R.



There are three main packages for LMM's in R that I intend to describe. Each have there own pros and cons and therefore, a working understanding of each implementation is still necessary.

lme (nlme) lmer (lme4) glmmTMB (glmmTMB)

We will start by fitting the linear mixed effects model.

```
data.hier.lme <- lme(y ~ x, random = ~1  | block, data.hier,
    method = "REML")
```

The hierarchical random effects structure is defined by the `random=` parameter. In this case, `random=~1|block` indicates that blocks are random effects and that the intercept should be allowed to vary per block. If we wished to allow the intercept and slope to vary for each block then the argument would have been something like `random=~x|block`. Note, that the constrained scatterplot above indicated that the slopes were similar for each group, so there was no real need for us to fit a random slope and intercepts model.

Nevertheless, we could explore whether there is a statistical basis to use the more complex model, by

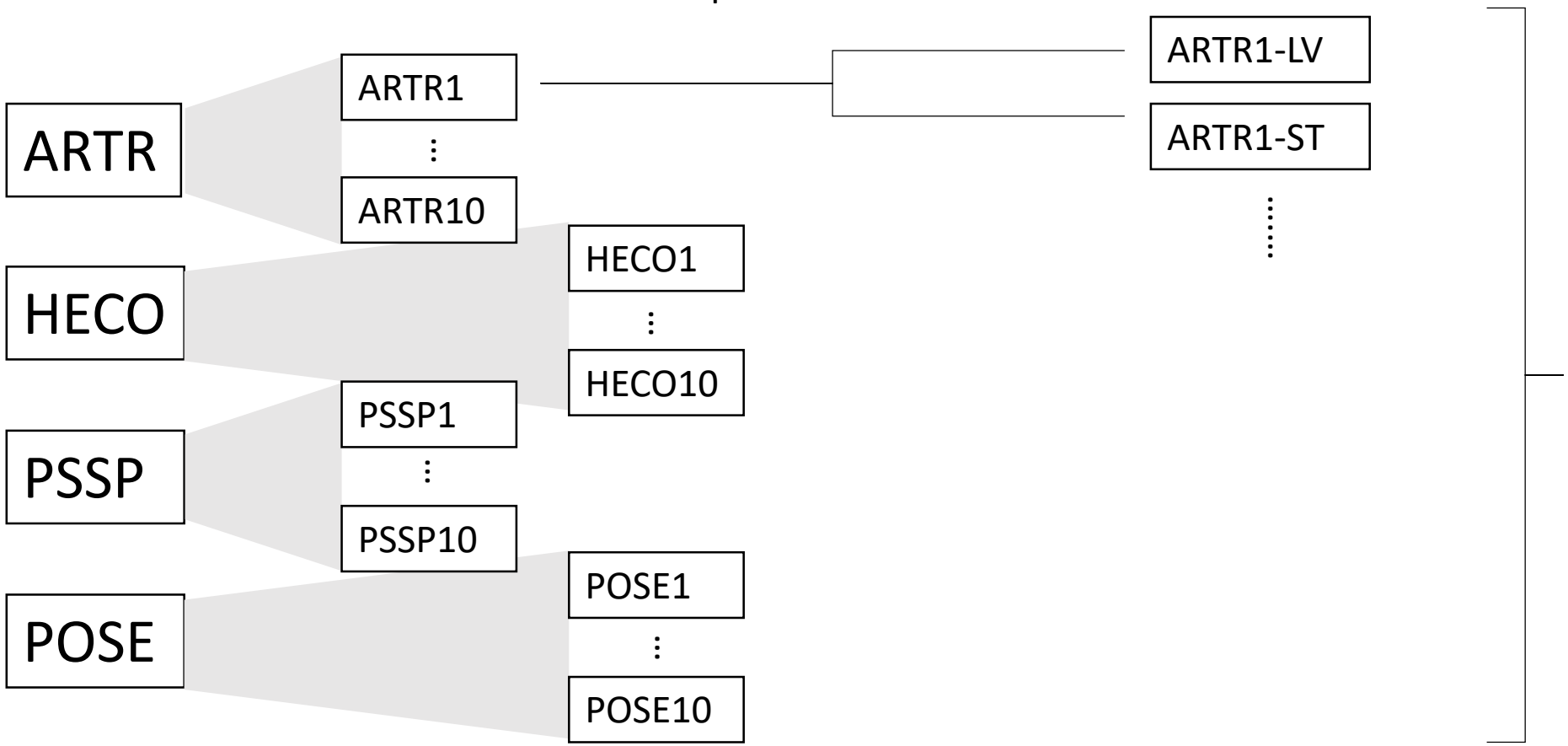
A simple example: Idaho soil chemistry data

Research Q: What is the effect of soil source and soil sterilization on soil chemistry?

4 different plant species as soil source levels

Collect 10 soil samples associated with each species

Each soil sample is split in half, and one half is sterilized



In total, 4*10*2 = 80 soil samples analyzed for a suite of soil chemistry response variables

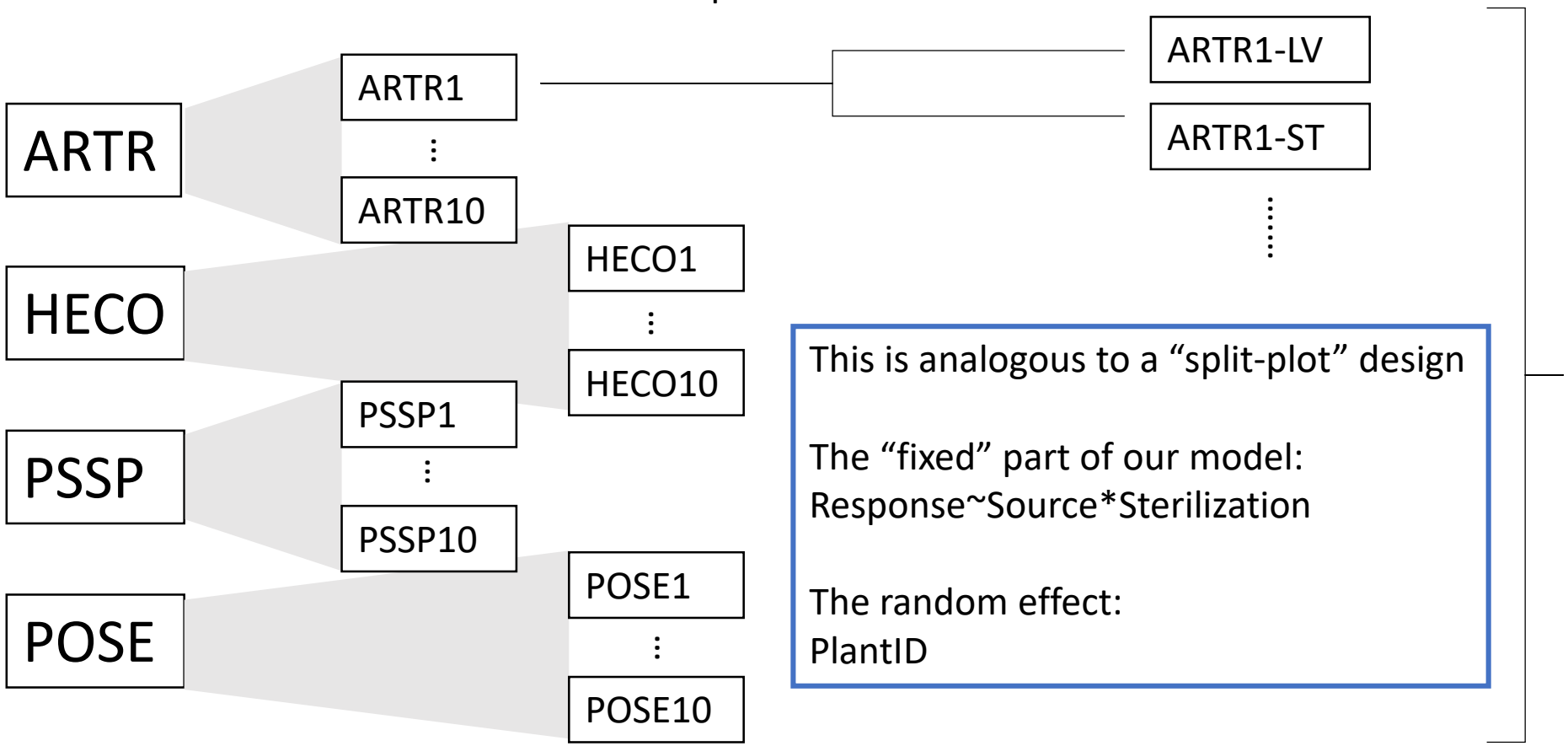
A simple example: Idaho soil chemistry data

Research Q: What is the effect of soil source and soil sterilization on soil chemistry?

4 different plant species as soil source levels

Collect 10 soil samples associated with each species

Each soil sample is split in half, and one half is sterilized



This is analogous to a "split-plot" design

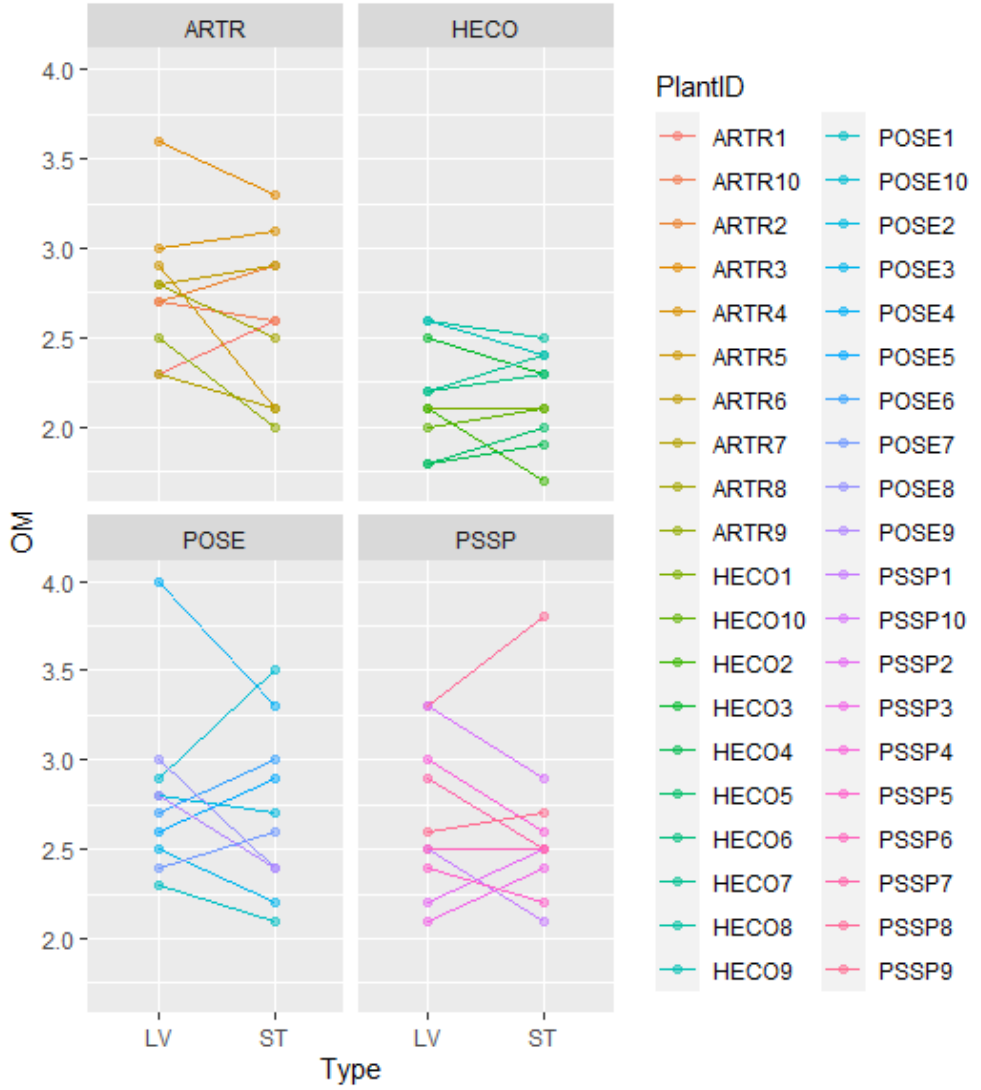
The "fixed" part of our model:
Response~Source*Sterilization

The random effect:
PlantID

In total, 4*10*2 = 80 soil samples analyzed for a suite of soil chemistry response variables

Example 1: What is the effect of soil source and soil sterilization on organic matter (OM)?

- What happens if we fit only a fixed effects model: Response~Source*Sterilization
- Random intercepts?
- Random slopes?
- How to specify random effects?



Basic syntax: function lmer() in the lme4 package

- General form: dependent ~ independent|grouping
  - grouping: usually the random effect

- Usually it is helpful to write the fixed part and the random part separately (put it in parentheses), so we can have transparency and flexibility

- Three basic variants:
  - Intercepts only by random factor: (1 | random.factor)
  - Slopes only by random factor: (0 + fixed.factor | random.factor)
  - Intercepts and slopes by random factor: (1 + fixed.factor | random.factor)

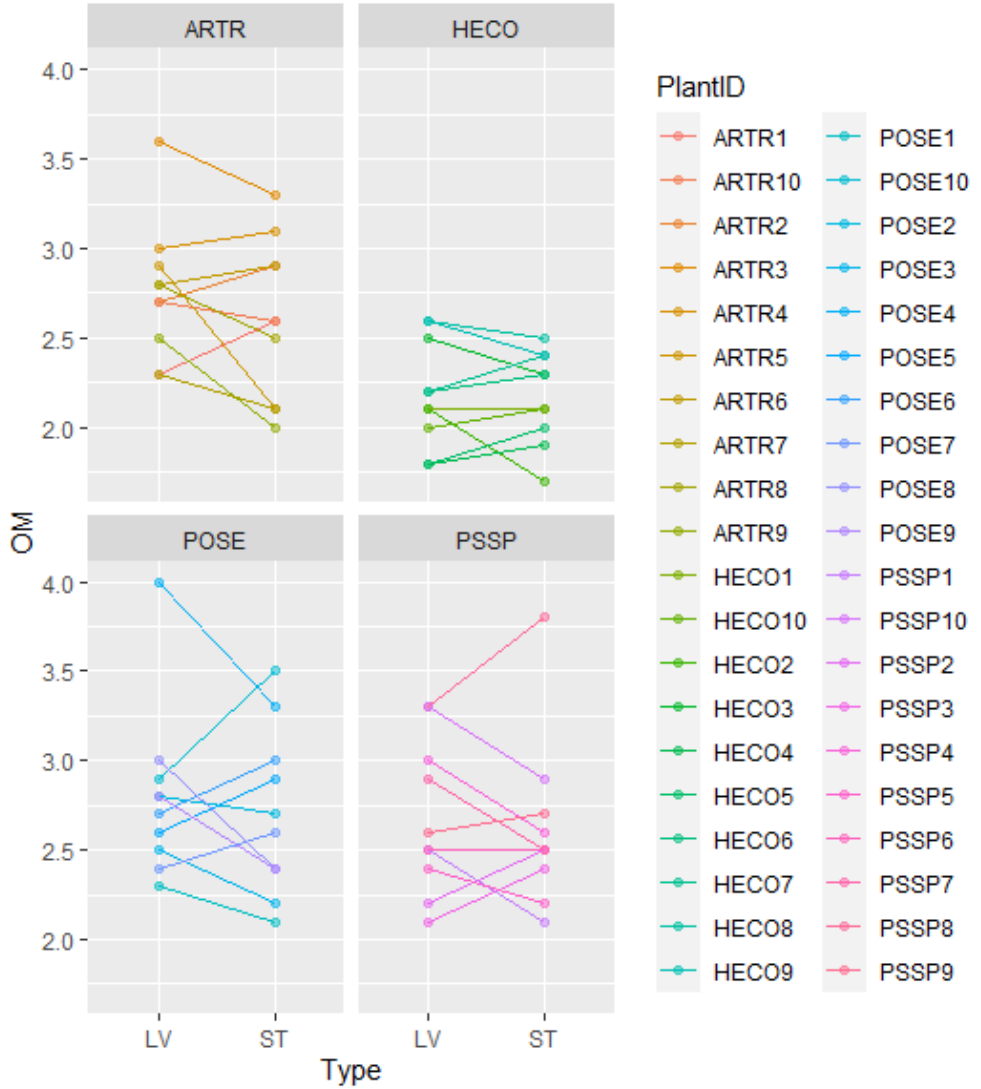
Example 1: What is the effect of soil source and soil sterilization on organic matter (OM)?

lmer() syntax dependent independent random

- Fixed effects only: OM~Source*Sterilization
- Random intercepts only: OM~Source*Sterilization+(1|PlantID)
- Random slopes:

  OM~Source*Sterilization+(0+Sterilization|PlantID)
- Random intercepts and slopes:

  OM~ Source*Sterilization +(1+Sterilization|PlantID)



```
> summary(m.OM)

Call:
lm(formula = OM ~ Type * Source, data = soil)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6100 -0.2725 -0.0600  0.2300  1.2000

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.7600     0.1294  21.322  < 2e-16 ***
TypeST             -0.1500     0.1831  -0.819  0.41526
SourceHECO         -0.5700     0.1831  -3.114  0.00265 **
SourcePOSE          0.0400     0.1831   0.219  0.82765
SourcePSSP         -0.0800     0.1831  -0.437  0.66341
TypeST:SourceHECO   0.1300     0.2589   0.502  0.61709
TypeST:SourcePOSE   0.0600     0.2589   0.232  0.81738
TypeST:SourcePSSP   0.0900     0.2589   0.348  0.72912
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 0.4093 on 72 degrees of freedom
Multiple R-squared:  0.2624,    Adjusted R-squared:  0.1907
F-statistic: 3.659 on 7 and 72 DF,  p-value: 0.001964
```

```
> summary(rI.OM)
Linear mixed model fit by REML. t-tests use Satterthwaite's method
Formula: OM ~ Type * Source + (1 | PlantID)
   Data: soil

REML criterion at convergence: 73.8

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.51596 -0.54637 -0.03082  0.49753  2.04514

Random effects:
 Groups   Name        Variance Std.Dev.
 PlantID  (Intercept) 0.11014  0.3319
 Residual             0.05742  0.2396
Number of obs: 80, groups:  PlantID, 40

Fixed effects:
                  Estimate Std. Error      df t value Pr(>|t|)
(Intercept)         2.7600     0.1294 50.2765  21.322  < 2e-16 ***
TypeST             -0.1500     0.1072 36.0000  -1.400  0.17014
SourceHECO         -0.5700     0.1831 50.2765  -3.114  0.00305 **
SourcePOSE          0.0400     0.1831 50.2765   0.219  0.82792
SourcePSSP         -0.0800     0.1831 50.2765  -0.437  0.66397
TypeST:SourceHECO   0.1300     0.1515 36.0000   0.858  0.39667
TypeST:SourcePOSE   0.0600     0.1515 36.0000   0.396  0.69450
TypeST:SourcePSSP   0.0900     0.1515 36.0000   0.594  0.55631
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) TypeST SrHECO SrPOSE SrPSSP TST:SH TST:SPO
TypeST      -0.414
SourceHECO  -0.707  0.293
SourcePOSE  -0.707  0.293  0.500
SourcePSSP  -0.707  0.293  0.500  0.500
TypST:SHECO  0.293 -0.707 -0.414 -0.207 -0.207
TypST:SPOSE  0.293 -0.707 -0.207 -0.414 -0.207  0.500
TypST:SPSSP  0.293 -0.707 -0.207 -0.207 -0.414  0.500  0.500
>
```

```
> summary(m.OM)

Call:
lm(formula = OM ~ Type * Source, data = soil)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6100 -0.2725 -0.0600  0.2300  1.2000

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.7600     0.1294  21.322  < 2e-16 ***
TypeST             -0.1500     0.1831  -0.819  0.41526
SourceHECO         -0.5700     0.1831  -3.114  0.00265 **
SourcePOSE          0.0400     0.1831   0.219  0.82765
SourcePSSP         -0.0800     0.1831  -0.437  0.66341
TypeST:SourceHECO   0.1300     0.2589   0.502  0.61709
TypeST:SourcePOSE   0.0600     0.2589   0.232  0.81738
TypeST:SourcePSSP   0.0900     0.2589   0.348  0.72912
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 0.4093 on 72 degrees of freedom
Multiple R-squared:  0.2624,    Adjusted R-squared:  0.1907
F-statistic: 3.659 on 7 and 72 DF,  p-value: 0.001964
```

- The PlantID random intercept effect explained quite a lot of variance compared to residual
- Fixed effects estimates stay pretty similar, but errors and p-values can change. In this case, some of the SE’s around the estimates decreased (more powerful test)

```
> summary(rI.OM)
Linear mixed model fit by REML. t-tests use Satterthwaite's method
Formula: OM ~ Type * Source + (1 | PlantID)
   Data: soil

REML criterion at convergence: 73.8

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.51596 -0.54637 -0.03082  0.49753  2.04514

Random effects:
 Groups   Name        Variance Std.Dev.
 PlantID  (Intercept) 0.11014  0.3319
 Residual             0.05742  0.2396
Number of obs: 80, groups:  PlantID, 40

Fixed effects:
                  Estimate Std. Error      df t value Pr(>|t|)
(Intercept)         2.7600     0.1294 50.2765  21.322  < 2e-16 ***
TypeST             -0.1500     0.1072 36.0000  -1.400  0.17014
SourceHECO         -0.5700     0.1831 50.2765  -3.114  0.00305 **
SourcePOSE          0.0400     0.1831 50.2765   0.219  0.82792
SourcePSSP         -0.0800     0.1831 50.2765  -0.437  0.66397
TypeST:SourceHECO   0.1300     0.1515 36.0000   0.858  0.39667
TypeST:SourcePOSE   0.0600     0.1515 36.0000   0.396  0.69450
TypeST:SourcePSSP   0.0900     0.1515 36.0000   0.594  0.55631

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) TypeST SrHECO SrPOSE SrPSSP TST:SH TST:SPO
TypeST      -0.414
SourceHECO  -0.707  0.293
SourcePOSE  -0.707  0.293  0.500
SourcePSSP  -0.707  0.293  0.500  0.500
TypST:SHECO  0.293 -0.707 -0.414 -0.207 -0.207
TypST:SPOSE  0.293 -0.707 -0.207 -0.414 -0.207  0.500
TypST:SPSSP  0.293 -0.707 -0.207 -0.207 -0.414  0.500  0.500
>
```

```
> Anova(m.OM)
Anova Table (Type II tests)

Response: OM
            Sum Sq Df F value    Pr(>F)
Type        0.1280  1  0.7639    0.3850
Source      4.1185  3  8.1933 9.168e-05 ***
Type:Source 0.0450  3  0.0895    0.9656
Residuals  12.0640 72
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Anova(rI.OM)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: OM
             Chisq Df Pr(>Chisq)
Type        2.2293  1   0.135414
Source     14.8310  3   0.001967 **
Type:Source 0.7837  3   0.853351
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

In this case, testing the ANOVA-type hypotheses on the fixed-effects only model (m.OM) and the mixed model (rI.OM) leads to similar conclusions

Use a similar type of Analysis of Deviance test (Wald chisquare) as with glm()

```
> Anova(rI.OM)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: OM
             Chisq Df Pr(>Chisq)
Type        2.2293  1   0.135414
Source     14.8310  3   0.001967 **
Type:Source 0.7837  3   0.853351
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

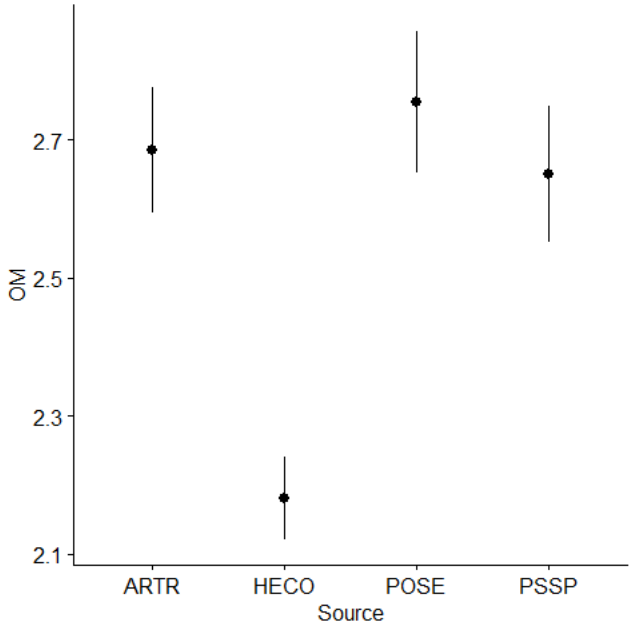
```
> emmeans(rI.OM,pairwise~Source, adjust="Tukey")
NOTE: Results may be misleading due to involvement in interactions
Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

$contrasts
contrast    estimate SE    df t.ratio p.value
ARTR - HECO 0.505    0.167 36 3.030   0.0223
ARTR - POSE -0.070   0.167 36 -0.420  0.9747
ARTR - PSSP 0.035    0.167 36 0.210   0.9967
HECO - POSE -0.575   0.167 36 -3.451  0.0075
HECO - PSSP -0.470   0.167 36 -2.820  0.0372
POSE - PSSP 0.105    0.167 36 0.630   0.9217

Results are averaged over the levels of: Type
 Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 4 estimates
```

We can similarly use emmeans() to do post-hoc pairwise comparisons of the significant fixed effect to understand which group level means are significantly different from each other

Example 1: What is the effect of soil source and soil sterilization on organic matter (OM)?

Analysis Methods

Soil organic matter content was analyzed using a linear mixed model which included: the fixed effects of plant species source, sterilization, and their interaction, and plant ID as a random intercept (function lmer() in the lme4 package, Bates et al. 2015). We performed *post-hoc* pairwise comparisons of group means using estimated marginal means, using the Tukey method to correct for multiple comparisons (function emmeans() in the emmeans package, Lenth 2020).



Example 1: What is the effect of soil source and soil sterilization on organic matter (OM)?

Results

Soil organic matter content was significantly different among plant species sources (X^2= 14.83, *df* = 3, *P* = 0.002), but not due to soil sterilization (X^2= 2.23, *df* = 1, *P* = 0.135), or its interaction with species source (X^2= 0.78, *df* = 3, *P* = 0.853). *Post-hoc* comparisons among species source levels revealed that soils from under HECO plants were significantly lower in organic matter compared to soils from all other species.

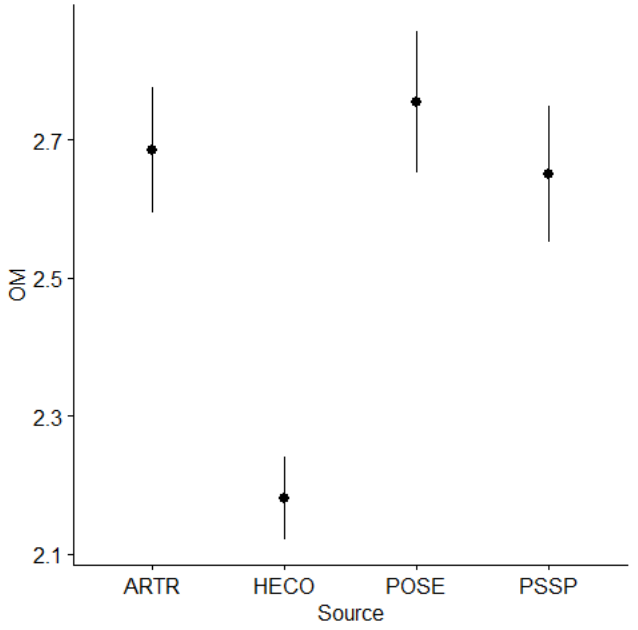


Fig. 1 Mean (SE) organic matter content in soils collected under each plant species.

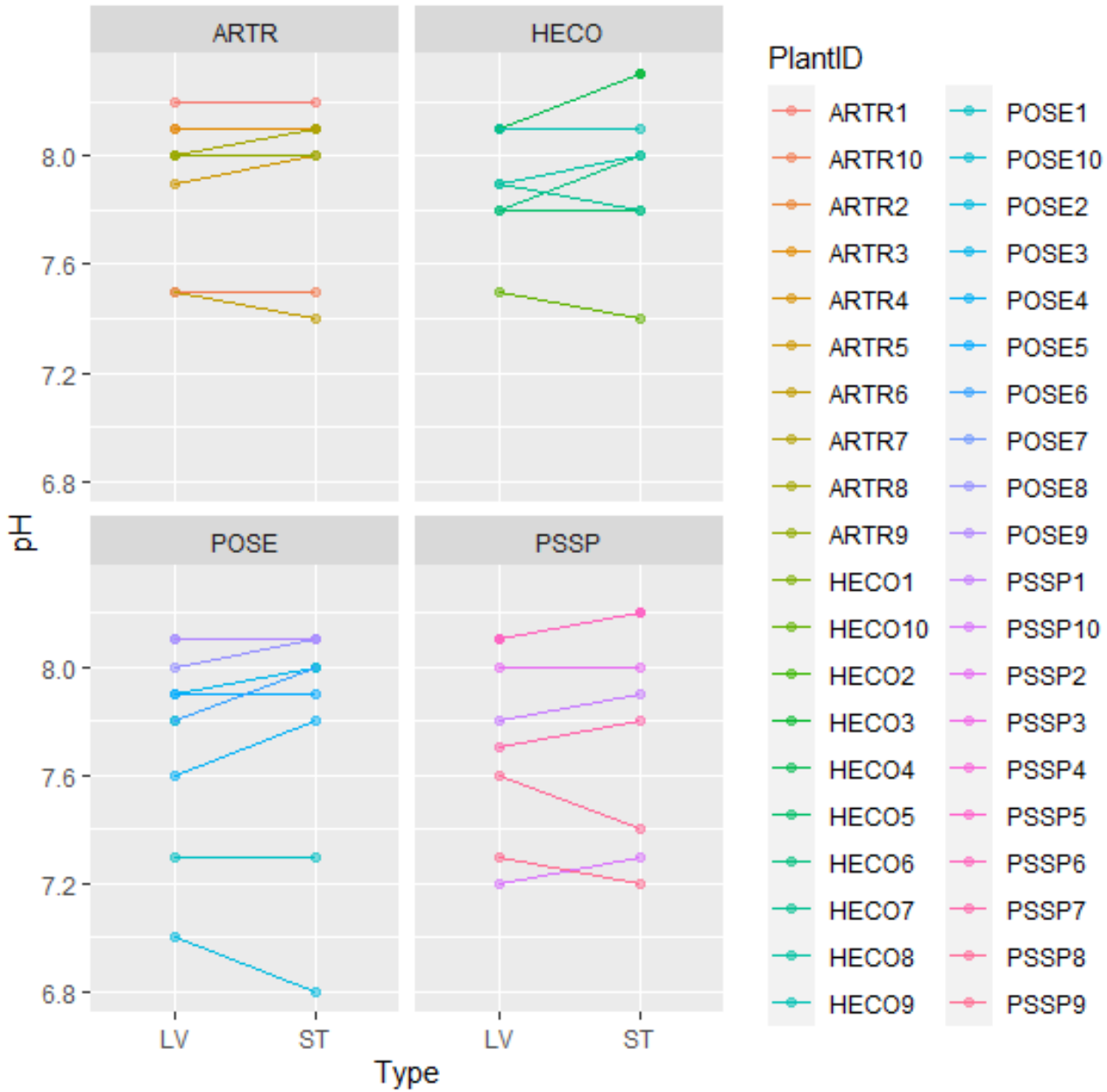
Example 2: What is the effect of soil source and soil sterilization on soil pH?

```
> summary(m.pH)

Call:
lm(formula = pH ~ Type * Source, data = soil)

Residuals:
   Min     1Q Median     3Q    Max
 -1.01  -0.14   0.08   0.21   0.36

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         7.9300     0.1032  76.842   <2e-16 ***
TypeST              0.0300     0.1459   0.206    0.838
SourceHECO          0.0100     0.1459   0.069    0.946
SourcePOSE         -0.1800     0.1459  -1.233    0.221
SourcePSSP         -0.1300     0.1459  -0.891    0.376
TypeST:SourceHECO   0.0600     0.2064   0.291    0.772
TypeST:SourcePOSE   0.0300     0.2064   0.145    0.885
TypeST:SourcePSSP   0.0100     0.2064   0.048    0.961
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3263 on 72 degrees of freedom
Multiple R-squared:  0.07787,   Adjusted R-squared:  -0.01179
F-statistic: 0.8685 on 7 and 72 DF,  p-value: 0.5356
```

```
> summary(rI.pH)
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: pH ~ Type * Source + (1 | PlantID)
   Data: soil

REML criterion at convergence: -17.6

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.99847 -0.40963 -0.02636  0.50575  1.41792

Random effects:
 Groups    Name        Variance Std.Dev.
 PlantID   (Intercept) 0.100417 0.3169
 Residual              0.006083 0.0780
Number of obs: 80, groups:  PlantID, 40

Fixed effects:
                  Estimate Std. Error       df t value Pr(>|t|)
(Intercept)        7.93000    0.10320 38.11497  76.842   <2e-16 ***
TypeST             0.03000    0.03488 36.00000   0.860    0.395
SourceHECO         0.01000    0.14594 38.11497   0.069    0.946
SourcePOSE        -0.18000    0.14594 38.11497  -1.233    0.225
SourcePSSP        -0.13000    0.14594 38.11497  -0.891    0.379
TypeST:SourceHECO  0.06000    0.04933 36.00000   1.216    0.232
TypeST:SourcePOSE  0.03000    0.04933 36.00000   0.608    0.547
TypeST:SourcePSSP  0.01000    0.04933 36.00000   0.203    0.840
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> Anova(m.pH)
Anova Table (Type II tests)

Response: pH
            Sum Sq Df F value Pr(>F)
Type        0.0605  1  0.5681 0.4535
Source      0.5765  3  1.8044 0.1540
Type:Source 0.0105  3  0.0329 0.9919
Residuals   7.6680 72
There were 12 warnings (use warnings() to see them)
> Anova(rI.pH)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: pH
             Chisq Df Pr(>Chisq)
Type        9.9452  1   0.001613 **
Source      2.7861  3   0.425786
Type:Source 1.7260  3   0.631163
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Only by including the random effect of PlantID do we reach the correct conclusion that sterilized soils have higher pH.


ARTR
HECO
POSE
PSSP
pH
8.0
7.6
7.2
6.8
LV
ST
Type
PlantID
ARTR1
ARTR10
ARTR2
ARTR3
ARTR4
ARTR5
ARTR6
ARTR7
ARTR8
ARTR9
HECO1
HECO10
HECO2
HECO3
HECO4
HECO5
HECO6
HECO7
HECO8
HECO9
POSE1
POSE10
POSE2
POSE3
POSE4
POSE5
POSE6
POSE7
POSE8
POSE9
PSSP1
PSSP10
PSSP2
PSSP3
PSSP4
PSSP5
PSSP6
PSSP7
PSSP8
PSSP9

break

Rstats tweet of the week

Allison Horst Retweeted

Mike Kearney 📊
@kearneymw

My accidental #rstats art is going up on the office wall! Big day!

2:38 PM · Mar 10, 2021 · Twitter for iPhone

25 Retweets **3** Quote Tweets **553** Likes

A quick overview of syntax extensions for lmer

formula	meaning
(1|group)	random group intercept
(x|group) = (1+x|group)	random slope of x within group with correlated intercept
(0+x|group) = (-1+x|group)	random slope of x within group: no variation in intercept
(1|group) + (0+x|group)	uncorrelated random intercept and random slope within group
(1|site/block) = (1|site)+(1|site:block)	intercept varying among sites and among blocks within sites (nested random effects)
site+(1|site:block)	*fixed* effect of sites plus random variation in intercept among blocks within sites
(x|site/block) = (x|site)+(x|site:block) = (1 + x|site)+(1+x|site:block)	slope and intercept varying among sites and among blocks within sites
(x1|site)+(x2|block)	two different effects, varying at different levels
x*site+(x|site:block)	fixed effect variation of slope and intercept varying among sites and random variation of slope and intercept among blocks within sites
(1|group1)+(1|group2)	intercept varying among crossed random effects (e.g. site, year)

(Bolker GLMM FAQ)

A quick overview of syntax extensions for lmer

formula	meaning
(1|group)	random group intercept
(x|group) = (1+x|group)	random slope of x within group with correlated intercept
(0+x|group) = (-1+x|group)	random slope of x within group: no variation in intercept
(1|group) + (0+x|group)	uncorrelated random intercept and random slope within group
(1|site/block) = (1|site)+(1|site:block)	intercept varying among sites and among blocks within sites (nested random effects)
site+(1|site:block)	*fixed* effect of sites plus random variation in intercept among blocks within sites
(x|site/block) = (x|site)+(x|site:block) = (1 + x|site)+(1+x|site:block)	slope and intercept varying among sites and among blocks within sites
(x1|site)+(x2|block)	two different effects, varying at different levels
x*site+(x|site:block)	fixed effect variation of slope and intercept varying among sites and random variation of slope and intercept among blocks within sites
(1|group1)+(1|group2)	intercept varying among crossed random effects (e.g. site, year)

(Bolker GLMM FAQ)

Whether nesting is explicit in the model specification depends on how you code your data

"Whether you explicitly specify a random effect as nested or not depends (in part) on the way the levels of the random effects are coded. If the 'lower-level' random effect is coded with unique levels, then the two syntaxes (1|a/b) (or (1|a)+(1|a:b)) and (1|a)+(1|b) are equivalent. If the lower-level random effect has the same labels within each larger group (e.g. blocks 1, 2, 3, 4 within sites A, B, and C) then the explicit nesting (1|a/b) is required. It seems to be considered best practice to code the nested level uniquely (e.g. A1, A2, …, B1, B2, …) so that confusion between nested and crossed effects is less likely." (Bolker GLMM FAQ)

Random effects uniquely coded

Source	PlantID	Type	Response1 …
ARTR	ARTR1	LV	
ARTR	ARTR1	ST	
ARTR	ARTR2	LV	
ARTR	ARTR2	ST	
…	…	…	
HECO	HECO1	LV	
HECO	HECO1	ST	
…	…	…	

Random effects not uniquely coded

Source	PlantID	Type	Response1 …
ARTR	1	LV	
ARTR	1	ST	
ARTR	2	LV	
ARTR	2	ST	
…	…	…	
HECO	1	LV	
HECO	1	ST	
…	…	…	

Whether nesting is explicit in the model specification depends on how you code your data

These two models do the same thing, but are specified differently because of the way data are coded

Response~Source*Type+(1|PlantID)

Random effects uniquely coded

Source	PlantID	Type	Response1 …
ARTR	ARTR1	LV	
ARTR	ARTR1	ST	
ARTR	ARTR2	LV	
ARTR	ARTR2	ST	
…	…	…	
HECO	HECO1	LV	
HECO	HECO1	ST	
…	…	…	

Response~Source*Type+(1|Source/PlantID)

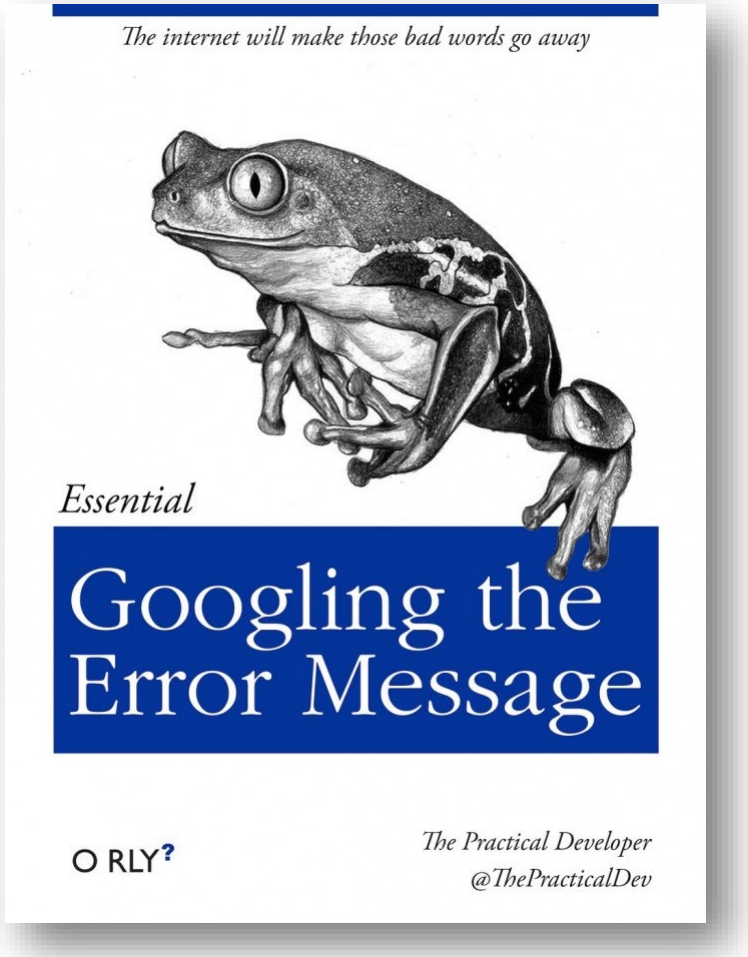
Random effects not uniquely coded

Source	PlantID	Type	Response1 …
ARTR	1	LV	
ARTR	1	ST	
ARTR	2	LV	
ARTR	2	ST	
…	…	…	
HECO	1	LV	
HECO	1	ST	
…	…	…	

Troubleshooting common issues with fitting mixed models

- Model did not converge
  - Centering/scaling predictors

- Singular model/fit

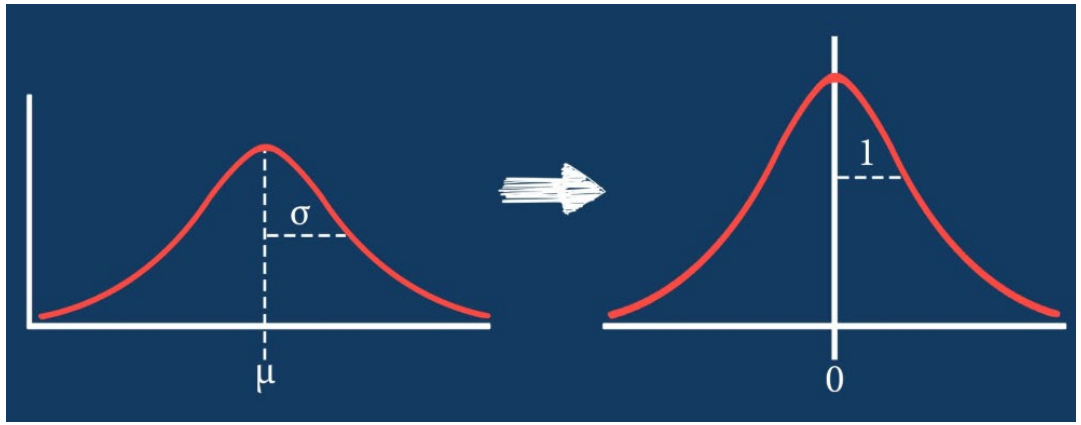


Convergence

- Refers to convergence of the algorithm used to estimate the model parameters
- “No convergence” doesn’t necessarily mean that the fit is bad or the model is wrong
- One way to potentially avoid this is make sure that the scale of your predictors are comparable, and if not, center/scale them
- Can also try other optimizer options
  - allFit()

Centering or standardizing predictors

- With continuous predictors, the scale of multiple predictors should be comparable, and not overly large compared to the response
- For example: plant height~precipitation*nitrogen+(1|site)
  - Range of plant: 0.5 - 1.8 m
  - Range of precipitation: 1000 ml – 1500 ml
  - Range of nitrogen: 5 ppm – 25 ppm
- Solution: rescale predictors so that they have mean=0 and SD=1
  - Easy to do using scale() in R
  - AKA z-score normalization

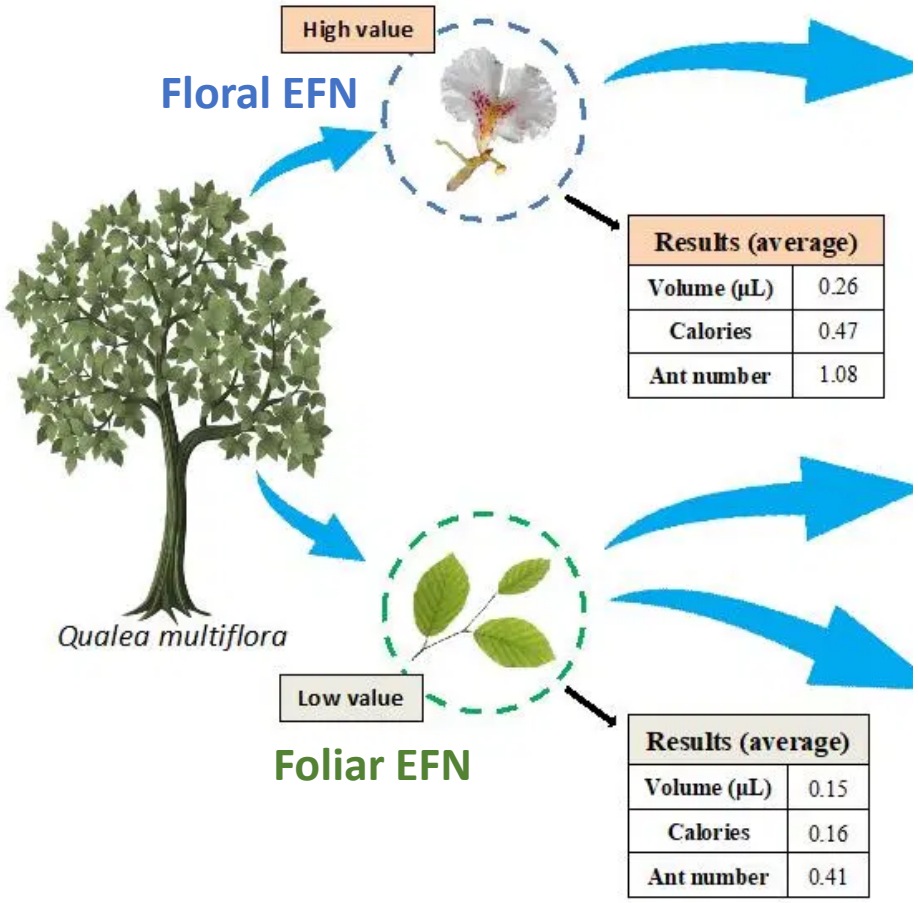


Boundary/singular fit error

- A boundary/singular fit happens when random effect variances are estimated as zero, or correlations estimated as +/- 1
- This often happens when:
  - Random effect has too few levels (e.g. <5)
  - Random effects structure is too complex (e.g. fitting a lot of random slopes and intercepts, and not enough data to do so)
- How to proceed?
  - Think about why you have the random effects in the first place…do you need them? Can you simplify them?
  - If a random effect variance component is zero, then dropping it from the model will have no effect on the model output

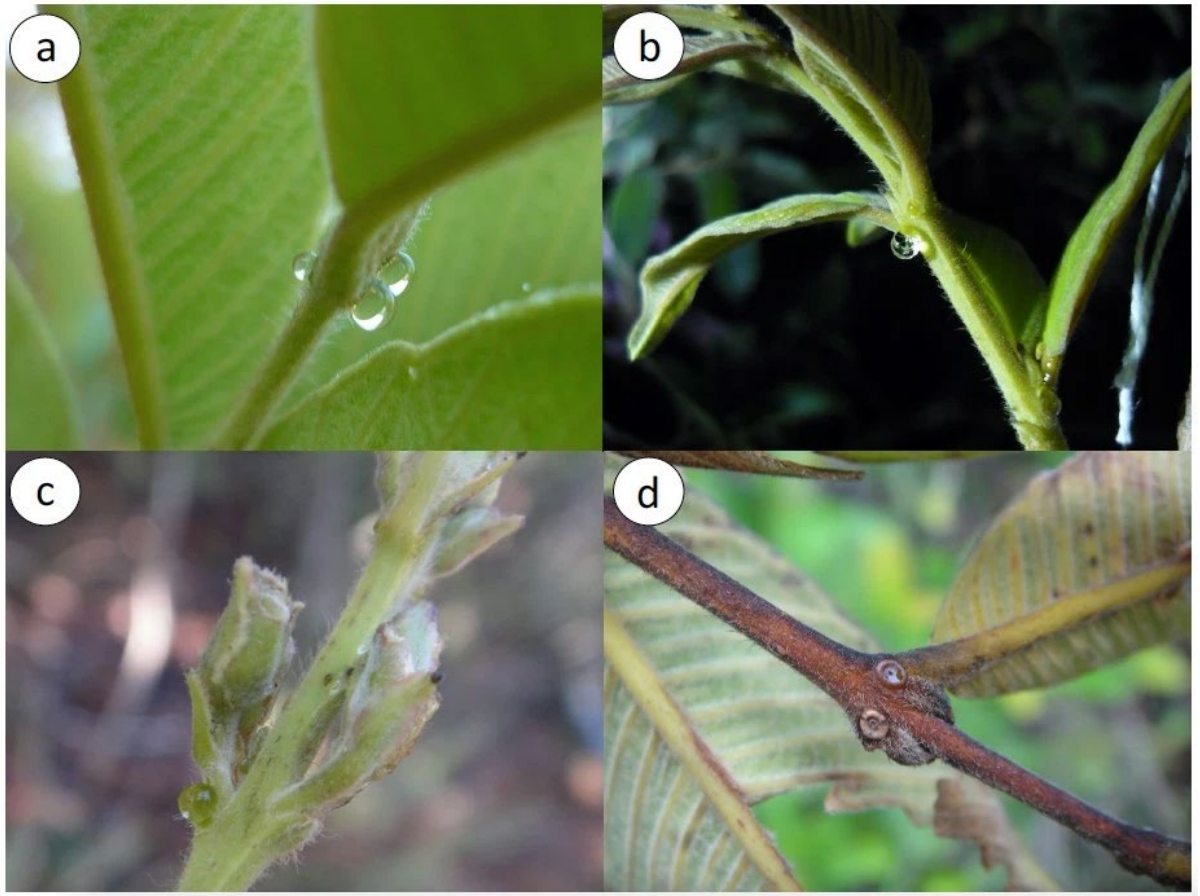
Boundary fit example

Prediction 1: Nectar produced in inflorescence extrafloral nectaries (EFNs) will have higher volumes and calories and will attract more ants than extrafloral nectar produced in leaf EFNs, given the relative value of these tissues



Results (average)	
Volume (µL)	0.26
Calories	0.47
Ant number	1.08

Results (average)	
Volume (µL)	0.15
Calories	0.16
Ant number	0.41



In their experimental design, they sampled multiple floral and foliar EFNs per plant

Calixto et al. 2020

Does nectar volume differ between floral and foliar EFNs?

```
> m1<-lmer(Volume~EFNtype + (1|Plant), data=EFN)
boundary (singular) fit: see ?isSingular
> summary(m1)
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: Volume ~ EFNtype + (1 | Plant)
   Data: EFN

REML criterion at convergence: 82.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.0451 -0.7250  0.0277  0.2353  3.3152

Random effects:
 Groups   Name        Variance Std.Dev.
 Plant    (Intercept) 0.0000   0.0000
 Residual             0.9278   0.9632
Number of obs: 30, groups:  Plant, 19

Fixed effects:
             Estimate Std. Error      df t value Pr(>|t|)
(Intercept)     2.0067     0.2487 28.0000   8.068 8.73e-09 ***
EFNtypeFoliar  -1.1333     0.3517 28.0000  -3.222  0.00322 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Boundary (singular) fit warning

Variance estimated for the random effect is zero

Does nectar volume differ between floral and foliar EFNs?

```
> m1<-lmer(Volume~EFNtype + (1|Plant), data=EFN)
boundary (singular) fit: see ?isSingular
> summary(m1)
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: Volume ~ EFNtype + (1 | Plant)
   Data: EFN

REML criterion at convergence: 82.8

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.0451 -0.7250  0.0277  0.2353  3.3152

Random effects:
 Groups   Name        Variance Std.Dev.
 Plant    (Intercept) 0.0000   0.0000
 Residual             0.9278   0.9632
Number of obs: 30, groups:  Plant, 19

Fixed effects:
             Estimate Std. Error      df t value Pr(>|t|)
(Intercept)     2.0067     0.2487 28.0000   8.068 8.73e-09 ***
EFNtypeFoliar  -1.1333     0.3517 28.0000  -3.222  0.00322 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

What to do?

Option 1: Remove random effect (1|Plant) and just fit it as a regular lm() since random effect is not explaining any variance

Option 2: Keep going, interpret the mixed model since that random effect won’t change any coefficient estimates or hypothesis tests

In this case I would probably pick Option 2 for consistency since I know I’ll be analyzing other response variables (nectar calories and ant number) in this dataset with similar models

Model fit and model adequacy

- Linear model assumptions (inspect residuals)
- Overdispersion if fitting a GLMM
- R squared can be tricky to calculate/approximate
- Stability of variance components
  - Singularity (variance component collapses to zero, can not be estimated)

A few additional thoughts about hypothesis testing for mixed models

- lme4 does not automatically produce p-values for fixed effects (if you don't simultaneously run lmerTest) because assigning degrees of freedom (among other things) get fishy quite quickly with complicated random effects

- How can you (and should you) try to get p-values for random effects?

- Often because of the iffyness/restrictions of frequentist hypothesis testing with complicated mixed models, folks often use mixed models within the context of model selection

A few additional thoughts about hypothesis testing for mixed models

Methods for testing single parameters

From worst to best:

- Wald Z-tests
- For balanced, nested LMMs where degrees of freedom can be computed according to classical rules: Wald t-tests
- Likelihood ratio test, either by setting up the model so that the parameter can be isolated/dropped (via anova or drop1, or via computing likelihood profiles
- Markov chain Monte Carlo (MCMC) or parametric bootstrap confidence intervals

Tests of effects (i.e. testing that several parameters are simultaneously zero)

From worst to best:

- Wald chi-square tests (e.g. car::Anova)
- Likelihood ratio test (via anova or drop1)
- For balanced, nested LMMs where df can be computed: conditional F-tests
- For LMMs: conditional F-tests with df correction (e.g. Kenward-Roger in pbkrtest package: see notes on K-R etc below.
- MCMC or parametric, or nonparametric, bootstrap comparisons (nonparametric bootstrapping must be implemented carefully to account for grouping factors)

(Bolker GLMM FAQ)

REML vs. ML

- ML: maximum likelihood
- REML: restricted maximum likelihood

- What you need to know:
  - ML methods can produced biased estimates of random effects variances, whereas REML is unbiased (which is why REML is the default)
  - However, REML cannot be used to compare models that differ in their fixed effects, but you can do it with ML
  - So if you are going to do any model comparison/selection, fit models using ML, find your best model, and re-fit that model with REML to get estimates

Questions?

- How to implement linear mixed models in R
  - Different packages
  - How to specify the models and random effects structure (syntax)
- Hypothesis testing in the mixed model framework
- Midterm survey
- Start thinking about final projects! Time for general questions next Tuesday.