# LMM_Non_Gap_Filled

**Created January 28, 2025**

Changes

- 4/9/2025: copy script from LMM_NvS.Rmd and update to use old, non-gap-filled data

```r
library(lme4)
```

```
## Loading required package: Matrix
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
```

## Load Data

```r
filepath = "https://raw.githubusercontent.com/shabanm2/Utqiagvik/main/elizabeth%20seasonal%20analysis/s
gtfile = "grndtemp_daily.csv"
airfile = "airtemp_daily.csv"
vwcfile = "vwc_daily.csv"
solfile = "solar_daily.csv"
windspeed = "windspeed_daily.csv"
winddir = "winddir_daily.csv"

grndtmp = read.csv(paste0(filepath, gtfile))
airtmp = read.csv(paste0(filepath, airfile))
vwc = read.csv(paste0(filepath, vwcfile))
solar = read.csv(paste0(filepath, solfile))
wind = read.csv(paste0(filepath, windspeed))

combined = full_join(grndtmp, airtmp, by=c("date", "station", "sensor")) %>% select(c(2:5,6,9))
```

```
## Warning in full_join(grndtmp, airtmp, by = c("date", "station", "sensor")): Detected an unexpected ma
## i Row 38 of `x` matches multiple rows in `y`.
## i Row 34 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
colnames(combined)[2:6] = c("station","sensor", "grounddepth", "groundtemp", "airtemp")
combined = full_join(combined, wind, by=c("date", "station", "sensor")) %>% select(c(1:6,9))
colnames(combined)[7] = c("windspeed")
combined = full_join(combined, solar, by=c("date", "station", "sensor")) %>% select(c(1:7,10))
```

```
## Warning in full_join(combined, solar, by = c("date", "station", "sensor")): Detected an unexpected ma
## i Row 38 of `x` matches multiple rows in `y`.
## i Row 75 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
colnames(combined)[8] = c("solar")

combined <- combined %>% mutate(depth = case_when(grounddepth == 7 ~ 1,
          grounddepth == 8 ~ 1,
          grounddepth == 9 ~ 2,
          grounddepth == 10 ~ 3,
          grounddepth == 11 ~ 3,
          grounddepth == 12 ~ 4,
          grounddepth == 13 ~ 4,
          grounddepth == 14 ~ 5,
          grounddepth == 15 ~ 6,
          grounddepth == 16 ~ 6,
          grounddepth == 17 ~ 6)) %>% filter(grounddepth != 1)

combined = full_join(combined, vwc, by=c("date", "station", "sensor", "depth")) %>% select(c(1:9,11))
```

```
## Warning in full_join(combined, vwc, by = c("date", "station", "sensor", : Detected an unexpected many
## i Row 38 of `x` matches multiple rows in `y`.
## i Row 38 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
colnames(combined)[c(2:3,9:10)] = c("site", "fullname", "vwcdepth", "vwc")
```

```
# File from github
# filepath = "https://raw.githubusercontent.com/shabanm2/Utqiagvik/main/Meteorological_Seasons_Data/"
#
# df <- read.csv(paste0(filepath, "daily_2022_2024.csv"))
# df <- df %>% select(-X) %>% select(-X.1)
df <- combined
df$Date <- as.POSIXct(df$date, format="%Y-%m-%d")
df <- df %>% filter(windspeed != -888.88) %>% filter(winddir != -888.88) %>% filter(site != "BUECI")
```

### Select and Transform Data

North vs South

TNHA:

North = TNHA-SC

South = TNHA-SA

SSMH:

North = SSMH-SB

South = SSMH-SA

BEO (Control): does not have different aspects

```
nvs <- df %>% filter(fullname == "TNHA-SA" | fullname == "TNHA-SC" | fullname == "SSMH-SB" | fullname ==
```

### Filter Date Function

```
# input:
# df - original df to get values from
# depth - what depth you want to filter for (as a number, 7-17)
# datemin - the start date you want to look at (inclusive) as a string
# datemax - the day after the last day you want to look at (i.e. exclusive) as a string
get_dates <- function(df, depth, datemin, datemax) {
  new_df <- df %>% filter(grounddepth == depth) %>% filter(Date >= datemin) %>% filter(Date < datemax)
}
```

```
# filter out data from before data collection
# filter to get only depth of 10cm for now
df_10cm_summer_2022 <- get_dates(nvs, 8, "2022-06-19", "2022-09-01")
```

## LMM Functions

### Fit LMM - Random Intercept

```
fit_lmm <- function(df) {
  lmm <- lmer(groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1|site), data = df)
  return(lmm)
}
```

### Fit LMM - Correlated Random Intercepts

```
fit_lmm_correlated <- function(df) {
  lmm <- lmer(groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (airtemp|site) + (vwc|site) + (
  return(lmm)
}
```

# LMM Example with Summer 2022

```r
summer2022_lmm <- fit_lmm(df_10cm_summer_2022)
summary(summer2022_lmm)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1 |
##     site)
##    Data: df
##
## REML criterion at convergence: 741.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9324 -0.5049  0.0018  0.6239  3.2721
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  site     (Intercept) 0.0318   0.1783
##  Residual             1.5332   1.2382
## Number of obs: 222, groups:  site, 2
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.595641   0.404354   3.946
## airtemp      0.468055   0.031928  14.660
## vwc         -4.030063   0.978126  -4.120
## solar        0.010074   0.001851   5.443
## windspeed    0.070931   0.095430   0.743
## aspectSouth  1.650172   0.225232   7.327
##
## Correlation of Fixed Effects:
##             (Intr) airtmp vwc    solar  wndspd
## airtemp     -0.288
## vwc         -0.630 -0.031
## solar       -0.253 -0.457  0.186
## windspeed   -0.388  0.207 -0.115 -0.038
## aspectSouth -0.322  0.136 -0.143 -0.306  0.412
```

It seems like there is not too much of a difference between the two sites, but there is still presence of a difference as shown by the two different intercepts for SSMH vs TNHA.

# Graphing linear models

Random effects have different intercepts, and the fixed effects have different slopes

```r
coeffs = coef(summer2022_lmm)$site
coeffs
```

```
##      (Intercept)   airtemp       vwc      solar  windspeed aspectSouth
```

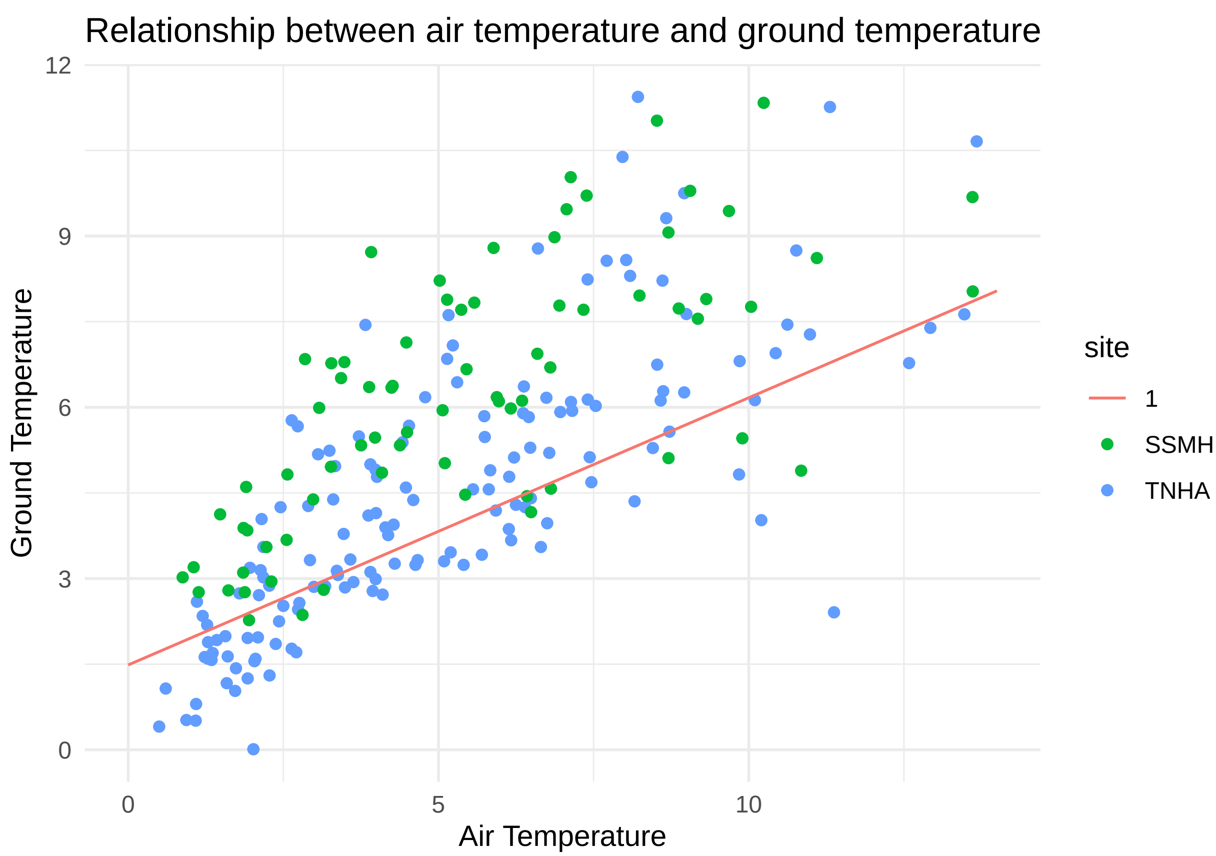
```
## SSMH     1.508517 0.4680553 -4.030063 0.01007437 0.07093062     1.650172
## TNHA     1.682766 0.4680553 -4.030063 0.01007437 0.07093062     1.650172
```

```
library(ggeffects)  # install the package first if you haven't already, then load it

# Extract the prediction data frame
pred.mm <- ggpredict(summer2022_lmm, terms = c("airtemp"))  # this gives overall predictions for the mo

# Plot the predictions

(ggplot(pred.mm) +
    geom_point(data = df_10cm_summer_2022,                      # adding the raw data (scaled values)
               aes(x = airtemp, y = groundtemp, colour = site)) +
   geom_line(aes(x = x, y = predicted, color = group)) +          # slope
   #geom_ribbon(aes(x = x, ymin = predicted - std.error, ymax = predicted + std.error),
               #fill = "lightgrey", alpha = 0.5) +  # error band
   labs(x = "Air Temperature", y = "Ground Temperature",
        title = "Relationship between air temperature and ground temperature") +
   theme_minimal()
)
```



# Table Output

```r
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```r
stargazer(summer2022_lmm, type = "text",
          digits = 3,
          star.cutoffs = c(0.05, 0.01, 0.001),
          digit.separator = "")
```

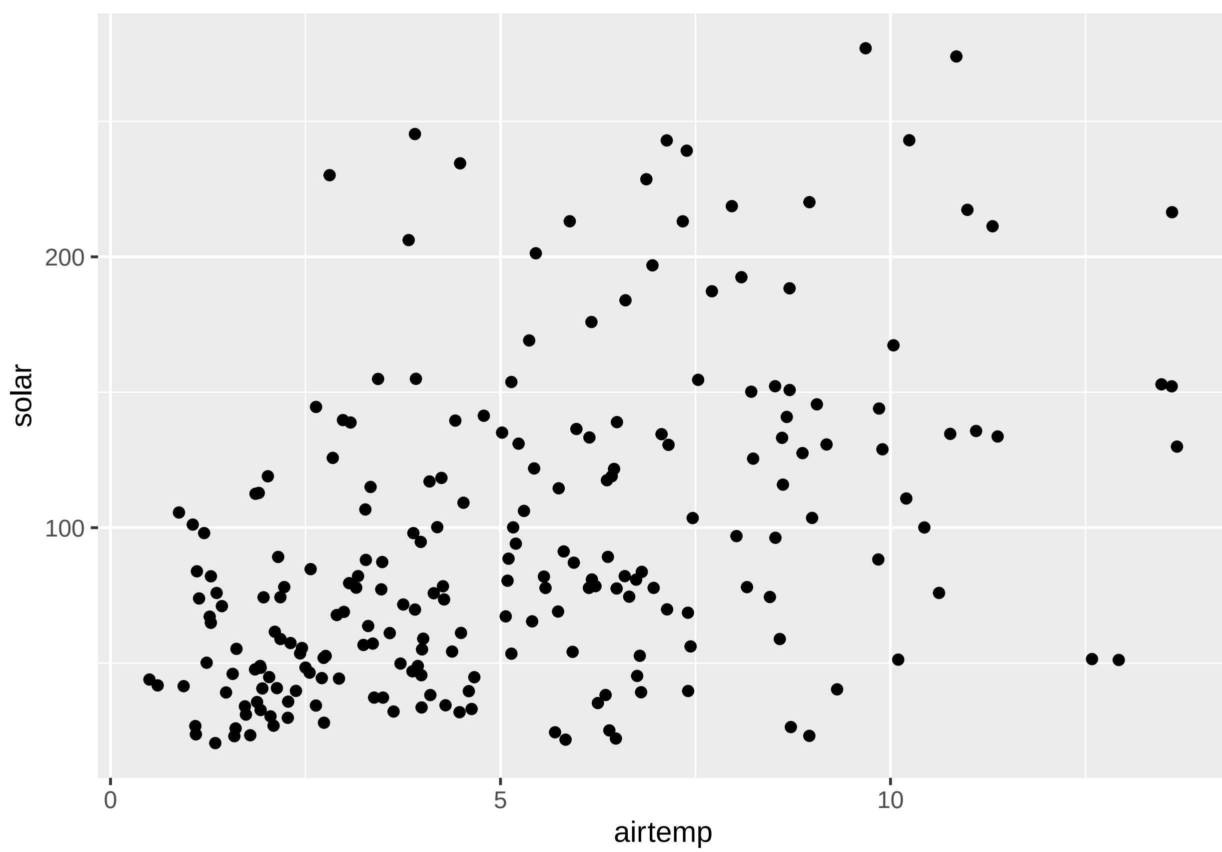
```
##
## ===================================================
##                           Dependent variable:
##                       -----------------------------
##                               groundtemp
## ---------------------------------------------------
## airtemp                        0.468***
##                                 (0.032)
##
## vwc                            -4.030***
##                                 (0.978)
##
## solar                          0.010***
##                                 (0.002)
##
## windspeed                        0.071
##                                 (0.095)
##
## aspectSouth                    1.650***
##                                 (0.225)
##
## Constant                       1.596***
##                                 (0.404)
##
## ---------------------------------------------------
## Observations                      222
## Log Likelihood                 -370.890
## Akaike Inf. Crit.               757.779
## Bayesian Inf. Crit.             785.000
## ===================================================
## Note:                 *p<0.05; **p<0.01; ***p<0.001
```

Getting an error from this code chunk because of the aspect variable

```r
#library(effects)
#est<-Effect("airtemp", partial.residuals=T, summer2022_lmm)
#plot(est)
#
#plot(summer2022_lmm)

#library(remef)
#y_partial <- remef(model, fix = "x2", ran = "all")
```

```r
ggplot(df_10cm_summer_2022, aes(x=airtemp, y=solar)) + geom_point()
```




```r
var_df = df_10cm_summer_2022 %>% select(groundtemp, airtemp, vwc, solar, windspeed)
round(cor(var_df),
  digits = 2 # rounded to 2 decimals
)
```

```
##            groundtemp airtemp   vwc solar windspeed
## groundtemp       1.00    0.73 -0.32  0.66     -0.31
## airtemp          0.73    1.00 -0.10  0.46     -0.25
## vwc             -0.32   -0.10  1.00 -0.30      0.11
## solar            0.66    0.46 -0.30  1.00     -0.22
## windspeed       -0.31   -0.25  0.11 -0.22      1.00
```

# Analysis

We are particularly interested in five micrometeorological variables to conduct our LMM: ground temperature (`groundtemp`), air temperature (`airtemp`), volumetric water content or ground moisture (`vwc`), solar radiation (`solar`), and wind speed (`windspeed`). We are comparing how the five micrometerological variables may correlate at each of our three sites with varying levels of infrastructure. Our sites are SSMH (commercial/hospital), TNHA (residential), and BEO (tundra control). For our LMM, we will only be comparing SSMH and TNHA because we want to look at differences on the north and south sides of the buildings, and BEO does not have any infrastructure to block or otherwise impact the micrometeorology at the site.

Our model will compare our fixed effects against our response variable of interest, `groundtemp` The fixed effects will include `airtemp`, `vwc`, `solar`, and `windspeed.`

Additionally, want to look at differences in our five variables between the north-facing and south-facing sides of our buildings. This variable, `aspect`, will be the last of our fixed effects.

Our model only has one random effect, which is `site`. This accounts for any "random" differences observed between SSMH and TNHA due to having different geographic locations. At the time, we are not looking at differences between the sites as predictors of ground temperature.

The formula for our LMM is $groundtemp = airtemp + vwc + solar + windspeed + aspect + (1|site)$

## Summer 2022

### Correlation Matrix of Explanatory Variables

```
var_df = df_10cm_summer_2022 %>% select(groundtemp, airtemp, vwc, solar, windspeed)
round(cor(var_df),
  digits = 2 # rounded to 2 decimals
)
```

```
##            groundtemp airtemp   vwc solar windspeed
## groundtemp       1.00    0.73 -0.32  0.66     -0.31
## airtemp          0.73    1.00 -0.10  0.46     -0.25
## vwc             -0.32   -0.10  1.00 -0.30      0.11
## solar            0.66    0.46 -0.30  1.00     -0.22
## windspeed       -0.31   -0.25  0.11 -0.22      1.00
```

### Random Effects

```
lmm = summary(summer2022_lmm)
```

Variance of Random Effects

```
lmm$varcor
```

```
##  Groups   Name        Std.Dev.
##  site     (Intercept) 0.17833
##  Residual             1.23821
```

The variance of the random effects show the variation of data (1) between sites and (2) within sites. Note that the standard deviation is the square root of the variance. The standard deviation is the spread of a group of data from the mean.

From our output, we see that there is more variation in values within sites than across from sites. However, there still is a difference in variation between sites.

**We may want to look into other ways of calculating our random effects to see if there are differences in slopes for our variables**

<table>
<tr><th>Formula</th><th>Alternative</th><th>Meaning</th></tr>
<tr><td><code>(1 | g)</code></td><td><code>1 + (1 | g)</code></td><td>Random intercept with fixed mean.</td></tr>
<tr><td><code>0 + offset(o) + (1 | g)</code></td><td><code>-1 + offset(o) + (1 | g)</code></td><td>Random intercept with a priori means.</td></tr>
<tr><td><code>(1 | g1/g2)</code></td><td><code>(1 | g1)+(1 | g1:g2)</code></td><td>Intercept varying among <code>g1</code> and <code>g2</code> within <code>g1</code>.</td></tr>
<tr><td><code>(1 | g1) + (1 | g2)</code></td><td><code>1 + (1 | g1) + (1 | g2).</code></td><td>Intercept varying among <code>g1</code> and <code>g2</code>.</td></tr>
<tr><td><code>x + (x | g)</code></td><td><code>1 + x + (1 + x | g)</code></td><td>Correlated random intercept and slope.</td></tr>
<tr><td><code>x + (x || g)</code></td><td><code>1 + x + (1 | g) + (0 + x | g)</code></td><td>Uncorrelated random intercept and slope.</td></tr>
</table>

Table 2: Examples of the right-hand-sides of mixed-effects model formulas. The names of grouping factors are denoted `g`, `g1`, and `g2`, and covariates and *a priori* known offsets as `x` and `o`.

Figure 1: table of random effects formulas

Source: Fitting Linear Mixed-Effects Models Using lme4 Bates et al., 2015

I think we should try doing it with the correlated random intercept and slope, $x + (x|g)$ such that for each variable we have a fixed effect slope and a random effect slope. The full model would be $groundtemp = airtemp + vwc + solar + windspeed + aspect + (airtemp|site) + (vwc|site) + (solar|site) + (windspeed|site) + (aspect|site)$.

Site Intercepts

```
coef(summer2022_lmm)$site
```

```
##      (Intercept)   airtemp       vwc      solar  windspeed aspectSouth
## SSMH    1.508517 0.4680553 -4.030063 0.01007437 0.07093062    1.650172
## TNHA    1.682766 0.4680553 -4.030063 0.01007437 0.07093062    1.650172
```

We can look at our intercepts for our two sites to assess whether or not there is a difference in our variables between our two sites. Based on our intercepts, it appears that there is some level of difference in ground temperature between the two sites.

This reinforces the use of correlated random intercept and slope as opposed to uncorrelated random intercept and slope in a subsequent model.

### Fixed Effects

```r
lmm$coefficients
```

```
##                 Estimate  Std. Error    t value
## (Intercept)  1.59564127 0.404353741  3.9461519
## airtemp      0.46805529 0.031928159 14.6596392
## vwc         -4.03006346 0.978125766 -4.1201894
## solar        0.01007437 0.001850946  5.4428237
## windspeed    0.07093062 0.095430378  0.7432708
## aspectSouth  1.65017228 0.225232347  7.3265332
```

### Correlation of Fixed Effects

Note: this one is not super important. We can suppress this output by setting corr=FALSE in our lmer model.

From Clay Ford at the UVA StatLab:

"These are not correlations of the variables and this is not an assessment of collinearity. Instead it's meant to give you some sense of the uncertainty in the estimated coefficients. For example, the solar coefficient is 0.008 with a standard error of 0.001. In repeated samples of this data, we expect the coefficient estimate to be between about 0.007 and 0.009. Likewise the estimated coefficient for airtmp is 0.504 with a standard error of 0.029. In repeated samples of this data, we expect the coefficient estimate to be between about 0.47 and 0.53. The correlation of those coefficients is about -0.464. This says in repeated samples we would expect one coefficient to go slightly down as the other goes slightly up. This is probably not that important in the grand scheme of things."

### Correlated Random Intercept and Slope Model

As previously stated, I wanted to try our model with the correlated random intercept and slope, $x + (x|g)$ such that for each variable we have a fixed effect slope and a random effect slope. The full model would be $groundtemp = airtemp + vwc + solar + windspeed + aspect + (airtemp|site) + (vwc|site) + (solar|site) + (windspeed|site) + (aspect|site)$.

```r
summer2022_lmm_correlated <- fit_lmm_correlated(df_10cm_summer_2022)
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
summary(summer2022_lmm_correlated)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (airtemp |
##     site) + (vwc | site) + (solar | site) + (windspeed | site) +
##     (aspect | site)
##    Data: df
##
## REML criterion at convergence: 684.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.7209 -0.5163  0.0516  0.5486  3.6377
```

```
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.  Corr
##  site     (Intercept) 1.555e+00  1.247061
##           airtemp     1.176e-05  0.003429 0.25
##  site.1   (Intercept) 4.200e-01  0.648048
##           vwc         2.657e+02 16.301205 0.99
##  site.2   (Intercept) 1.880e+00  1.371192
##           solar       3.601e-06  0.001898 -1.00
##  site.3   (Intercept) 1.933e+00  1.390481
##           windspeed   6.008e-03  0.077511 -1.00
##  site.4   (Intercept) 1.322e+00  1.149745
##           aspectSouth 1.786e+00  1.336516 0.16
##  Residual             1.123e+00  1.059921
## Number of obs: 222, groups:  site, 2
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)   2.745454   2.072727   1.325
## airtemp       0.488572   0.027612  17.694
## vwc         -13.120272  11.607094  -1.130
## solar         0.010607   0.002083   5.092
## windspeed    -0.031997   0.102994  -0.311
## aspectSouth   2.010465   1.240834   1.620
##
## Correlation of Fixed Effects:
##             (Intr) airtmp vwc    solar  wndspd
## airtemp     -0.032
## vwc          0.205 -0.009
## solar       -0.328 -0.346  0.006
## windspeed   -0.329  0.137  0.003 -0.015
## aspectSouth -0.194  0.020 -0.006 -0.037  0.076
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

# Fall 2022

```r
df_10cm_fall_2022 <- get_dates(nvs, 8, "2022-09-01", "2022-12-01")
fall2022_lmm <- fit_lmm(df_10cm_fall_2022)
summary(fall2022_lmm)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1 |
##     site)
##    Data: df
##
## REML criterion at convergence: 812.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6101 -0.6786 -0.1005  0.7237  2.4933
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  site     (Intercept) 3.9608   1.9902
##  Residual             0.9842   0.9921
## Number of obs: 279, groups:  site, 2
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -0.551275   1.423448  -0.387
## airtemp      0.167098   0.016177  10.329
## vwc          9.215003   0.743088  12.401
## solar        0.025251   0.002876   8.779
## windspeed   -0.159968   0.063484  -2.520
## aspectSouth -0.265601   0.177916  -1.493
##
## Correlation of Fixed Effects:
##             (Intr) airtmp vwc    solar  wndspd
## airtemp      0.086
## vwc         -0.051 -0.474
## solar       -0.042 -0.444 -0.214
## windspeed   -0.093 -0.123 -0.062  0.074
## aspectSouth -0.071  0.083 -0.307 -0.134  0.292
```

## Spring 2023

ERROR: missing values from SSMH

```r
df_10cm_spring_2023 <- get_dates(nvs, 8, "2023-03-01", "2023-06-01")
# spring2023_lmm <- fit_lmm(df_10cm_spring_2023)
# summary(spring2023_lmm)
```

## Summer 2023?

```r
df_10cm_summer_2023 <- get_dates(nvs, 8, "2023-06-01", "2023-09-01")
summer2023_lmm <- fit_lmm(df_10cm_summer_2023)
summary(summer2023_lmm)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1 |
##     site)
##    Data: df
##
## REML criterion at convergence: 134.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.96822 -0.69777 -0.04858  0.58601  2.31873
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  site     (Intercept) 0.5884   0.7671
##  Residual             1.9588   1.3996
## Number of obs: 37, groups:  site, 2
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) -1.850719   1.273955  -1.453
## airtemp      0.521589   0.085499   6.101
## vwc          0.451051   3.388744   0.133
## solar        0.005832   0.004042   1.443
## windspeed    0.520239   0.388362   1.340
## aspectSouth  2.750036   0.815016   3.374
##
## Correlation of Fixed Effects:
##             (Intr) airtmp vwc    solar  wndspd
## airtemp      0.015
## vwc         -0.480 -0.476
## solar       -0.639 -0.183  0.354
## windspeed   -0.585  0.172  0.233  0.184
## aspectSouth  0.230  0.267 -0.712 -0.469  0.060
```