LMM_NvS

Created January 28, 2025

Changes

• 1/28/25: loading data

```
library(lme4)

## Loading required package: Matrix

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

library(ggplot2)
```

Load Data

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

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" | fu
```

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")</pre>
```

LMM Functions

Fit LMM - Random Intercept

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

Fit LMM - Correlated Random Intercepts

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

LMM Example with Summer 2022

```
summer2022_lmm <- fit_lmm(df_10cm_summer_2022)</pre>
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: 955.7
##
## Scaled residuals:
##
       Min
                1Q Median
  -4.1222 -0.6453 -0.0094 0.5647
                                    2.5418
##
##
## Random effects:
  Groups
             Name
                         Variance Std.Dev.
##
  site
             (Intercept) 0.3026
                                  0.5501
## Residual
                         1.3784
                                  1.1741
## Number of obs: 296, groups: site, 2
## Fixed effects:
##
                Estimate Std. Error t value
## (Intercept) 3.063267
                           0.524690
                                     5.838
## airtemp
                           0.028961 17.406
                0.504097
## VWC
               -3.909724
                           0.877143
                                     -4.457
## solar
               0.008013
                           0.001483
                                      5.402
## windspeed
               -0.461289
                           0.123771
                                    -3.727
## aspectSouth 0.412346
                           0.152185
                                      2.710
## Correlation of Fixed Effects:
##
               (Intr) airtmp vwc
                                    solar wndspd
## airtemp
               -0.169
               -0.542 0.021
## VWC
               -0.165 -0.464
## solar
                              0.112
## windspeed
               -0.375
                      0.000
                             0.247
## aspectSouth -0.167 0.051
                             0.058 -0.288
                                            0.293
```

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

Relationship between air temperature and ground temperature

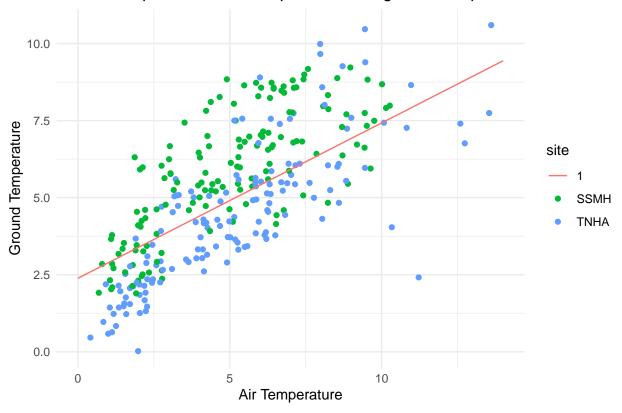


Table Output

```
library(stargazer)
```

##

```
## Please cite as:
```

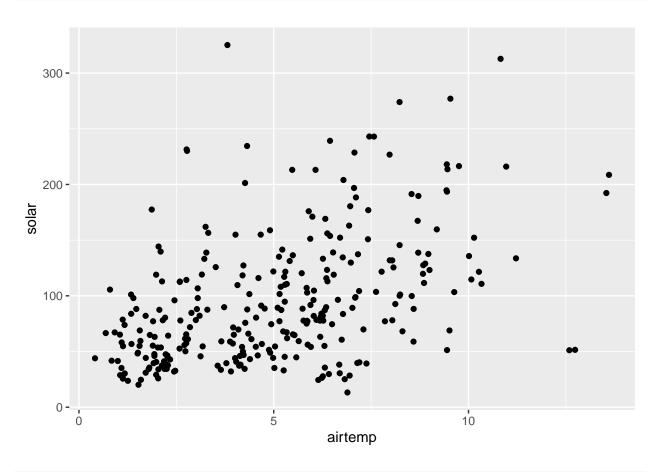
- ## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
- ## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

```
##
Dependent variable:
##
                            {\tt groundtemp}
## airtemp
                             0.504***
##
                              (0.029)
##
                             -3.910***
## VWC
##
                              (0.877)
##
                             0.008***
## solar
##
                              (0.001)
##
## windspeed
                             -0.461***
##
                              (0.124)
##
## aspectSouth
                              0.412**
                              (0.152)
##
##
## Constant
                             3.063***
##
                              (0.525)
##
## Observations
                              296
## Log Likelihood
                           -477.861
## Akaike Inf. Crit.
                            971.722
## Bayesian Inf. Crit.
                           1001.245
## Note:
                   *p<0.05; **p<0.01; ***p<0.001
```

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

```
#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")</pre>
```



```
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
                             0.72 - 0.33
                                        0.62
                                                   -0.24
                    0.72
                                                   -0.06
## airtemp
                             1.00 -0.05 0.46
## VWC
                    -0.33
                            -0.05 1.00 -0.23
                                                   -0.10
## solar
                    0.62
                             0.46 -0.23 1.00
                                                   -0.20
## windspeed
                   -0.24
                            -0.06 -0.10 -0.20
                                                    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
                            0.72 -0.33 0.62
                                                  -0.24
## airtemp
                    0.72
                            1.00 -0.05 0.46
                                                  -0.06
## VWC
                   -0.33
                           -0.05 1.00 -0.23
                                                  -0.10
## solar
                    0.62
                            0.46 -0.23 1.00
                                                  -0.20
## windspeed
                           -0.06 -0.10 -0.20
                                                   1.00
                   -0.24
```

Random Effects

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

Variance of Random Effects

lmm\$varcor

```
## Groups Name Std.Dev.
## site (Intercept) 0.55012
## Residual 1.17406
```

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

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

Formula	Alternative	Meaning
(1 g)	1 + (1 g)	Random intercept
0	4	with fixed mean.
0 + offset(o) + (1 g)	-1 + offset(o) + (1 g)	Random intercept with a priori means.
(1 g1/g2)	(1 g1)+(1 g1:g2)	Intercept varying
		among g1 and g2
		within g1.
(1 g1) + (1 g2)	1 + (1 g1) + (1 g2).	Intercept varying
		among $g1$ and $g2$.
$x + (x \mid g)$	1 + x + (1 + x g)	Correlated random
		intercept and slope.
x + (x g)	1 + x + (1 g) + (0 + x g)	Uncorrelated random
		intercept and slope.

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

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 3.443036 0.5040974 -3.909724 0.008012815 -0.4612893 0.4123458
## TNHA 2.683499 0.5040974 -3.909724 0.008012815 -0.4612893 0.4123458
```

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

lmm\$coefficients

```
## Estimate Std. Error t value
## (Intercept) 3.063267226 0.524690334 5.838238
## airtemp 0.504097359 0.028961303 17.405893
```

```
## vwc -3.909723577 0.877142635 -4.457341

## solar 0.008012815 0.001483397 5.401666

## windspeed -0.461289251 0.123771258 -3.726950

## aspectSouth 0.412345834 0.152184931 2.709505
```

Correlation of Fixed Effects

##

##

##

site.1

site.2

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).

```
summer2022_lmm_correlated <- fit_lmm_correlated(df_10cm_summer_2022)</pre>
## boundary (singular) fit: see help('isSingular')
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: 900.5
##
## Scaled residuals:
##
                1Q
                    Median
                                 3Q
                                        Max
  -4.5680 -0.5265
                   0.0222 0.5486
                                    2.9725
##
##
## Random effects:
##
    Groups
             Name
                         Variance Std.Dev.
##
    site
             (Intercept) 6.201e-01 7.875e-01
##
             airtemp
                         1.500e-06 1.225e-03 0.93
```

1.173e+01 3.426e+00 -0.73

(Intercept) 5.561e-01 7.457e-01

(Intercept) 4.383e-01 6.621e-01

```
##
                        1.032e-09 3.212e-05 -1.00
##
   site.3 (Intercept) 5.345e-01 7.311e-01
            windspeed
##
                       1.156e+00 1.075e+00 -0.07
            (Intercept) 5.179e-01 7.197e-01
##
  site.4
##
            aspectSouth 7.023e-01 8.381e-01 0.06
## Residual
                        1.106e+00 1.052e+00
## Number of obs: 296, groups: site, 2
## Fixed effects:
##
               Estimate Std. Error t value
## (Intercept) 4.038519 1.326243
                                    3.045
                          0.026112 19.857
              0.518517
## airtemp
## VWC
              -6.065754 2.565580 -2.364
## solar
                          0.001330
                                   6.035
               0.008028
## windspeed -0.771307
                          0.858868 -0.898
## aspectSouth -0.358224
                          0.736921
                                   -0.486
##
## Correlation of Fixed Effects:
##
              (Intr) airtmp vwc
                                   solar wndspd
## airtemp
              -0.040
## VWC
              -0.338 -0.003
## solar
              -0.066 -0.461 0.032
             -0.233 -0.006 0.012 0.013
## windspeed
## aspectSouth -0.235 0.000 0.012 -0.050 0.257
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Fall 2022

```
df_10cm_fall_2022 <- get_dates(nvs, 8, "2022-09-01", "2022-12-01")
fall2022_lmm <- fit_lmm(df_10cm_fall_2022)</pre>
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: 956.5
##
## Scaled residuals:
       Min
                1Q Median
                                3Q
                                       Max
## -2.7897 -0.6158 0.0441 0.6022 3.5196
##
## Random effects:
## Groups Name
                         Variance Std.Dev.
             (Intercept) 0.0802
## site
                                  0.2832
## Residual
                         0.7589
                                  0.8711
## Number of obs: 364, groups: site, 2
## Fixed effects:
```

```
##
               Estimate Std. Error t value
## (Intercept) -1.462173 0.283823 -5.152
## airtemp
              0.360554
                          0.015180 23.752
## VWC
               4.307437
                          0.680994
                                   6.325
## solar
               0.033969
                        0.003001 11.319
## windspeed
              0.420061
                          0.077563
                                   5.416
## aspectSouth -0.395202
                         0.118565 -3.333
##
## Correlation of Fixed Effects:
##
             (Intr) airtmp vwc
                                  solar wndspd
## airtemp
              0.539
              -0.576 -0.673
## VWC
## solar
              -0.134 -0.282 -0.124
## windspeed
              -0.454 -0.294 0.417 -0.143
## aspectSouth -0.135 0.081 0.190 -0.620 0.221
```

Spring 2023

ERROR: missing values from SSMH

```
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)</pre>
```

Summer 2023?

```
df_10cm_summer_2023 <- get_dates(nvs, 8, "2023-06-01", "2023-09-01")
summer2023_lmm <- fit_lmm(df_10cm_summer_2023)</pre>
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: 396
##
## Scaled residuals:
##
       Min
               1Q Median
                                3Q
                                       Max
## -2.8225 -0.7145 -0.0301 0.7924 2.0870
## Random effects:
## Groups Name
                         Variance Std.Dev.
## site
             (Intercept) 1.970
                                  1.404
## Residual
                         1.236
                                  1.112
## Number of obs: 124, groups: site, 2
##
## Fixed effects:
##
                Estimate Std. Error t value
```

```
## (Intercept) -0.058059
                        1.091224 -0.053
## airtemp 0.788353 0.055685 14.157
              2.699529 1.077593
                                  2.505
## VWC
             0.001368
                         0.001930
## solar
                                  0.709
## windspeed
             -0.075125
                         0.136552 -0.550
## aspectSouth 0.891824
                         0.248065
                                 3.595
## Correlation of Fixed Effects:
             (Intr) airtmp vwc
##
                                 solar wndspd
             -0.102
## airtemp
## vwc
             -0.249 -0.402
             -0.213 0.157 -0.028
## solar
## windspeed -0.233 0.086 0.269 0.155
## aspectSouth -0.114 -0.370 0.492 -0.300 0.134
```

Fall 2023

```
df_10cm_fall_2023 <- get_dates(nvs, 8, "2023-09-01", "2023-12-01")
fall2023_lmm <- fit_lmm(df_10cm_fall_2023)
summary(fall2023_lmm)</pre>
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1 |
      site)
##
     Data: df
##
## REML criterion at convergence: 1366.2
## Scaled residuals:
             1Q Median
                               30
      Min
                                      Max
## -2.5437 -0.7604 -0.0752 0.6154 4.3434
##
## Random effects:
## Groups
            Name
                        Variance Std.Dev.
## site
            (Intercept) 1.024
                                 1.012
## Residual
                        3.250
                                 1.803
## Number of obs: 336, groups: site, 2
##
## Fixed effects:
##
               Estimate Std. Error t value
## (Intercept) 2.744061 0.841239 3.262
## airtemp
              0.755909 0.031316 24.138
## vwc
              -8.522193 1.237836 -6.885
## solar
               0.021576
                        0.007817
                                    2.760
## windspeed -0.046308
                        0.091456 -0.506
## aspectSouth -0.617648 0.257866 -2.395
##
## Correlation of Fixed Effects:
##
             (Intr) airtmp vwc
                                   solar wndspd
## airtemp
              0.291
              -0.431 -0.271
## VWC
```

```
## solar
             -0.162 -0.686 -0.042
## windspeed -0.177 0.003 0.045 0.019
## aspectSouth -0.314 0.039 0.528 -0.275 0.293
```

Spring 2024

```
df_10cm_spring_2024 <- get_dates(nvs, 8, "2024-03-01", "2024-06-01")
spring2024_lmm <- fit_lmm(df_10cm_spring_2024)</pre>
## Warning: Some predictor variables are on very different scales: consider
## rescaling
summary(spring2024_lmm)
## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1 |
##
      site)
##
     Data: df
##
## REML criterion at convergence: 1692.7
## Scaled residuals:
##
      Min
           1Q Median
                                      Max
## -2.0736 -0.4768 -0.0900 0.4493 4.2002
##
## Random effects:
## Groups Name
                        Variance Std.Dev.
            (Intercept) 0.1437 0.379
## site
                        10.2671 3.204
## Residual
## Number of obs: 326, groups: site, 2
##
## Fixed effects:
                Estimate Std. Error t value
## (Intercept) -10.337632 1.023318 -10.102
## airtemp
               0.291803 0.033425
                                      8.730
## VWC
               28.572782
                          5.020347
                                      5.691
## solar
                0.007475
                          0.003757
                                     1.989
## windspeed
                0.308392
                          0.190959
                                     1.615
## aspectSouth 2.166333
                          0.404173
                                      5.360
##
## Correlation of Fixed Effects:
             (Intr) airtmp vwc
                                   solar wndspd
## airtemp
              0.769
## VWC
              -0.698 -0.425
## solar
              -0.565 -0.517 0.055
## windspeed -0.323 -0.099 0.212 0.032
## aspectSouth -0.208  0.001  0.305 -0.338  0.132
## fit warnings:
## Some predictor variables are on very different scales: consider rescaling
```

Summer 2024

```
df 10cm summer 2024 <- get dates(nvs, 8, "2024-06-01", "2024-09-01")
summer2024_lmm <- fit_lmm(df_10cm_summer_2024)</pre>
summary(summer2024_lmm)
## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1 |
##
      site)
##
     Data: df
##
## REML criterion at convergence: 1218.5
##
## Scaled residuals:
##
      Min
               1Q Median
                               3Q
                                      Max
## -4.0939 -0.5414 -0.0330 0.6532 3.0557
##
## Random effects:
## Groups Name
                       Variance Std.Dev.
## site
         (Intercept) 0.2058
                               0.4537
## Residual
                        1.5082
                                 1.2281
## Number of obs: 368, groups: site, 2
##
## Fixed effects:
                Estimate Std. Error t value
## (Intercept) -0.7778817 0.4726349 -1.646
## airtemp
             0.8047200 0.0261341 30.792
## VWC
              3.0331240 0.6386741
                                     4.749
             0.0001559 0.0012104 0.129
## solar
## windspeed 0.0844563 0.0744593
                                    1.134
## aspectSouth 1.3166946 0.1778824
                                    7.402
## Correlation of Fixed Effects:
##
             (Intr) airtmp vwc
                                  solar wndspd
## airtemp
             -0.117
              -0.588 -0.272
## VWC
              -0.399 0.170 0.197
## solar
## windspeed -0.328 0.063 0.256 0.061
## aspectSouth -0.418 -0.305 0.650 -0.068 0.255
```

Fall 2024

```
df_10cm_fall_2024 <- get_dates(nvs, 8, "2024-09-01", "2024-12-01")
fall2024_lmm <- fit_lmm(df_10cm_fall_2024)
summary(fall2024_lmm)

## Linear mixed model fit by REML ['lmerMod']
## Formula: groundtemp ~ airtemp + vwc + solar + windspeed + aspect + (1 |
## site)</pre>
```

```
Data: df
##
##
## REML criterion at convergence: 145.3
## Scaled residuals:
          1Q Median 3Q
##
      Min
                                  Max
## -3.9397 -0.4401 0.0954 0.5860 2.6190
##
## Random effects:
## Groups Name
                     Variance Std.Dev.
## site
           (Intercept) 0.2234 0.4727
## Residual
                      0.2243 0.4736
## Number of obs: 92, groups: site, 2
##
## Fixed effects:
##
              Estimate Std. Error t value
## (Intercept) -0.050057 0.528148 -0.095
## airtemp 0.781066 0.048210 16.201
            1.021078 0.605391
## VWC
                                1.687
             ## solar
## windspeed 0.085388 0.069052 1.237
## aspectSouth 0.956131 0.188392 5.075
##
## Correlation of Fixed Effects:
##
          (Intr) airtmp vwc solar wndspd
## airtemp
            -0.077
## VWC
            -0.665 -0.321
## solar
            -0.424 -0.376 0.502
## windspeed -0.366 0.052 0.366 0.076
## aspectSouth -0.578 -0.307 0.841 0.332 0.346
```