Linear model analysis for temperature difference size

2024-10-14 07:03:21.239061

```r
library(viridisLite)
```

Temperature difference.

```r
d2$dtemp <- d2$Itemp - d2$Otemp
```

```r
fit1 <- lm(Itemp ~ poly(glorad, 3) + poly(Otemp, 3) + poly(wv2, 3), data = d2)
fit2 <- lm(dtemp ~ poly(glorad, 3) + poly(Otemp, 3) + poly(wv2, 3), data = d2)
```

Check results

```r
summary(fit1)
```

```
##
## Call:
## lm(formula = Itemp ~ poly(glorad, 3) + poly(Otemp, 3) + poly(wv2,
##     3), data = d2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8227 -0.5742  0.0287  0.5749  7.5529
##
## Coefficients:
##                    Estimate Std. Error  t value Pr(>|t|)
## (Intercept)        15.07577    0.01404 1073.801  < 2e-16 ***
## poly(glorad, 3)1   47.94051    1.59253   30.103  < 2e-16 ***
## poly(glorad, 3)2   -6.16850    1.24507   -4.954 7.42e-07 ***
## poly(glorad, 3)3    5.05989    1.21319    4.171 3.07e-05 ***
## poly(Otemp, 3)1   472.01562    1.38080  341.841  < 2e-16 ***
## poly(Otemp, 3)2    62.94090    1.23672   50.894  < 2e-16 ***
## poly(Otemp, 3)3    37.86700    1.22338   30.953  < 2e-16 ***
## poly(wv2, 3)1      -0.85992    1.41834   -0.606 0.544345
## poly(wv2, 3)2      -1.13793    1.25875   -0.904 0.366016
## poly(wv2, 3)3      -4.73877    1.22117   -3.881 0.000105 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.195 on 7230 degrees of freedom
## Multiple R-squared:  0.9607, Adjusted R-squared:  0.9606
## F-statistic: 1.963e+04 on 9 and 7230 DF,  p-value: < 2.2e-16
```

```r
summary(fit2)
```

```
##
## Call:
## lm(formula = dtemp ~ poly(glorad, 3) + poly(Otemp, 3) + poly(wv2,
##     3), data = d2)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -7.8227 -0.5742  0.0287  0.5749  7.5529 
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)    
## (Intercept)       1.11794    0.01404  79.627  < 2e-16 ***
## poly(glorad, 3)1 47.94051    1.59253  30.103  < 2e-16 ***
## poly(glorad, 3)2 -6.16850    1.24507  -4.954 7.42e-07 ***
## poly(glorad, 3)3  5.05989    1.21319   4.171 3.07e-05 ***
## poly(Otemp, 3)1  80.78311    1.38080  58.504  < 2e-16 ***
## poly(Otemp, 3)2  62.94090    1.23672  50.894  < 2e-16 ***
## poly(Otemp, 3)3  37.86700    1.22338  30.953  < 2e-16 ***
## poly(wv2, 3)1    -0.85992    1.41834  -0.606 0.544345    
## poly(wv2, 3)2    -1.13793    1.25875  -0.904 0.366016    
## poly(wv2, 3)3    -4.73877    1.22117  -3.881 0.000105 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.195 on 7230 degrees of freedom
## Multiple R-squared:  0.6411, Adjusted R-squared:  0.6406 
## F-statistic:  1435 on 9 and 7230 DF,  p-value: < 2.2e-16
```

Temperature difference model is better–lower residual standard error. (R-squared is lower but that is just because we have already removed a lot of the variation by calculating a difference.)

Let's generate scaled predictor variables for standardized coefficients (relative to 1 standard deviation of predictor variable). This will show which predictors are the most important compared to how much they vary.

```r
fit3 <- lm(dtemp ~ poly(scale(glorad), 3) + poly(scale(Otemp), 3) + poly(scale(wv2), 3), data = d2)
```

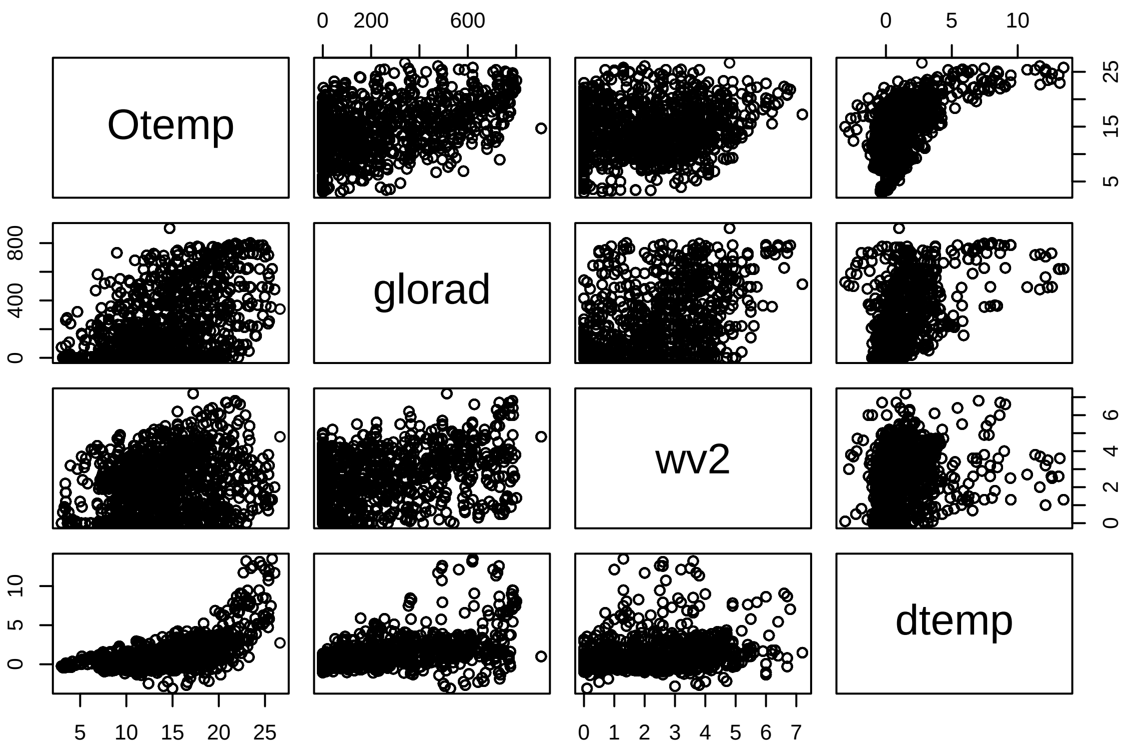
```r
summary(fit3)
```

```
## 
## Call:
## lm(formula = dtemp ~ poly(scale(glorad), 3) + poly(scale(Otemp), 
##     3) + poly(scale(wv2), 3), data = d2)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -7.8227 -0.5742  0.0287  0.5749  7.5529 
## 
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)    
## (Intercept)              1.11794    0.01404  79.627  < 2e-16 ***
## poly(scale(glorad), 3)1 47.94051    1.59253  30.103  < 2e-16 ***
## poly(scale(glorad), 3)2 -6.16850    1.24507  -4.954 7.42e-07 ***
## poly(scale(glorad), 3)3  5.05989    1.21319   4.171 3.07e-05 ***
## poly(scale(Otemp), 3)1  80.78311    1.38080  58.504  < 2e-16 ***
## poly(scale(Otemp), 3)2  62.94090    1.23672  50.894  < 2e-16 ***
## poly(scale(Otemp), 3)3  37.86700    1.22338  30.953  < 2e-16 ***
## poly(scale(wv2), 3)1    -0.85992    1.41834  -0.606 0.544345    
## poly(scale(wv2), 3)2    -1.13793    1.25875  -0.904 0.366016    
## poly(scale(wv2), 3)3    -4.73877    1.22117  -3.881 0.000105 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.195 on 7230 degrees of freedom
## Multiple R-squared:  0.6411, Adjusted R-squared:  0.6406
## F-statistic:  1435 on 9 and 7230 DF,  p-value: < 2.2e-16
```

It looks like temperature (`Otemp`) is the most important. Is that supported by the measurements?

```r
pairs(d2[, .(Otemp, glorad, wv2, dtemp)])
```



Seem so, yes.

Let's see how much worse the model is without the other two.

```r
fit4 <- lm(dtemp ~ poly(Otemp, 3), data = d2)
```

```r
summary(fit4)
```

```
##
## Call:
## lm(formula = dtemp ~ poly(Otemp, 3), data = d2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2102 -0.7387 -0.0205  0.5493  7.8603
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.11794    0.01519   73.60   <2e-16 ***
## poly(Otemp, 3)1 103.12430    1.29249   79.79   <2e-16 ***
## poly(Otemp, 3)2  69.13284    1.29249   53.49   <2e-16 ***
## poly(Otemp, 3)3  35.25896    1.29249   27.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.292 on 7236 degrees of freedom
## Multiple R-squared:  0.5795, Adjusted R-squared:  0.5793
## F-statistic:  3324 on 3 and 7236 DF,  p-value: < 2.2e-16
```

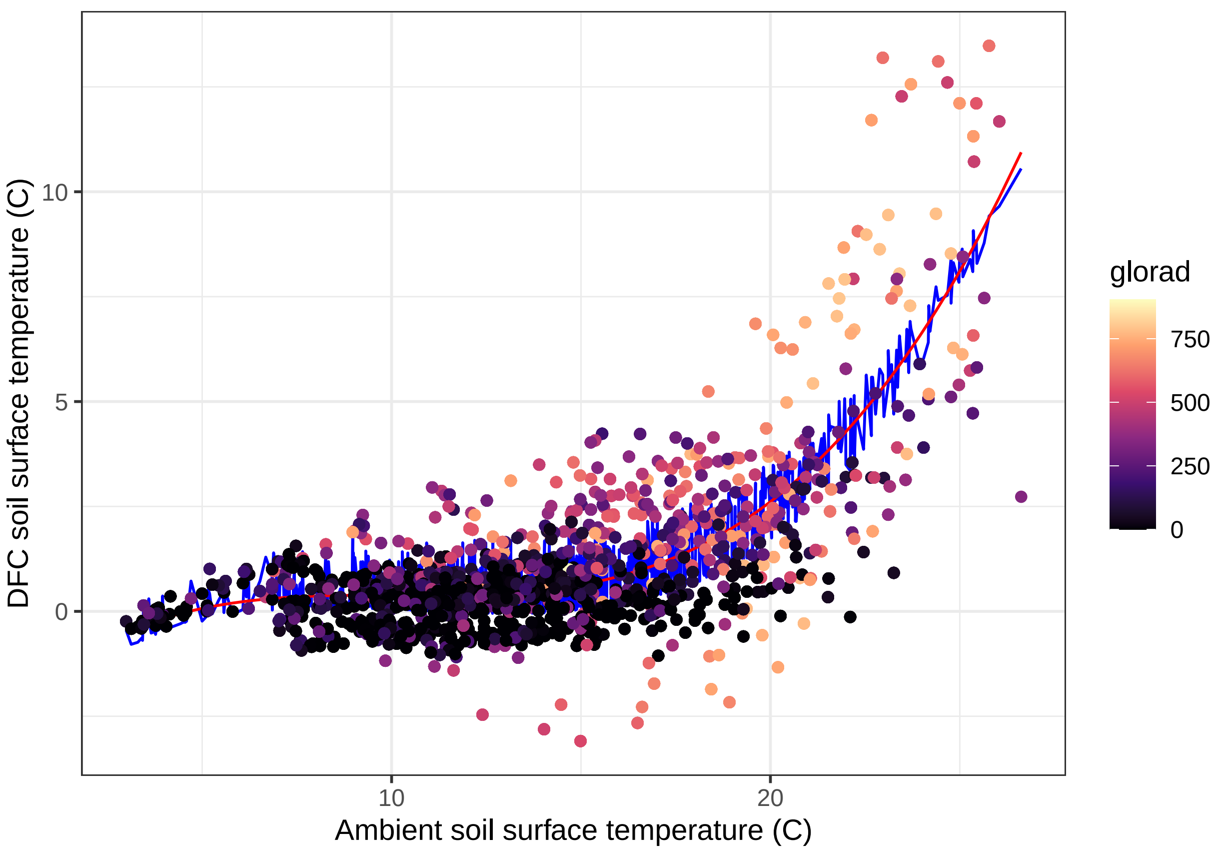
It is almost the same as 2. So effects of radiation and wind look small, surprisingly.

Generate predictions for plotting.

```
d2$Itemp.pred <- predict(fit1)
d2$dtemp.pred2 <- predict(fit2)
d2$dtemp.pred4 <- predict(fit4)
```

And take a look.

```
ggplot(d2, aes(Otemp, dtemp, colour = glorad)) +
  geom_line(aes(Otemp, dtemp.pred2), colour = 'blue') +
  geom_line(aes(Otemp, dtemp.pred4), colour = 'red') +
  geom_point() +
  scale_color_viridis_c(option = 'magma') +
  theme_bw() +
  xlab('Ambient soil surface temperature (C)') + ylab('DFC soil surface temperature (C)')
```



So, temperature alone indeed does as well as the most complete model. But both miss a lot of the variation. How about effects of earlier weather? For that we need to add lagged predictor variables.

```
d2[, etime := as.numeric(difftime(t.start, min(t.start), units = 'hours'))]
d2
```

```
## Key: <t.start>
```

```
##                  t.start  pos.x    Itemp  pos.y    Otemp  temp       date
##                   <POSc> <char>    <num> <char>    <num> <num>     <IDat>
##    1: 2024-04-24 19:00:00     in  7.11750    out  6.239167   5.5 2024-04-24
##    2: 2024-04-24 19:00:00     in  7.11750    out  6.239167   5.5 2024-04-24
##    3: 2024-04-24 19:00:00     in  7.11750    out  6.239167   5.5 2024-04-24
##    4: 2024-04-24 19:00:00     in  7.11750    out  6.239167   5.5 2024-04-24
##    5: 2024-04-24 19:00:00     in  7.11750    out  6.239167   5.5 2024-04-24
##   ---
## 7236: 2024-07-05 05:00:00     in 11.22333    out 11.042500  11.8 2024-07-05
## 7237: 2024-07-05 05:00:00     in 11.22333    out 11.042500  11.8 2024-07-05
## 7238: 2024-07-05 05:00:00     in 11.22333    out 11.042500  11.8 2024-07-05
## 7239: 2024-07-05 05:00:00     in 11.22333    out 11.042500  11.8 2024-07-05
## 7240: 2024-07-05 05:00:00     in 11.22333    out 11.042500  11.8 2024-07-05
##           time  prec surfwet glorad  metp megrtp mesotp10 mesotp30 meanrh
##         <char> <num>   <int>  <num> <num>  <num>    <num>    <num>  <num>
##    1: 19:00:00   0.0       0    7.1   5.2    4.9      7.4      7.0   86.4
##    2: 19:00:00   0.0       0    7.1   5.2    4.9      7.4      7.0   86.4
##    3: 19:00:00   0.0       0    7.1   5.2    4.9      7.4      7.0   86.4
##    4: 19:00:00   0.0       0    7.1   5.2    4.9      7.4      7.0   86.4
##    5: 19:00:00   0.0       0    7.1   5.2    4.9      7.4      7.0   86.4
##   ---
## 7236: 05:00:00   0.3       0  120.6   9.6    9.4     14.3     14.9   91.3
## 7237: 05:00:00   0.3       0  120.6   9.6    9.4     14.3     14.9   91.3
## 7238: 05:00:00   0.3       0  120.6   9.6    9.4     14.3     14.9   91.3
## 7239: 05:00:00   0.3       0  120.6   9.6    9.4     14.3     14.9   91.3
## 7240: 05:00:00   0.3       0  120.6   9.6    9.4     14.3     14.9   91.3
##       meanwd meanwv   wd2   wv2  pres netrad heatflux Itemp.pred     dtemp
##        <num>  <num> <num> <num> <num>  <num>    <num>      <num>     <num>
##    1:  234.1      0 234.1   0.0 997.5  -19.7        8   6.337001 0.8783333
##    2:  234.1      0 234.1   0.0 997.5  -19.7        8   6.337001 0.8783333
##    3:  234.1      0 234.1   0.0 997.5  -19.7        8   6.337001 0.8783333
##    4:  234.1      0 234.1   0.0 997.5  -19.7        8   6.337001 0.8783333
##    5:  234.1      0 234.1   0.0 997.5  -19.7        8   6.337001 0.8783333
##   ---
## 7236:  194.8      0 194.8   2.5 991.5   27.7      -10  11.574261 0.1808333
## 7237:  194.8      0 194.8   2.5 991.5   27.7      -10  11.574261 0.1808333
## 7238:  194.8      0 194.8   2.5 991.5   27.7      -10  11.574261 0.1808333
## 7239:  194.8      0 194.8   2.5 991.5   27.7      -10  11.574261 0.1808333
## 7240:  194.8      0 194.8   2.5 991.5   27.7      -10  11.574261 0.1808333
##        dtemp.pred dtemp.pred2 dtemp.pred4 etime
##             <num>       <num>       <num> <num>
##    1: -0.04605882  0.09783421   0.2499027     0
##    2: -0.04605882  0.09783421   0.2499027     0
##    3: -0.04605882  0.09783421   0.2499027     0
##    4: -0.04605882  0.09783421   0.2499027     0
##    5: -0.04605882  0.09783421   0.2499027     0
##   ---
## 7236:  0.55985289  0.53176128   0.3800066  1714
## 7237:  0.55985289  0.53176128   0.3800066  1714
## 7238:  0.55985289  0.53176128   0.3800066  1714
## 7239:  0.55985289  0.53176128   0.3800066  1714
## 7240:  0.55985289  0.53176128   0.3800066  1714
```

```r
head(table(d2$etime))
```

```
##
## 0 1 2 3 4 5
## 5 7 6 7 6 6
```

```r
tail(table(d2$etime))
```

```
##
## 1709 1710 1711 1712 1713 1714
##    7    5    7    6    7    6
```

Hmm, why do we have multiple observations for each unique value? Oh, were there 7 different DFCs? In that case we need a DFC ID/key in this data frame.

```r
names(d2)
```

```
##  [1] "t.start"     "pos.x"       "Itemp"       "pos.y"       "Otemp"
##  [6] "temp"        "date"        "time"        "prec"        "surfwet"
## [11] "glorad"      "metp"        "megrtp"      "mesotp10"    "mesotp30"
## [16] "meanrh"      "meanwd"      "meanwv"      "wd2"         "wv2"
## [21] "pres"        "netrad"      "heatflux"    "Itemp.pred"  "dtemp"
## [26] "dtemp.pred"  "dtemp.pred2" "dtemp.pred4" "etime"
```