Linear model analysis for temperature difference size

2024-11-08 20:17:06.996192

```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)
```

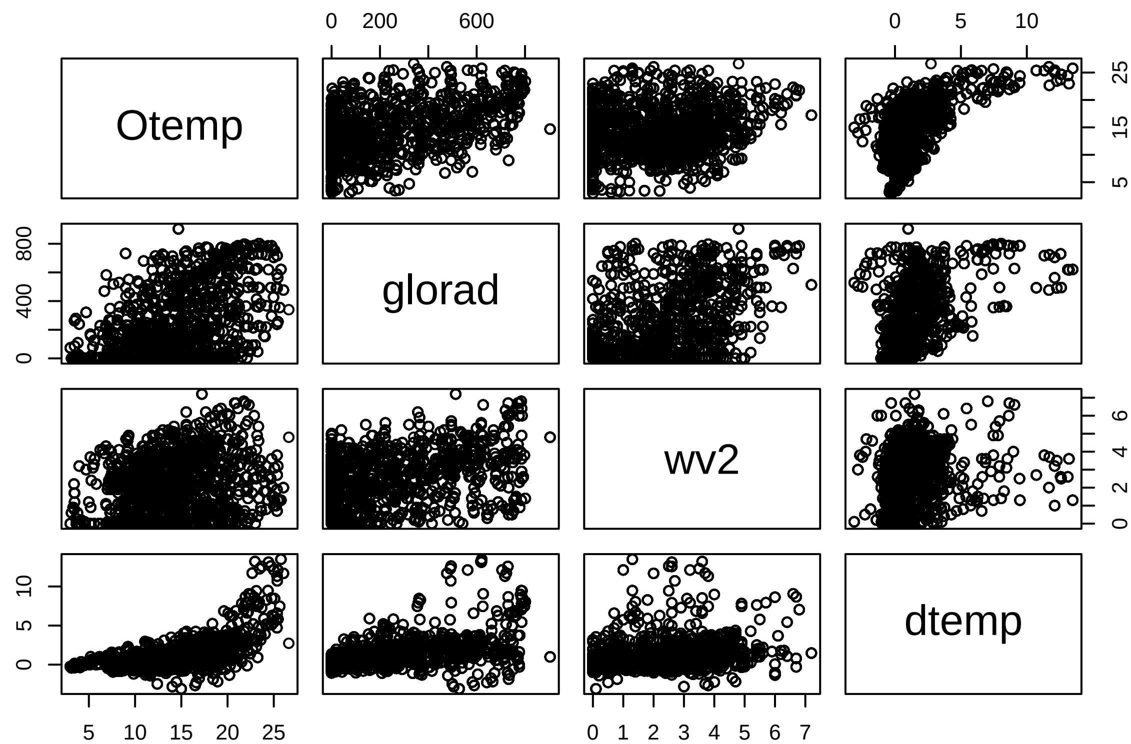
```r
summary(fit3)
```

```
##
## 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.

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

```
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
```

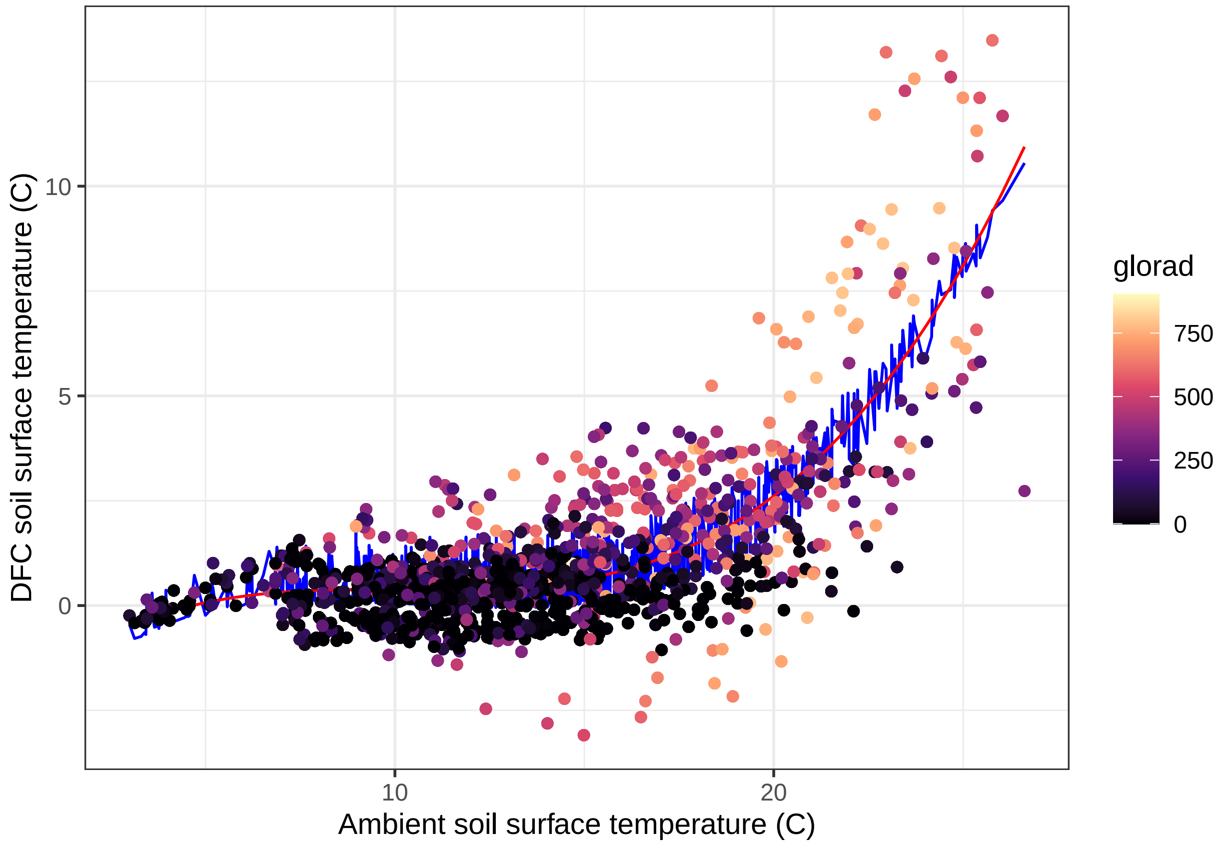
It is almost the same as 2. So effects of radiation and wind look small, surprisingly.

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. Try previous hour.

```
wthr <- unique(d2[, .(t.start, Otemp, glorad, wv2)])
wthr[, t.start := t.start - 3600]
d2.orig <- d2
d2 <- merge(d2, wthr, by = 't.start', suffixes = c('', '.lag1'))
wthr[, t.start := t.start - 3600]
d2 <- merge(d2, wthr, by = 't.start', suffixes = c('', '.lag2'))
```

```
fit5 <- lm(dtemp ~ poly(glorad, 3) + poly(Otemp, 3) + poly(wv2, 3) + poly(glorad.lag1, 3) + poly(Otemp.l
summary(fit5)
```

```
##
## Call:
## lm(formula = dtemp ~ poly(glorad, 3) + poly(Otemp, 3) + poly(wv2,
##     3) + poly(glorad.lag1, 3) + poly(Otemp.lag1, 3) + poly(wv2.lag1,
##     3), data = d2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2062 -0.5393  0.0354  0.5314  6.6285
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)            1.13238    0.01311  86.350  < 2e-16 ***
## poly(glorad, 3)1       95.80272    3.89008  24.627  < 2e-16 ***
## poly(glorad, 3)2      -16.59054    2.00486  -8.275  < 2e-16 ***
## poly(glorad, 3)3        4.82763    1.39403   3.463 0.000537 ***
## poly(Otemp, 3)1       116.77069    6.91358  16.890  < 2e-16 ***
## poly(Otemp, 3)2        -8.41262    4.28916  -1.961 0.049875 *
## poly(Otemp, 3)3       -11.16063    3.15370  -3.539 0.000404 ***
## poly(wv2, 3)1          19.06307    3.15036   6.051 1.51e-09 ***
## poly(wv2, 3)2           4.55615    2.04077   2.233 0.025609 *
## poly(wv2, 3)3          -0.40015    1.57065  -0.255 0.798913
## poly(glorad.lag1, 3)1 -42.18464    3.57016 -11.816  < 2e-16 ***
## poly(glorad.lag1, 3)2  -8.54178    1.91974  -4.449 8.74e-06 ***
## poly(glorad.lag1, 3)3   0.33860    1.37231   0.247 0.805121
## poly(Otemp.lag1, 3)1  -44.78975    7.42097  -6.036 1.66e-09 ***
## poly(Otemp.lag1, 3)2   73.11074    4.36503  16.749  < 2e-16 ***
## poly(Otemp.lag1, 3)3   54.94227    3.20825  17.125  < 2e-16 ***
## poly(wv2.lag1, 3)1    -24.15847    3.18324  -7.589 3.63e-14 ***
## poly(wv2.lag1, 3)2     -4.35724    2.04170  -2.134 0.032866 *
## poly(wv2.lag1, 3)3     -6.75304    1.58355  -4.264 2.03e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.107 on 7111 degrees of freedom
## Multiple R-squared:  0.6951, Adjusted R-squared:  0.6944
## F-statistic: 900.8 on 18 and 7111 DF,  p-value: < 2.2e-16
```

```
summary(fit2)
```

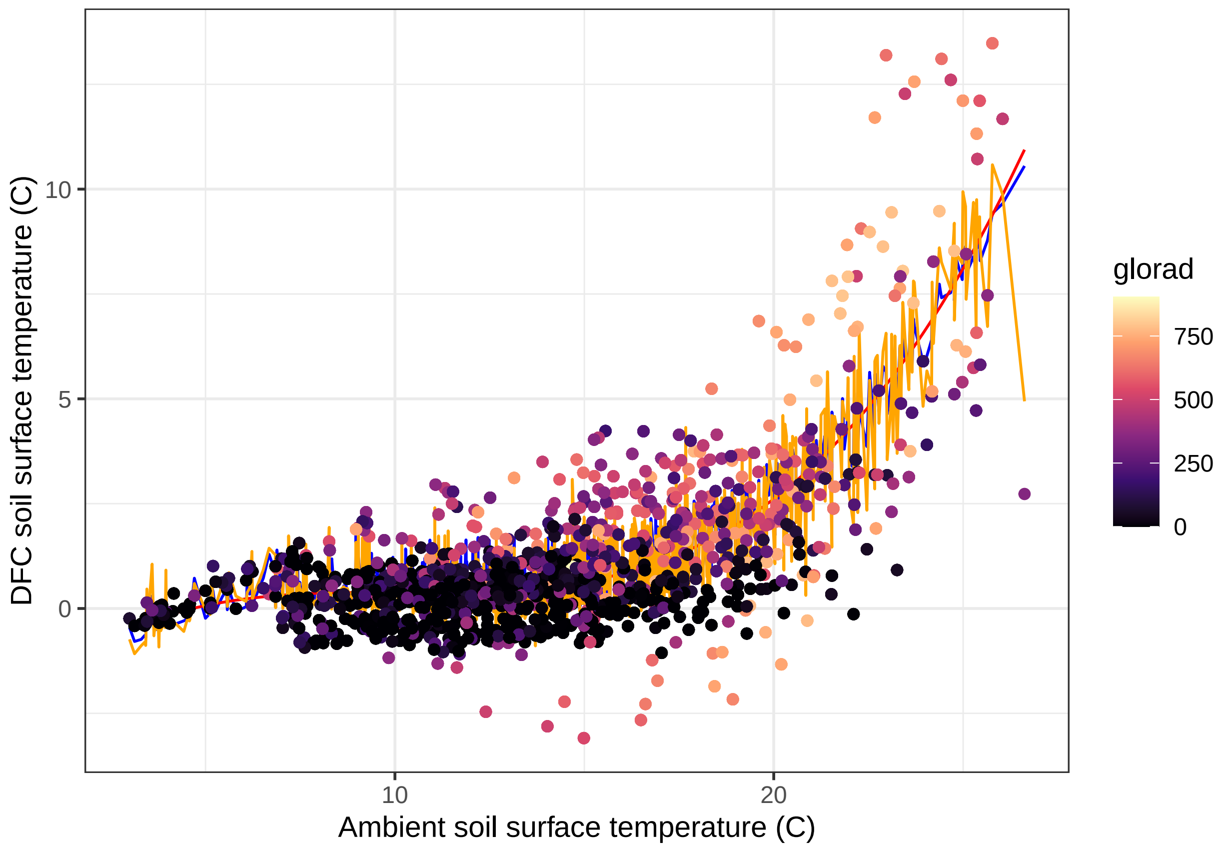
```
##
## Call:
## lm(formula = dtemp ~ poly(glorad, 3) + poly(Otemp, 3) + poly(wv2,
##     3), data = d2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8227 -0.5742  0.0287  0.5749  7.5529
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       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
```

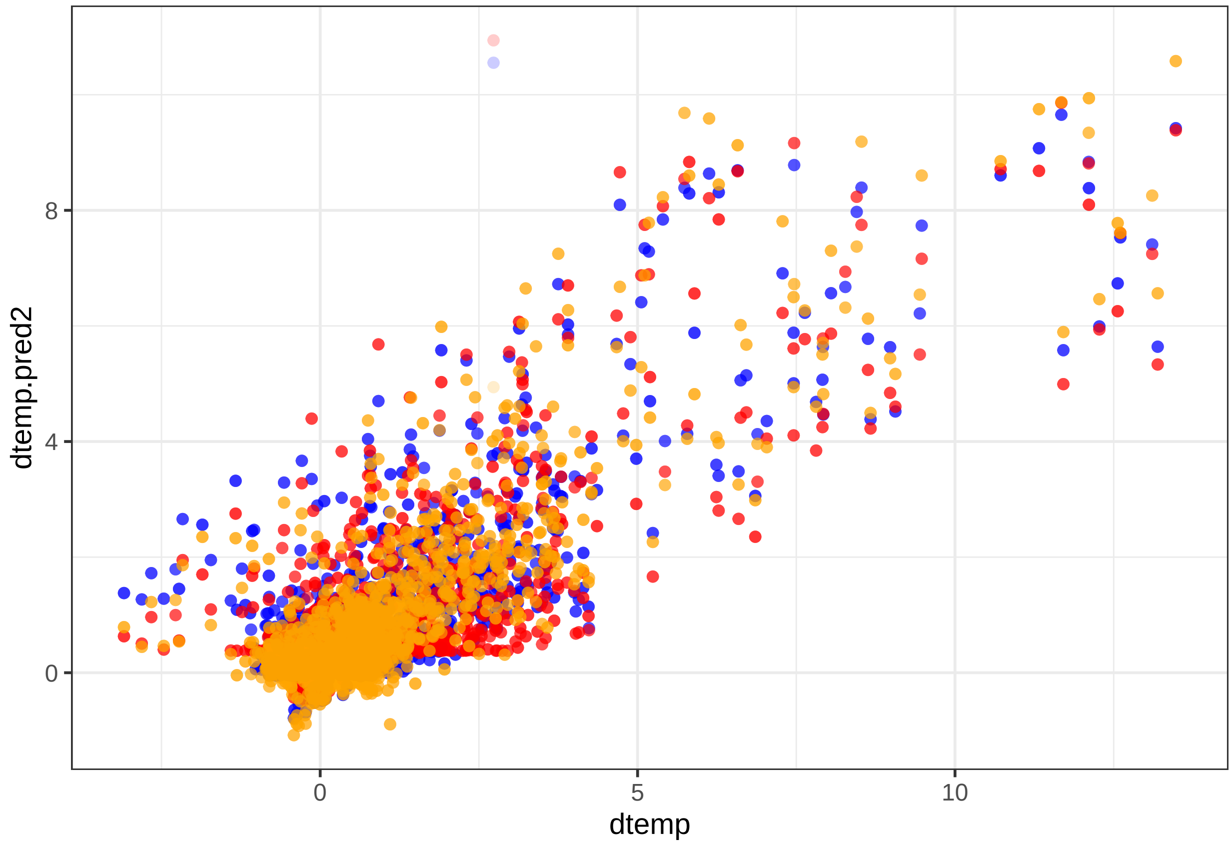
Not a lot of improvement really.

```r
d2$dtemp.pred5 <- predict(fit5)
```

```r
ggplot(d2, aes(Otemp, dtemp, colour = glorad)) +
  geom_line(aes(Otemp, dtemp.pred2), colour = 'blue') +
  geom_line(aes(Otemp, dtemp.pred4), colour = 'red') +
  geom_line(aes(Otemp, dtemp.pred5), colour = 'orange') +
  geom_point() +
  scale_color_viridis_c(option = 'magma') +
  theme_bw() +
  xlab('Ambient soil surface temperature (C)') + ylab('DFC soil surface temperature (C)')
```




```r
ggplot(d2) +
  geom_point(aes(dtemp, dtemp.pred2), colour = 'blue', alpha = 0.2) +
  geom_point(aes(dtemp, dtemp.pred4), colour = 'red', alpha = 0.2) +
  geom_point(aes(dtemp, dtemp.pred5), colour = 'orange', alpha = 0.2) +
  theme_bw()
```




```r
newdat <- expand.grid(glorad = 0:10 * 100, Otemp = 0:10 * 3, wv2 = 2)
setDT(newdat)
newdat$dtemp.pred5 <- predict(fit2, newdata = newdat)
newdat[dtemp.pred5 > 5, ]
```

```
##      glorad Otemp   wv2 dtemp.pred5
##       <num> <num> <num>       <num>
##   1:   1000    21     2    5.332794
##   2:      0    24     2    5.366631
##   3:    100    24     2    5.784040
##   4:    200    24     2    6.079520
##   5:    300    24     2    6.289129
##   6:    400    24     2    6.448925
##   7:    500    24     2    6.594966
##   8:    600    24     2    6.763309
##   9:    700    24     2    6.990013
##  10:    800    24     2    7.311135
##  11:    900    24     2    7.762733
##  12:   1000    24     2    8.380865
##  13:      0    27     2   10.241125
##  14:    100    27     2   10.658533
##  15:    200    27     2   10.954013
##  16:    300    27     2   11.163623
##  17:    400    27     2   11.323419
##  18:    500    27     2   11.469460
```

```
## 19:    600    27     2   11.637803
## 20:    700    27     2   11.864507
## 21:    800    27     2   12.185629
## 22:    900    27     2   12.637227
## 23:   1000    27     2   13.255359
## 24:      0    30     2   17.356997
## 25:    100    30     2   17.774406
## 26:    200    30     2   18.069886
## 27:    300    30     2   18.279495
## 28:    400    30     2   18.439291
## 29:    500    30     2   18.585332
## 30:    600    30     2   18.753675
## 31:    700    30     2   18.980379
## 32:    800    30     2   19.301501
## 33:    900    30     2   19.753099
## 34:   1000    30     2   20.371231
##     glorad Otemp   wv2 dtemp.pred5
```

Looks at some subsets.

```r
newdat[dtemp.pred5 > 3, ]
```

```
##     glorad Otemp   wv2 dtemp.pred5
##      <num> <num> <num>       <num>
##  1:   1000     9     2    3.105617
##  2:   1000    15     2    3.056096
##  3:    900    18     2    3.078057
##  4:   1000    18     2    3.696189
##  5:    200    21     2    3.031449
##  6:    300    21     2    3.241058
##  7:    400    21     2    3.400854
##  8:    500    21     2    3.546895
##  9:    600    21     2    3.715238
## 10:    700    21     2    3.941942
## 11:    800    21     2    4.263064
## 12:    900    21     2    4.714662
## 13:   1000    21     2    5.332794
## 14:      0    24     2    5.366631
## 15:    100    24     2    5.784040
## 16:    200    24     2    6.079520
## 17:    300    24     2    6.289129
## 18:    400    24     2    6.448925
## 19:    500    24     2    6.594966
## 20:    600    24     2    6.763309
## 21:    700    24     2    6.990013
## 22:    800    24     2    7.311135
## 23:    900    24     2    7.762733
## 24:   1000    24     2    8.380865
## 25:      0    27     2   10.241125
## 26:    100    27     2   10.658533
## 27:    200    27     2   10.954013
## 28:    300    27     2   11.163623
## 29:    400    27     2   11.323419
```

```
## 30:    500    27     2   11.469460
## 31:    600    27     2   11.637803
## 32:    700    27     2   11.864507
## 33:    800    27     2   12.185629
## 34:    900    27     2   12.637227
## 35:   1000    27     2   13.255359
## 36:      0    30     2   17.356997
## 37:    100    30     2   17.774406
## 38:    200    30     2   18.069886
## 39:    300    30     2   18.279495
## 40:    400    30     2   18.439291
## 41:    500    30     2   18.585332
## 42:    600    30     2   18.753675
## 43:    700    30     2   18.980379
## 44:    800    30     2   19.301501
## 45:    900    30     2   19.753099
## 46:   1000    30     2   20.371231
##      glorad Otemp   wv2 dtemp.pred5
```

```
newdat[dtemp.pred5 > 5, ]
```

```
##      glorad Otemp   wv2 dtemp.pred5
##       <num> <num> <num>       <num>
##  1:   1000    21     2    5.332794
##  2:      0    24     2    5.366631
##  3:    100    24     2    5.784040
##  4:    200    24     2    6.079520
##  5:    300    24     2    6.289129
##  6:    400    24     2    6.448925
##  7:    500    24     2    6.594966
##  8:    600    24     2    6.763309
##  9:    700    24     2    6.990013
## 10:    800    24     2    7.311135
## 11:    900    24     2    7.762733
## 12:   1000    24     2    8.380865
## 13:      0    27     2   10.241125
## 14:    100    27     2   10.658533
## 15:    200    27     2   10.954013
## 16:    300    27     2   11.163623
## 17:    400    27     2   11.323419
## 18:    500    27     2   11.469460
## 19:    600    27     2   11.637803
## 20:    700    27     2   11.864507
## 21:    800    27     2   12.185629
## 22:    900    27     2   12.637227
## 23:   1000    27     2   13.255359
## 24:      0    30     2   17.356997
## 25:    100    30     2   17.774406
## 26:    200    30     2   18.069886
## 27:    300    30     2   18.279495
## 28:    400    30     2   18.439291
## 29:    500    30     2   18.585332
## 30:    600    30     2   18.753675
## 31:    700    30     2   18.980379
```

```
## 32:     800    30     2   19.301501
## 33:     900    30     2   19.753099
## 34:    1000    30     2   20.371231
##      glorad Otemp   wv2 dtemp.pred5
```

```r
newdat[dtemp.pred5 > 8, ]
```

```
##      glorad Otemp   wv2 dtemp.pred5
##       <num> <num> <num>       <num>
##  1:    1000    24     2    8.380865
##  2:       0    27     2   10.241125
##  3:     100    27     2   10.658533
##  4:     200    27     2   10.954013
##  5:     300    27     2   11.163623
##  6:     400    27     2   11.323419
##  7:     500    27     2   11.469460
##  8:     600    27     2   11.637803
##  9:     700    27     2   11.864507
## 10:     800    27     2   12.185629
## 11:     900    27     2   12.637227
## 12:    1000    27     2   13.255359
## 13:       0    30     2   17.356997
## 14:     100    30     2   17.774406
## 15:     200    30     2   18.069886
## 16:     300    30     2   18.279495
## 17:     400    30     2   18.439291
## 18:     500    30     2   18.585332
## 19:     600    30     2   18.753675
## 20:     700    30     2   18.980379
## 21:     800    30     2   19.301501
## 22:     900    30     2   19.753099
## 23:    1000    30     2   20.371231
##      glorad Otemp   wv2 dtemp.pred5
```

How common were the high temperatures and temperature differences?

```r
quantile(d2.orig$Otemp)
```

```
##       0%      25%      50%      75%     100% 
##  2.99250 10.79750 13.58917 17.04000 26.62083
```

```r
quantile(d2.orig$dtemp)
```

```
##         0%        25%        50%        75%       100% 
## -3.0908333  0.0225000  0.6333333  1.5491667 13.4783333
```