Result validation: Boulder trajectories vs. DIS and PC displacement

Doris Hermle

06-11-2021

Investigation of the data and its relationships

```
library("foreign")
options(scipen = 999)
# read in data
DataDIS_II <- read.dbf("IRIS_DIS_10cm_II_corr.dbf")
DataPC_II <- read.dbf("IRIS_PC_10cm_II.dbf")
datasets <- c('DataDIS_II', 'DataPC_II')
```

```
# mahalanobis distance for multivariate outlier detection
# Leys et al., 2019 (DOI: 10.5334/irsp.289)
```

In this part, we are investigating relationship for the dense inverse search (DIS) algorithm.

```
# Dataset Dense Inverse Search
# select only the variables we neeed
varsInterest_DISII <- DataDIS_II[, c('distanz', 'MEAN')]

# run the mahalanobis distance on it
varsInterest_DISII$Mahalanobis <- mahalanobis(varsInterest_DISII,
                                              colMeans(varsInterest_DISII),
                                              cov(varsInterest_DISII))


# check p-values associated with mahalanobis distance (df is degrees of freedom and it is
# the number of variables of interest minus 1)
varsInterest_DISII$pvalue <- pchisq(varsInterest_DISII$Mahalanobis, df=1, lower.tail=FALSE)

# kick-out of outliers, run 2nd regression
# add new regression lines to the plot

# kick outliers
varsInterest_DISII$Block_ID <- DataDIS_II$Block_ID
WO_Outliers_DISII <- varsInterest_DISII[!(varsInterest_DISII$pvalue < .001),]
outliers_DISII <- varsInterest_DISII[(varsInterest_DISII$pvalue < .001),]
```

In this part, we are investigating the same relationship for the phase correlation (PC) algorithm.

```r
# Dataset Phase Correlation
# select only the variables we neeed
varsInterest_PCII <- DataPC_II[, c('distanz', 'MEAN')]

# run the mahalanobis distance on it
varsInterest_PCII$Mahalanobis <- mahalanobis(varsInterest_PCII, colMeans(varsInterest_PCII),
                                             cov(varsInterest_PCII))


# check p-values associated with mahalanobis distance (df is degrees of freedom and it is
# the number of variables of interest minus 1)
varsInterest_PCII$pvalue <- pchisq(varsInterest_PCII$Mahalanobis, df=1, lower.tail=FALSE)

# kick-out of outliers, run 2nd regression
# add new regression lines to the plot

# kick outliers
varsInterest_PCII$Block_ID <- DataPC_II$Block_ID
WO_Outliers_PCII <- varsInterest_PCII[!(varsInterest_PCII$pvalue < .001),]
outliers_PCII <- varsInterest_DISII[(varsInterest_DISII$pvalue < .001),]
```

Summary of statistics and visualisation

Here we summarise the statistics and plot the DIS data with both regression lines, with and without outliers.

```r
# regression with the outliers
lm_DIS <- lm(MEAN ~ distanz, data = DataDIS_II)
summary(lm_DIS)
```

```
##
## Call:
## lm(formula = MEAN ~ distanz, data = DataDIS_II)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7308 -1.1590 -0.9890  0.1514 12.3797
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.5119     0.6897   2.192  0.03441 *
## distanz       0.4851     0.1787   2.715  0.00982 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.148 on 39 degrees of freedom
## Multiple R-squared:  0.159,  Adjusted R-squared:  0.1374
## F-statistic: 7.372 on 1 and 39 DF,  p-value: 0.009817
```

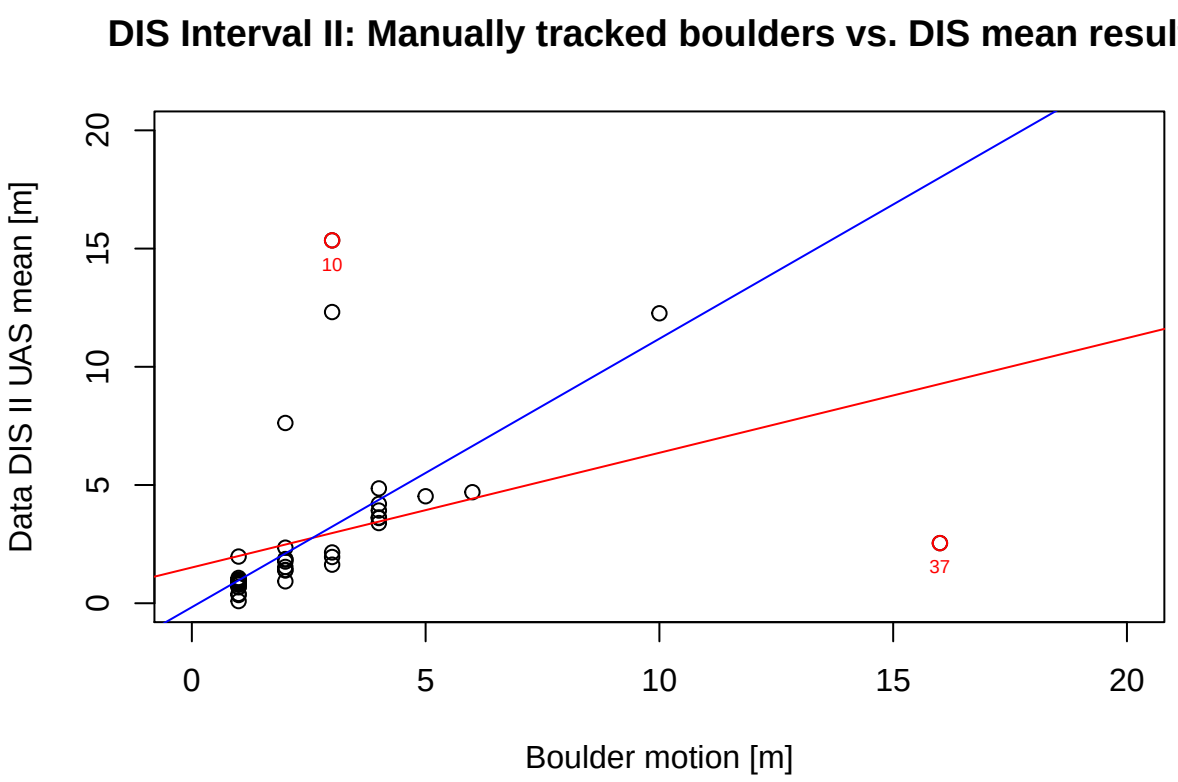
```r
# without outliers
lm_DISout <- lm(MEAN ~ distanz, data = WO_Outliers_DISII)
summary(lm_DISout)
```

```
##
## Call:
## lm(formula = MEAN ~ distanz, data = WO_Outliers_DISII)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9497 -0.7333 -0.2600 -0.0429  9.0761
##
## Coefficients:
##             Estimate Std. Error t value     Pr(>|t|)
## (Intercept)  -0.1671     0.4965  -0.337        0.738
## distanz       1.1355     0.1667   6.812 0.0000000505 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.893 on 37 degrees of freedom
## Multiple R-squared:  0.5564, Adjusted R-squared:  0.5444
## F-statistic:  46.4 on 1 and 37 DF,  p-value: 0.00000005052
```

```
# lets plot
plot(x = DataDIS_II$distanz , y = DataDIS_II$MEAN, xlab = "Boulder motion [m]",
     ylab = "Data DIS II UAS mean [m]", xlim = c(0, 20), ylim = c(0, 20),
     main = "DIS Interval II: Manually tracked boulders vs. DIS mean results")
abline(lm_DIS, col = "red") # plot regression line with outliers
abline(lm_DISout, col = "blue")
points(x = outliers_DISII$distanz , y = outliers_DISII$MEAN,
       labels = outliers_DISII$Block_ID, col="red")
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "labels" ist kein
## Grafikparameter
```

```
text(x = outliers_DISII$distanz , y = outliers_DISII$MEAN,
     labels = outliers_DISII$Block_ID, cex = 0.6, pos = 1, col = "red")
```



And here we summarise the statistics and plot the data for PC with both regression lines, with and without outliers.

```
# regression without the outliers
lm_PC <- lm(MEAN ~ distanz, data = DataPC_II)
summary(lm_PC)
```

```
##
## Call:
## lm(formula = MEAN ~ distanz, data = DataPC_II)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4874 -0.7926 -0.6362 -0.1099 20.3968
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.9972     0.7413   1.345   0.1863
## distanz       0.5500     0.1920   2.864   0.0067 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.384 on 39 degrees of freedom
## Multiple R-squared:  0.1738, Adjusted R-squared:  0.1526
## F-statistic: 8.203 on 1 and 39 DF,  p-value: 0.006699
```

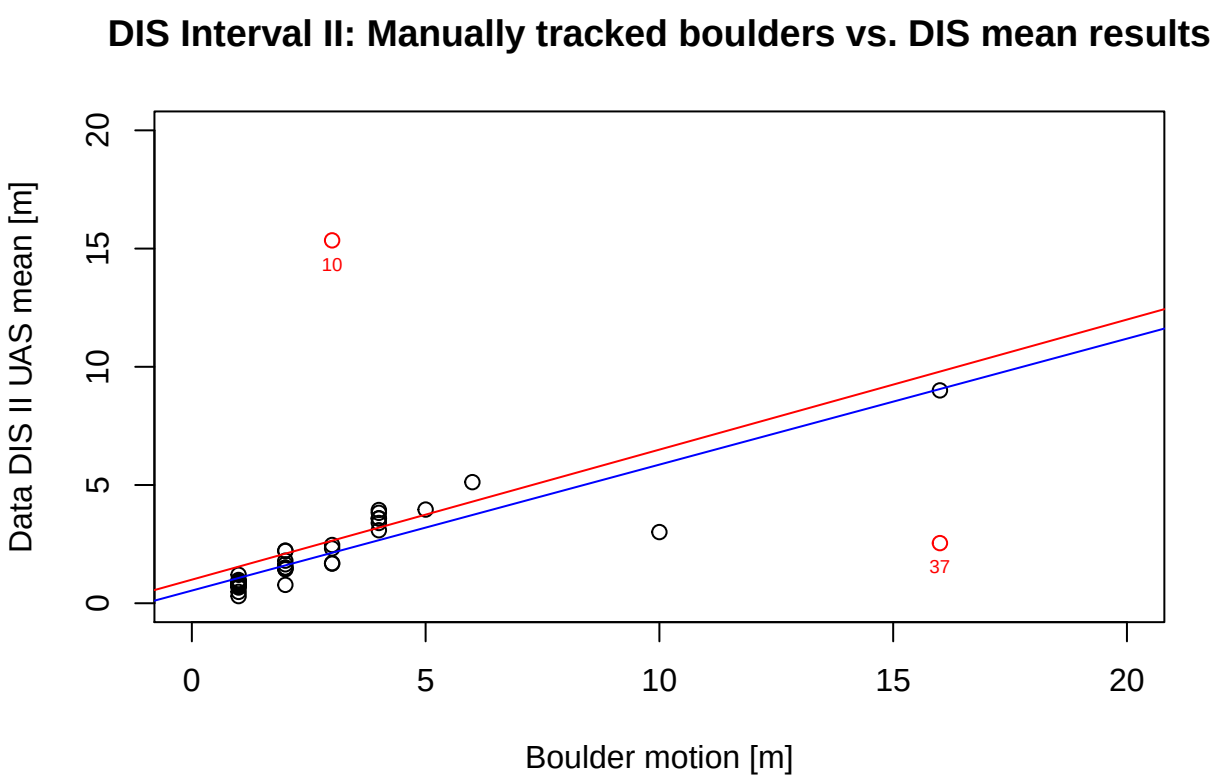
```r
# without outliers
lm_PCout <- lm(MEAN ~ distanz, data =WO_Outliers_PCII)
summary(lm_PCout)
```

```
##
## Call:
## lm(formula = MEAN ~ distanz, data = WO_Outliers_PCII)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8528 -0.2943 -0.1443  0.3753  1.3899
##
## Coefficients:
##             Estimate Std. Error t value       Pr(>|t|)
## (Intercept)  0.53537    0.19262   2.779        0.00851 **
## distanz      0.53272    0.06467   8.238 0.000000000683 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7344 on 37 degrees of freedom
## Multiple R-squared:  0.6471, Adjusted R-squared:  0.6376
## F-statistic: 67.86 on 1 and 37 DF,  p-value: 0.0000000006827
```

```r
# lets plot
plot(x = DataPC_II$distanz , y = DataPC_II$MEAN, xlab = "Boulder motion [m]",
     ylab = "Data DIS II UAS mean [m]", xlim = c(0, 20), ylim = c(0, 20),
     main = "DIS Interval II: Manually tracked boulders vs. DIS mean results")
abline(lm_PC, col = "red") # plot regression line with outliers
abline(lm_PCout, col = "blue")
points(x = outliers_PCII$distanz , y = outliers_PCII$MEAN,
       labels = outliers_PCII$Block_ID, col="red")
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "labels" ist kein
## Grafikparameter
```

```r
text(x = outliers_PCII$distanz , y = outliers_PCII$MEAN,
     labels = outliers_PCII$Block_ID, cex = 0.6, pos = 1, col = "red")
```


DIS Interval II: Manually tracked boulders vs. DIS mean results
Data DIS II UAS mean [m]
Boulder motion [m]
0
5
10
15
20
10
37