PAM Correction – Cropping of extremes

Andreas Schönberg M.Sc.

Method

- Using the best resulting model for correction (type and dialect).
- Detect extremes by absolute difference which occur due to the correction.
- Crop those extremes from the original data.
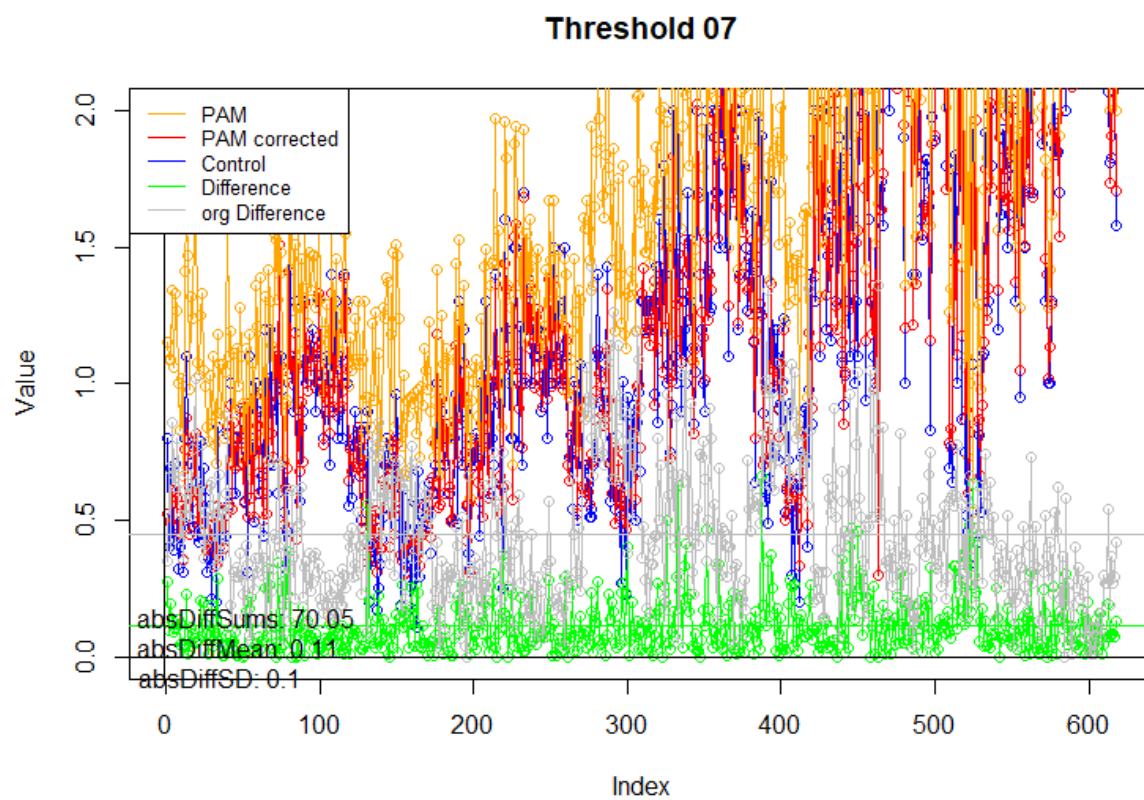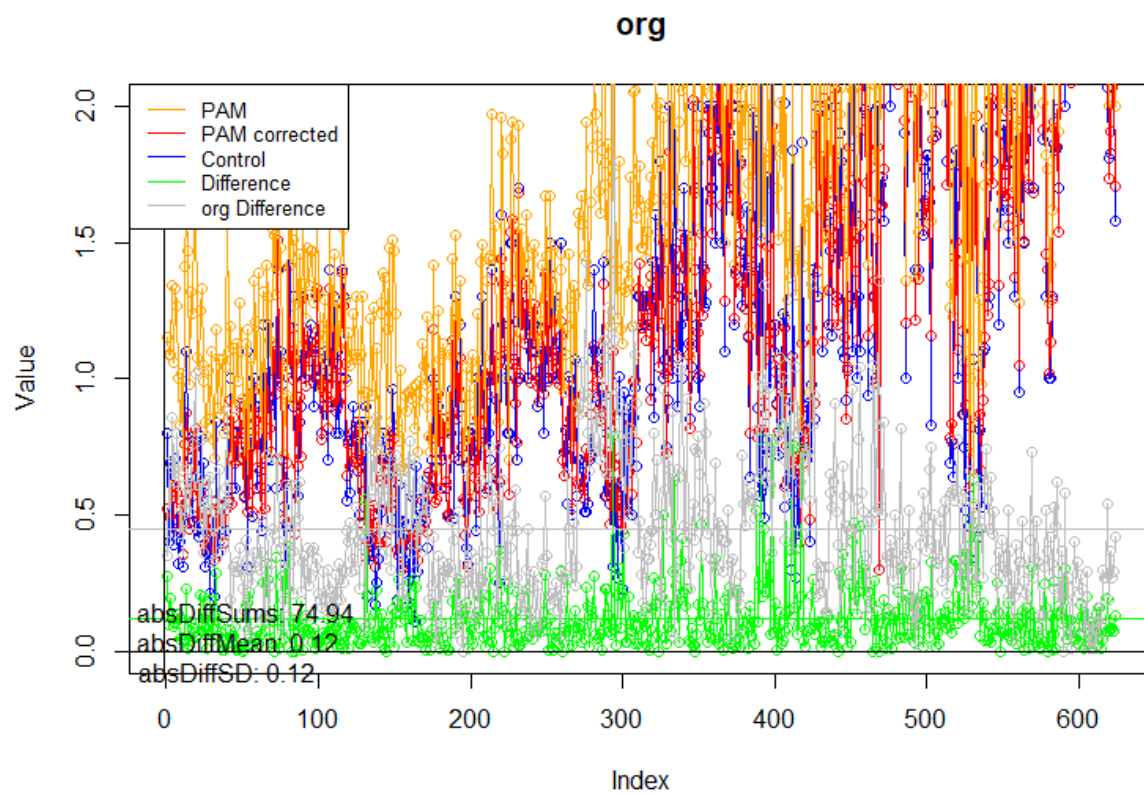- Again use the best model on cropped data.

Results

Cropping the values which lead to extremes after the correction, result in an improved perfromance of the model. By cropping the values which lead to differneces of more than 0.4 (up to 20 entries) the perfromance is slightly improved (6.5% to 16%). For data cropped by a threshold of 0.3 aswell as 0.2 the resulting models reach 29% up to 51% increase in perfromance.

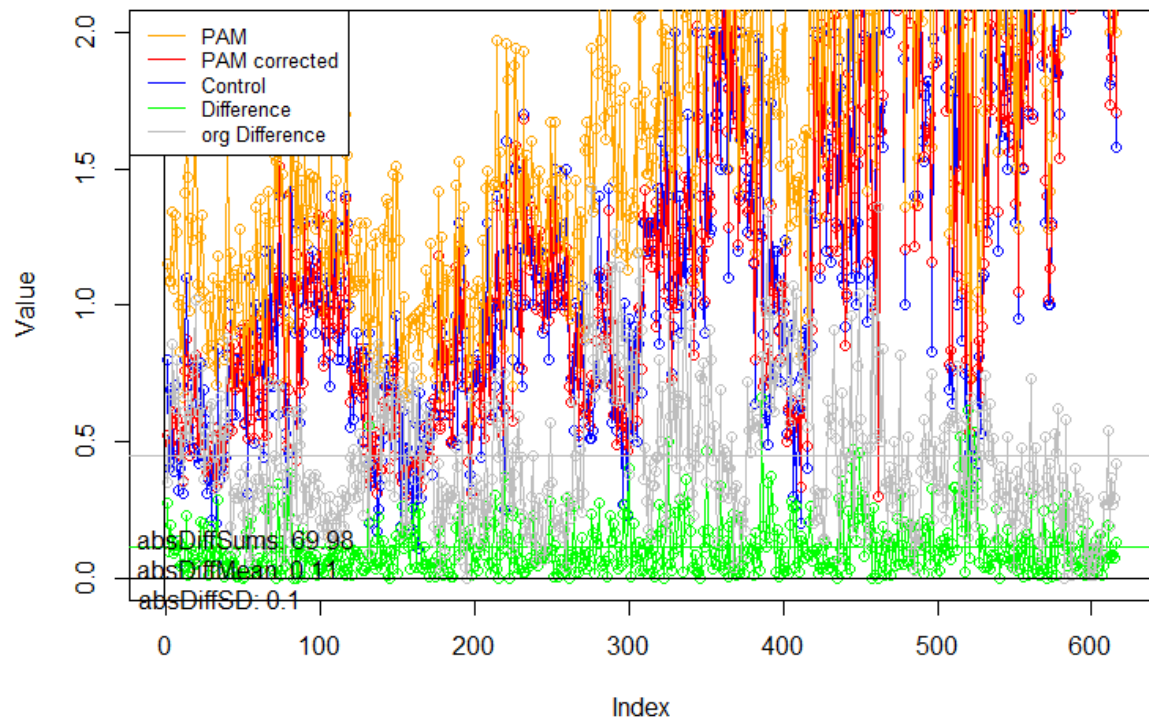*Table 1:* *Cropped extremes by threshold and error values.*

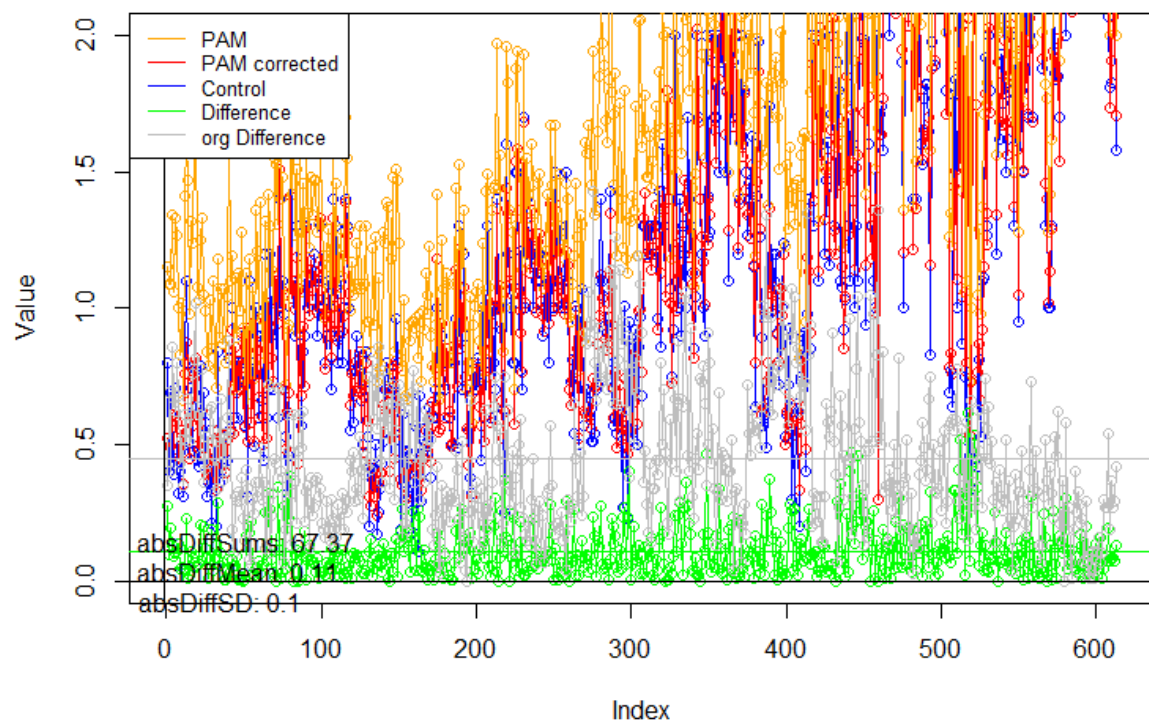| Threshold	Entries deleted	Error total Sum	Error Mean	Error SD	Max Error (Extrema)	Improvement (based on total error)
No (original)	-	74.94	0.12	0.12	0.82	-
0.7	6	70.05	0.11	0.10	0.65	6.5 %
0.6	8	69.98	0.11	0.10	0.65	6.6 %
0.5	11	67.37	0.11	0.10	0.61	10.0 %
0.4	20	62.39	0.11	0.09	0.42	16.7 %
0.3	43	52.93	0.09	0.07	0.30	29.3 %
0.2	112	36.10	0.07	0.05	0.27	51.8 %

Discussion

Lesser data entries lead to decreasing amount of values which are used to calculate the means for subgroups in the model. While this leads to better result it although leads to lesser usability for future generated PAM values due to lesser information which the model is based on. The results for cropping by a threshold of <0.4 seems overoptimistic.
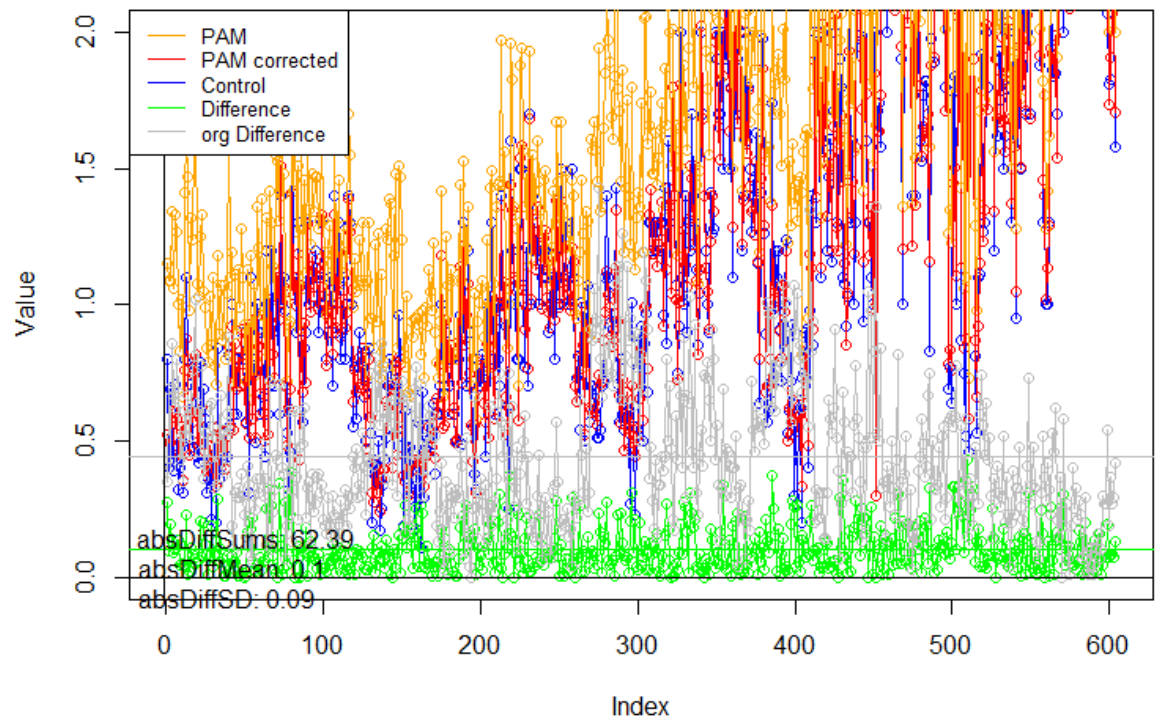
org

Threshold 07

Threshold 06

Legend:
- PAM (orange)
- PAM corrected (red)
- Control (blue)
- Difference (green)
- org Difference (grey)

absDiffSums: 69.98
absDiffMean: 0.11
absDiffSD: 0.1

Value

Index

Threshold 05

Legend:
- PAM (orange)
- PAM corrected (red)
- Control (blue)
- Difference (green)
- org Difference (grey)

absDiffSums: 67.37
absDiffMean: 0.11
absDiffSD: 0.1

Value

Index

Threshold 04

Legend:
- PAM (orange)
- PAM corrected (red)
- Control (blue)
- Difference (green)
- org Difference (grey)

absDiffSums: 62.39
absDiffMean: 0.1
absDiffSD: 0.09

Threshold 03

Legend:
- PAM (orange)
- PAM corrected (red)
- Control (blue)
- Difference (green)
- org Difference (grey)

absDiffSums: 52.93
absDiffMean: 0.09
absDiffSD: 0.07

Threshold 02