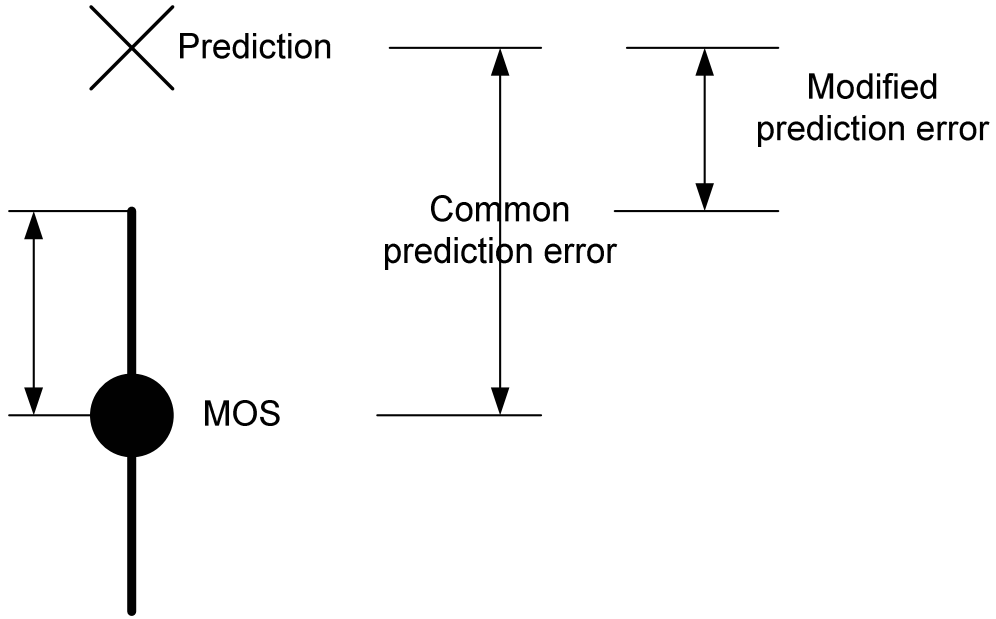
Epsilon-insensitive r.m.s.e

Basic idea: Taking into account the ci95 confidence interval for calculating prediction erros and resulting rmse

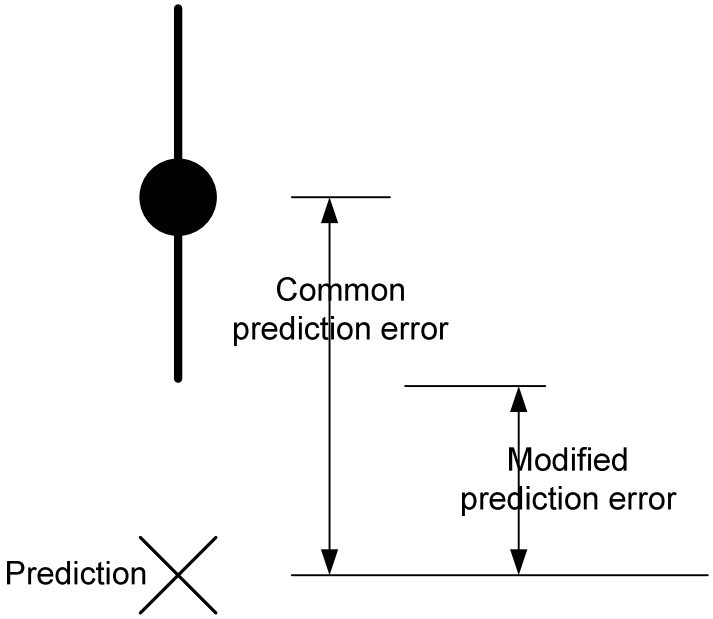


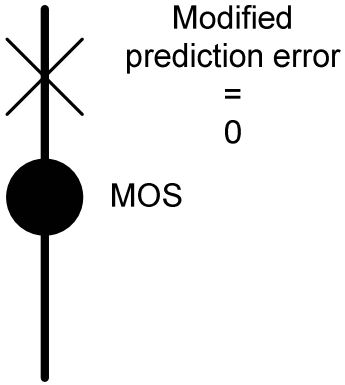
SwissQual

Jens Berger
jens.berger@swissqual.com

Epsilon-insensitive r.m.s.e

Basic idea: Taking into account the ci95 confidence interval for calculating prediction erros and resulting rmse



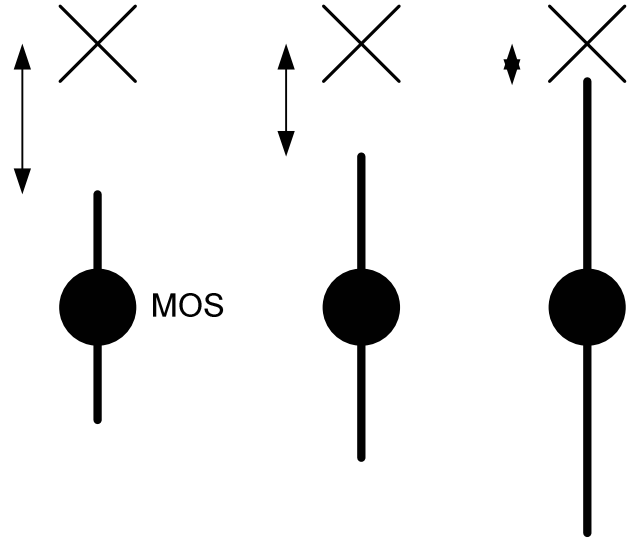


SwissQual

Jens Berger
jens.berger@swissqual.com

Epsilon-insensitive r.m.s.e

The Prediction Error becomes reduced in case the MOS is more inconfident (i.e. due to less votes or wide distribution of votes)



SwissQual

Jens Berger
jens.berger@swissqual.com

How to calculate rmse*?

$$Perror(i) = \max(0, \left|MOSLQS(i) - MOSLQO(i)\right| - ci_{95}(i))$$

$$rmse^* = \sqrt{\left(\frac{1}{N-d}\sum_{N} Perror(i)^2\right)}$$

Remark: There are some worth special rules in case the ci95 ➔0 at the scale boundaries.

SwissQual

Jens Berger
jens.berger@swissqual.com

How to compare models?

For each data set and model a Distance *d* between the best performing model and the others are calculated. It considers statistical significance too.

$$d_{k,v} = \max(0, rmse*_{k,v}{}^2 - rmse*_{k,b}{}^2 \times \mathrm{F}(0.05, N_k, N_k))$$

SwissQual

Jens Berger
jens.berger@swissqual.com

How to compare models?

The Distances for each model are averaged across the datasets. If desired the data sets can be weigthed.

Finally, models can be selected those are the best and statistically equivalent.

$$p_v = \sum_{k=1}^{M} w_k \times d_{k,v}$$

$$t_v = \max(0, \frac{p_v}{\left(p_{\min} + c\right)} - \mathrm{F}(0.05, K\ , K\ ))$$

SwissQual

Jens Berger
jens.berger@swissqual.com