Model Reference Analysis Tool

# Validation analysis

## Correlation coefficient

The correlation coefficient in the Validation analysis is calculated over:

* Observed values subsetted to dates with modeled values
* Modelled values subsetted to dates with observed values

These values are log transformed if required.

In the plot the modeled values are plotted against the observed values. In an optimal simulation this should give a straight line of measurements. To this plot an abline is added which is based on the Line of Best Fit. And finally the R-squared is added in top, indicating the percentage of the model that is predicted by the Line of Best Fit.

## Target diagram

The target diagram is calculated over:

* Observed values subsetted to dates with modeled values
* Modelled values subsetted to dates with observed values

These values are log transformed if required. Depending on the chosen technique the target diagram can also be calculated on the best fitting simulation value within a search window from the observed value or the averaging of simulation values within the search window form the observed value.

In the plot the unbiased Root Mean Square Difference and the Bias is plotted against each other. The unbiased Root Mean Square Difference gives an indication on the precision of the simulation. For example, when the model follows the trend of the observations, but is continuously under predicting, the precision is correct but there is a bias. When the model stays around the observations, but is not following the trend than there is no to a small bias but the precision is off.

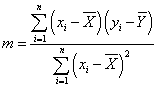
## Line of Best Fit

<http://hotmath.com/hotmath_help/topics/line-of-best-fit.html>

1. Calculate the mean of the observed values and the mean of the modeled values

1. Than substract the observed and the modeled values by their mean, the result is summed.
2. Than devide the summation by the result of the summed observed values mines their mean.
3. This results in *m*, which is the slope and is going to be an indicator for the starting point of the line.



1. Now compute the starting point of the line by substracting the mean of the simulated values by m times the mean of the observed values. This gives the starting point on the Y axis at X = 0.

http://hotmath.com/hotmath_help/topics/line-of-best-fit/b.gif

1. The formula of the line results in y = mx + b . M is also the measure for the correlation coeifficient.

## R-squared

<https://www.khanacademy.org/math/probability/regression/regression-correlation/v/calculating-r-squared>

1. After calculating the Line of Best Fit we compare the simulated value predicted by the Line of Best Fit to the actual simulated value. The actual simulate value minus the simulated value predicted by the line and then square the result. This gives the standard error that is not explained by the line.



1. Now we also calculate the total error (which induces the error by the variation in the observed values). This is calculated by the actual simulate value minus the mean of the actual simulated value and then square the result.



1. Now we would like to know the percentage of the error that is not explained by the Line of Best Fit. Therefor we divide the unpredicted error by the total error. To come to R-squared we sustract he obtained result form 1. This R-squared value should be close to 1.



1. The result is indicating what percentage of the simulated values is explained by the Line of Best Fit. The highest number is 1 (100%).

## Normalised Unbiased Root Mean Square Difference (uRMSD)

1. First calculate the standard deviation between the observed values and simulated values.



1. Determine whether the standard deviation of the simulation is larger or smaller than the standard deviation of observed.

Give -1 or 1 as indicator by a negative or positive result from:



1. Now calculate the unbiased Root Mean Square Difference. The uRMSD is an positive value when the standard deviation of the simulated values is greater than the observed values and an negative value when visa versa. The indicator can be 1 or -1:



1. After calculating the uRSMD the result still needs to be standardized against the observed values:



## Bias (nBias)

1. The Bias is calculated by subtracting the mean of the simulated values with the mean of the observed values. The result is divided by the standard deviation of the observed values:

**