**Indicators for model testing**

The mean bias error (MB) reflects the systematic deviation between observed (Y) and predicted values (X):



where *Y*i represents the ith measured value, *X*i corresponds to the ith model predicted value, and *n* is the number of observations.

The root mean squared error of prediction (RMSE) characterises the spread between observed and simulated values:



The normalised root mean squared error of prediction (RMSPE) also characterises the spread between observed and simulated values, but relative to the spread of observed values:



where SD is the standard deviation of measurements calculated as:



where is the mean of the observations.

The coefficient of determination (r2) is determined as:



Model efficiency (ME) is determined as:



High model efficiency, for a specific attribute, is represented by a ME value of 1. In contrast, an ME value that is <0 indicates that the mean of the measured values is a better estimator than the model-predicted values (Smith et al., 1997). The coefficient of determination is also a useful metric for assessing model precision. For any attribute, an r2 value of 1 indicates that there is a strong correlation between the simulated and the measured quantities and we say that the model is precise. Conversely, the lowest precision level is represented by 0. It is, however, important to stress that there are inherent limitations of the r2 statistic since a highly precise and inaccurate model does not predict the measured values well and hence less emphasis was placed on this statistic when model predictions were inaccurate. The RMSPE is inversely related to r2. The mean bias (MB) is also used to assess model accuracy, but reflects systematic deviations between observed and simulated values. If MB is negative, it means that, for the specific variable, the model overestimates and, if it is positive, it means that the model underestimates (Tedeschi, 2006).

**References**

Smith P, Smith JU, Powlson DS, McGill WB, Arah JRM, Chertov OG, Coleman K, Franko U, Frolking S, Jenkinson DS, Jensen LS, Kelly RH, Klein-Gunnewiek H, Komarov AS, Li C, Molina JAE, Mueller T, Parton WJ, Thornely JHM, Whitmore AP (1997) A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. Geoderma 81:153−225.

Tedeschi LO (2006) Assessment of the adequacy of mathematical models. Agric Syst 89:225−247.