Appendix F: Model to Data Comparisons

Appendix F-4: Statistical metrics definition

This appendix is in connection with Section 3.2.4.1.2 in the main report – Model-data comparison for individual Boat Run locations. It is from the Appendix B in the document "Instruction manual for R-calibration scripts", by John M. Davis, USEPA Region 4. 2019.

Appendix B:

Primer on interpreting goodness of fit statistics

Quantitative analysis of model fit can be assessed using several widely used goodness of fit statistics (Moriasi et al. 2007) that were calculated using the 'hydroGOF' package (Zambrano-Bigiarini 2017). An overview of the various statistical comparisons is provided below.

- Arithmetic Mean ($\overline{x}$) – On average, assesses how well the simulated values represent observed values. For both the observed and simulated dataset, an arithmetic mean is calculated for each parameter across the entire model simulation period.

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- Percentiles – This verifies the model is reasonably predicting extreme values in the observed data. Entails inspection of the 10^{th} and 90^{th} percentiles of both observed and simulated data.

- Mean Error (ME) – For each pair of measured and simulated values, measures the average difference (i.e., error) between observed and simulated data. Does not indicate if the simulated value is over or underpredicting the observed value and does not consider the natural variation in the observed data. For each paired observed and simulated record, the difference of the observed and simulated value is calculated, and subsequently averaged.

$$\text{ME} = \frac{1}{N}\sum_{i\,=\,1}^{N}(S_i - O_i)$$

- Mean Absolute Error (MAE) – Measures the average magnitude of the difference (i.e., error) between observed and simulated data. It does not consider the direction of those differences (i.e., whether the model is over or underpredicting) or natural variation in the observed data. Calculated similarly to Mean Error, but the absolute value of the difference is taken.

$$\text{MAE} = \frac{1}{N}\sum_{i\,=\,1}^{N}|S_i - O_i|$$

- Root Mean Square Error (RMSE) – Measures the difference (i.e., error) between observed and simulated data. This metric provides assurance that the model is matching the frequency, magnitude, and duration of water quality changes. However, it does not account for natural variability in observed data. Values of RMSE range from 0 to infinity, with RMSE = 0 indicating a perfect match between observed and simulated data.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i\,=\,1}^{N}(S_i - O_i)^2}$$

- Normalized Root Mean Square Error (NRMSE %) – Similar to RMSE; however, error is standardized relative to the range of the observed data. NRMSE is reported as a percent. Values of NRMSE range from -100% to 100%, with NRMSE = 0% indicating a perfect match.

$$\text{NRMSE\%} = 100 \times \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2}}{O_{Max} - O_{Min}}$$

- Coefficient of determination (R^2) – Assesses the strength of the linear relationship between observed and simulated data. Describes the proportion of variation in the observed data that is explained by a simple linear regression relating observed and simulated data. Values of R^2 range from 0 to 1, with better fitting models possessing higher R^2 values.

$$R^2 = \left[\frac{\sum_{i=1}^{N}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2} \times \sqrt{\sum_{i=1}^{N}(P_i - \overline{P})^2}}\right]^2$$

- Spearman Rank Correlation Coefficient (r) – Conceptually, similar to simple linear regression, but the relationship between simulated and observed values is assessed based on their rank value (i.e., highest value given a rank of 1). As the comparison is nonparametric, data do not need to meet assumptions of normality and equal variance. Values range from -1 to 1, with r = -1 indicating a perfect negative relationship between simulated and observed data and r =1 indicating a perfect positive relationship.

$$r_s = \frac{covariance(rg_{Sim}, rg_{Obs})}{\sigma_{rg_{Sim}}\sigma_{rg_{Obs}}} = \frac{\sum_{i=1}^{N}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2} \times \sqrt{\sum_{i=1}^{N}(P_i - \overline{P})^2}}$$

- Percent Bias (PBIAS) – Provides a measure of whether a model, on average, tends to over- or underestimate observed values. The magnitude of the difference in observed and simulated data is calculated relative to the sum of observed data. Values range from -100% to 100%, with more accurate models exhibiting PBIAS that approach 0%. Values of PBIAS > 0% indicates that the model is overestimating observed values, while PBIAS < 0% indicates the model is underestimating them.

$$\text{PBIAS\%} = 100 \times \frac{\sum_{i=1}^{N}(S_i - O_i)}{\sum_{i=1}^{N} O_i}$$

- Nash-Sutcliffe Coefficient (NSE) – This metric is closely related to mean square error and root mean square error. Using the mean of the observed data as a baseline, it assesses the magnitude of the difference in observed and simulated data relative to residual variance (i.e., natural variation) of observed data. This unitless metric indicates how well the linear fit of observed versus simulated data fits a 1:1 line. Values range from -Infinity to 1, whereby NSE = 1 represents a perfect match of simulated and observed data, NSE = 0 indicates that model predictions are as accurate as the mean of observed data, while NSE = -Infinity indicates that the mean of observed values is a better predictor than simulated data.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}$$

- Index of Agreement (d) – Provides a measure of model error relative to natural variability (i.e., error). Values range from 0 to 1, with an index of agreement = 1 indicating a perfect fit of simulated and observed data, and a value of 0 indicating no agreement between them.

$$\text{d} = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(|S_i - \bar{O}| + |O_i - \bar{O}|)^2}$$

- Modified Kling-Gupta Efficiency (KGE') – This unitless metric is similar to the Nash-Sutcliffe Coefficient (NSE) but attempts to assess fit by assigning equal weight to correlation, bias, and variability metrics. In contrast, optimal models indicated by NSE have the potential to overemphasize the linear correlation component and can exhibit reduced effectiveness when assessing highly seasonal parameters.

  Accordingly, KGE' is decomposed into three components, which are simultaneously reported out with the KGE'. It uses a Pearson Correlation Coefficient (r) to assess the linear correlation between measured and simulated values, the ratio of simulated means vs. observed means (β) to assess model bias, and the ratio of coefficient of variations (γ) to assess model variance. This decomposition helps to indicate whether a model is reproducing temporal dynamics (as represented by r) as well as flow distributions (as represented by β and γ). Values range from -Inf to 1, whereby an ideal fitting model is indicated by KGE' = 1.

$$\text{KGE}' = 1 - \sqrt{((\text{r} - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2)}$$

  - Kling-Gupta Pearson Correlation Coefficient (r) – The Pearson correlation coefficient between simulated and observed values. In contrast to the nonparametric Spearman rank correlation coefficient, the parameteric Pearson correlation coefficent is calculated based on the actual paired values rather than their ranked values. Values range from -1 to 1 and a perfectly fit model is indicated by r =1.

$$r_p = \frac{covariance(Sim, Obs)}{\sigma_{\text{Sim}}\sigma_{\text{Obs}}} = \frac{\sum_{i=1}^{N}(O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2} \times \sqrt{\sum_{i=1}^{N}(P_i - \overline{P})^2}}$$

  - Kling-Gupta Beta (β: Ratio of means) – The ratio of the simulated mean to the observed mean. An ideal model has β = 1.

$$\beta = \frac{\mu_s}{\mu_o}$$

- Kling-Gupta Gamma (γ : Ratio of coefficient of variation) – The ratio of the simulated coefficient of variation relative to the observed coefficient of variation. An ideal model has γ = 1.

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s / \mu_s}{\sigma_o / \mu_o}$$