Geophysical Research Abstracts
Vol. 19, EGU2017-3574, 2017
EGU General Assembly 2017

Mean absolute error and root mean square error: which is the better metric for assessing model performance?

Gary Brassington
Environment and Research Division, Bureau of Meteorology, Australia (g.brassington@bom.gov.au)

The mean absolute error (MAE) and root mean square error (RMSE) are two metrics that are often used interchangeably as measures of ocean forecast accuracy. Recent literature has debated which of these should be preferred though their conclusions have largely been based on empirical arguments. We note that in general,

$$RMSE^2 = MAE^2 + VAR_k\left[|\varepsilon|\right]$$

such that RMSE includes both the MAE as well as additional information related to the variance (biased estimator) of the errors ε with sample size k. The greater sensitivity of RMSE to a small number of outliers is directly attributable to the variance of absolute error. Further statistical properties for both metrics are derived and compared based on the assumption that the errors are Gaussian. For an unbiased (or bias corrected) model both MAE and RMSE are shown to estimate the total error standard deviation to within a constant coefficient such that

$$MAE \approx \sqrt{2/\pi}RMSE$$

. Both metrics have comparable behaviour in response to model bias and asymptote to the model bias as the bias increases. MAE is shown to be an unbiased estimator while RMSE is a biased estimator. MAE also has a lower sample variance compared with RMSE indicating MAE is the most robust choice. For real-time applications where there is a likelihood of "bad" observations we recommend

$$TESD = \sqrt{\frac{\pi}{2}}MAE \pm \frac{1}{\sqrt{k}}\sqrt{\frac{\pi}{2}-1}\sqrt{\frac{\pi}{2}}MAE$$

as an unbiased estimator of the total error standard deviation with error estimates (one standard deviation) based on the sample variance and defined as a scaling of the MAE itself. A sample size (k) on the order of 90 and 9000 provides an error scaling of 10% and 1% respectively. Nonetheless if the model performance is being analysed using a large sample of delayed-mode quality controlled observations then RMSE might be preferred where the second moment sensitivity to large model errors is important. Alternatively for model intercomparisons the information might compactly represented by a graph with axes of MAE

and $\sqrt{VAR_k\left[|\varepsilon|\right]}$

where radials from the origin represent RMSE

.