Geosci. Model Dev. Discuss., 7, C192–C195, 2014
www.geosci-model-dev-discuss.net/7/C192/2014/

Geoscientific Model Development
Discussions
Open Access

Interactive comment on "Root mean square error (RMSE) or mean absolute error (MAE)?" *by* T. Chai and R. R. Draxler

Anonymous Referee #1

Received and published: 25 March 2014

General comments

This is the third time that I have been asked to review a version of this paper, twice previously for the journal Atmospheric Environment which ultimately elected not to publish this paper. The authors of "Root mean square error (RMSE) or mean absolute error (MAE)?" argue that Willmott and collaborators' recommendations–to preferentially use the MAE, rather than the RMSE, as the appropriate measure of average error magnitude in model performance evaluations (articulated in papers published in 2005 and 2009)–is "not the solution to the problem." They claim to ". . .demonstrate that the RMSE is not ambiguous in its meaning. . .and. . .[it] is more appropriate to represent model performance than the MAE when the error distribution is expected to be Gaussian." They also report "that the RMSE satisfies the triangle inequality requirement

for a distance metric," contrary to interpretations made by Willmott et al. For reasons outlined below, most of their arguments are not compelling and I recommend that this paper not be accepted for publication.

If published, this paper will tend to mislead uncritical researchers about the use of the RMSE in the statistical evaluation of model performance. The only redeeming quality of this paper is that it will serve as a good example to cite when pointing out the all-too-common under appreciation of problems associated with the RMSE.

Specific comments

This manuscript falls well short of refuting Willmott et al.'s main points about the interpretational limitations associated with the RMSE; because its value is a function of three variables (the MAE, the variability among the error magnitudes and, in instances, the square root of n). Even though the authors claim to show that the RMSE is not ambiguous, they do not. They appear unable to appreciate that, if one does not know what portion of an RMSE is average error magnitude and what portion is error magnitude variability, it (an RMSE) really is "ambiguous" and impossible to adequately interpret or meaningfully compare with other RMSEs. Without additional information, all that can be said about the RMSE is that it inconsistently overestimates average error magnitude. And, if one knows the values of the above-mentioned variables that comprise an RMSE, it (the RMSE) provides no additional information and is superfluous. If, in addition to the average error magnitude (the MAE), variability among the error magnitudes is of interest, a measure of error magnitude variability should be calculated, reported and interpreted in addition to the MAE, as suggested by Willmott et al. (2009). The MAE, in contrast to the RMSE, is a straightforward measure of average error magnitude.

The authors want to assume errors are normally distributed and unbiased, and this seems to be their primary argument for evaluating the RMSE. It's difficult to understand why anyone would want to assume normality when they do not have to. Nonetheless, the authors say that “when the error distribution is expected to be Gaussian and there are enough samples, the RMSE has an advantage over the MAE to illustrate the error distribution.” This is a red herring, for at least three reasons. First, their example (see their Table 1) clearly shows that the MAE does equally well at recovering their (“Gaussian”) error distribution. Second, whether their preconditions are satisfied adequately is nearly always questionable. Third, regardless of whether these preconditions satisfied, one still should know the average error magnitude (the MAE), which cannot be teased out of the RMSE post hoc. In addition, let me mention that “modern statistics” increasingly tends to emphasize interpretability over “exact properties.” Subfields such as “exploratory data analysis” and “robust statistics” have produced a number of innovative approaches that enhance interpretability (e.g., through resistance to outliers or insensitivity to non-normal distributions) more than mathematical exactness. Classic texts by Tukey (1977) and Huber (1981), for example, introduce some intriguing approaches and I encourage the authors to explore such more flexible options.

Even though the triangular inequality issue is a relatively minor one, it is unfortunate that the authors of this paper misinterpret Willmott et al.’s correct point about combinations of elements within the squared-errors vector not satisfying the triangular inequality. Although Willmott et al. explained their invocation of the triangular inequality; perhaps their brief descriptions of this (in both 2005 and 2009) were insufficient, because the authors of this manuscript have misinterpreted them. In the 2005 paper, for instance, Willmott and Matsuura state that “…counter-intuitive values of RMSE are expected because |e_i|ˆ2 (or e_iˆ2) is not a metric; that is, |e_i|ˆ2 does not satisfy the triangle inequality of a metric (Mielke & Berry 2001).” In the 2009 paper, Willmott et al. similarly indicate that “…the relative influence of each squared error on the sum of the squared errors often is counterintuitive [in part, because of likely triangular inequalities among the individually squared errors], which undermines meaningful scientific interpretation of the sum of the squared errors” and, in turn, of the RMSE. In other words, relationships among the squared errors (within the set of squared errors from which the RMSE is evaluated) that do not satisfy the triangular inequality, when the squared

errors are summed, can help produce nonsensical sums of squares and RMSEs.

Interactive comment on Geosci. Model Dev. Discuss., 7, 1525, 2014.