Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-51-RC2, 2020

Geoscientific Model Development Discussions
Open Access

GMDD

Interactive comment

Interactive comment on "From R-squared to coefficient of model accuracy for assessing "goodness-of-fits"" *by* Charles Onyutha

Barry Croke (Referee)

barry.croke@anu.edu.au

Received and published: 1 November 2020

General comments

This paper explores how we go about measuring the "goodness of fit", proposing an alternative measure to R^2. The author cites Kvålseth's 1985 paper (Cautionary Note and R^2), but fails to acknowledge that there are more than 1 measures that are commonly used which are called R^2 (or the coefficient of determination). The most common of these are:

1. R^2 for a linear regression (square of the Pearson Product Moment Correlation Coefficient – note that for a sample of a population, this is traditionally called r). This is R_6^2 in Kvålseth (1985)

Printer-friendly version

Discussion paper

2. NSE (called R^2 in the Nash Sutcliffe paper – note that the Nash Sutcliffe paper built from this existing definition of R^2 to propose a way of evaluating incremental changes in a model by replacing the mean observed flow in the denominator with the previous version of the modelled flow – which they called r^2 to discriminate this from what Hydrologists refer to NSE). This is R_1^2 in Kvålseth (1985).

3. Fraction of variance explained (1-variance of model residuals/variations of observed output). This is identical to NSE if the bias is zero. It is very close to NSE for any reasonable model (i.e. small bias), though this measure will always be larger than NSE (indicating a slightly better fit) if the bias is not zero. This is R_4^2 in Kvålseth (1985).

The fact that R^2 (or the coefficient of determination) is multiply defined leads to a lot of confusion, and authors need to be clear about which version they are using, and readers need to ensure that they understand which version is being referred to. While there is growing use of the R^2 for a linear regression in Hydrology (something that really needs to stop), many of the papers that refer to R^2 or the coefficient of determination are actually using NSE. My view is that the linear regression version adds little to NSE, and if the linear regression form is used, the slope and intercept of the regression should always be reported as well.

The author also does not report on the conclusion of the analysis done by Kvålseth – that generally the best form of R^2 appears to be what hydrologists refer to as the NSE.

The author refers to example hydrology papers that have used R^2. Looking at some of these papers, they appear to use the NSE version of R^2. However, the author also refers to the regression-based version of R^2, specifically in the statement: "Regressing X on Y yields R^2 which is the same as that if Y is regressed on X thereby invalidating its use as a coefficient of determination" in the abstract and body of the paper. These two forms of R^2 are not equivalent (though related in the unbiased case – see Bardsley, 2013, Hydrol. Proc. DOI: 10.1002/hyp.9914), and the author needs to clarify which R^2 is being discussed (i.e. focus on NSE) and remove the discussion on the linear regression. Note that for NSE, the value doesn't remain the same if you swap the observed and modelled values, so the criticism of the linear regression form is irrelevant in regard to the NSE form.

Overall, this confusion about which R^2 is being talked about distracts the reader from the message the author is trying to convey. Subsequently the introduction requires extensive revision. The author would be better served by not focusing on the R^2 for a linear regression (other than to state that this should not be used), and focus on the NSE form of R^2. This would substantially change the discussion part of the paper also (as well as impacting on the abstract and conclusion).

Given the issue with the discussion of R^2 in the paper, there is merit in the coefficient of model accuracy being proposed by the author. Each performance criteria gives a different view of the model behavior. Some (e.g. NSE and RMSE) are very similar (in the case based on the sum of squared residuals), so little is gained from using both of these. The CMA proposed by the author is different from any performance criteria I have seen, being more closely related to the Spearman correlation coefficient (noting the penalty based on comparison of individual observed and modelled values), and therefore may be of benefit for modellers.

I spotted a few typos in the paper, but didn't check specifically for these. The author should carefully check the paper for errors.

Specific comments

1. Line 64: I find the use of the upper case to indicate the series of values and the lower case with subscript I to represent individual values in the series an unnecessary complication. Adding the subscript indicates that you are looking at individual values, so the data set could be lower case without confusion. If you want to use the upper case for the series, then logically, the mean values in the expressions should also be upper case as these indicate the mean value of the series). At the moment, this leads to a confusing mixture of notation in the paper (e.g. line 88, which refers to the dataset

Y, and gives values of y=. . .). It would make more sense to refer to one as the original (untransformed values), and the other as the transformed values (e.g. x=X-mean(X); y=Y-mean(X)).

2. Line 105-109: Presumably, the author is using x=observed-mean observed, and y=modelled-mean observed?

3. Line 113: Maybe I am missing something, but with the subscript i includes in equations 9 and 10, isn't min(h_i,x_i)=max(h_i,x_i) as h_i and x_i are scalars? This is pivotal for calculating the β, and so needs to be clarified. Also in equation 10, I assume this should be $\omega_{2,i}$?

4. Line 116: Why not use the absolute value rather than the squared value to handle the issue of opposing signs? Is there any reason for the choice, or is it just a personal preference? What impact does this have on the result?

5. Line 152: MAB is defined as the model average bias. However, the inclusion of the denominator means this is not an average bias. Please check the formula given in the paper. It should also be noted that this formulation is problematic if the observed values are too close to zero. Is this based on a published formation? If so, please add a citation.

6. Line 172: either "poorly fitted model" or "poor model fit"

7. Line 176: "tend to zero"

8. 190: Yes, R^2 for a linear regression can be high for poor models, if the model error can be well modelled by a linear relationship with the observed values. This is why the R^2 for a linear regression should never be used (unless you also report the slope and intercept). Again, I point out that many hydrological papers that refer to R^2 are using NSE rather than the R^2 for a linear regression.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-51,

2020.