Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-51-AC2, 2020

Geoscientific Model Development Discussions

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

Charles Onyutha

conyutha@kyu.ac.ug

Received and published: 27 September 2020

GENERAL

The author is thankful to the anonymous reviewer for acknowledging that the contribution of this paper falls in the above average category. The author is further grateful to the reviewer for recognizing that some part of the paper is interesting and well written.

The reviewer remarked that paper is too long for what it is, and does not make fair comparisons with other metrics. To address this comment the reviewer offered a number of suggestions.

COMMENT 1

Abstract: the abstract is ineffective as it does not compactly present the major findings. Much of this material reads like an "introduction", starting with the subjective first sentence. My suggested rewrite of this sentence is:

A new measure of "goodness of fit" eliminates several of the well known shortcomings of the widely used correlation coefficient R2, including its insensitivity to bias when models are compared to measurements.

The abstract should then go on to describe the new metric, list its specific properties, then list its similarities, differences, advantages and disadvantages as compared to other metrics. Each sentence should be compact and deliver new and interesting information. No fluff or opinion is appropriate.

REPLY TO COMMENT 1

The author agrees with the reviewer on the need to revise the abstract. In this line, the abstract will be revised as below.

A new measure of "goodness of fit" hereinafter referred to as coefficient of model accuracy (CMA) eliminates several of the well known shortcomings of the widely used coefficient of determination R2, including its insensitivity to bias when model outputs are compared to measurements. CMA can be computed as the product of correlation coefficient, measure of deviations of data points from comparison baseline, and ratio of squared means of observed data (X) to the average of modeled data (Y) or vice versa. Correlation coefficient quantifies the measure of statistical linear relationship between X and Y. Division of squared means of datasets measures the amount by which Y is biased. Measure of deviation of data points from a stipulated comparison baseline gives an insight on the differences in variances of X and Y. This is done while ensuring that CMA yields different values in the two cases (i) when we regress Y on X and (ii) if X is regressed on Y. CMA values ranges from zero to one. Based on large number of simulations, other metrics such as Index of Agreement, and Taylor Skill Score were found to get closer the maximum value of 1 faster than CMA. Comparison of CMA and other existing metrics can be found given with respect to several properties such as sensitivity to possible outliers, error quantification, and ease of computation. To allow applications of CMA, MATLAB and R codes as well as an illustrative MS Excel file to compute the CMA were provided for readers.

COMMENT 2

Line 45 ff This PP is very interesting; though I am concerned that misinterpretations may be included. Fundamentally, the reason that the values of R or R2 do not depend on whether y is plotted against x, or vice versa, is evident in the definition provided by Eq. 1; specifically, there is no difference in how x and y are treated in that mathematical definition, so that x and y are interchangeable and mathematically symmetric. That is, if all of the x's were replaced by y's, and all the y's were replaced by x's, the equation would look the same. Thus, given any two column table of data, either column could be defined as x and the other as y, and the result returned by eq 1 would be the same.

REPLY TO COMMENT 2

The author agrees with the reviewer regarding the explanation on why values of R or R2 do not depend on whether y is plotted against x, or vice versa. The clarification made by the reviewer will be included in the revised manuscript. Any possible misinterpretations regarding the statement in line 45 and Eq. (1) will be removed from the revised manuscript.

COMMENT 3

Equation 4 and line 75ff. I have confirmed that Eq. 4 is correct, but the claim that the "deviations of X and Y from their means to obtain are assessed independently" is not. Specifically, the author has mistakenly concealed that dependence in his substitution for "m". The reader can refer back to Eq. 2 to see that m depends on the PRODUCT of the x and y deviations, and Eq. 4 depends on m; this conclusively refutes the author's statement.

REPLY TO COMMENT 3

The author agrees with the reviewer that m from Eq. (2) depends on the product of x and y deviations and Eq. (4) depends on m. In this case the statement that "….deviations of X and Y from their means to obtain gamma are assessed independently" can no longer hold water and will be removed from the revised manuscript. The point the author wanted to make was that R2 does not take into account model errors because it does not comprise squared differences between x and y as considered directly by other metrics Nash Sutcliffe Efficiency (NSE) and Root Mean Squared Error (RMSE).

COMMENT 4

I don't have time to plough through the author's derivation, but it is clear that computing this would be very difficult compared to many of the simple, single-formula metrics that are currently available to compare models and measurements.

REPLY TO COMMENT 4

The author agrees with the reviewer that the version of CMA in the discussion paper can be more computationally difficult than some existing metrics such as NSE and RMSE. Although the derivation of CMA appears long, the formula of CMA is as simple as other existing "goodness-of-fit" metrics. In summary, there are basically two formulae the author was presenting to address shortcomings of Rˆ2. The first formula is measure of model efficiency (MME) and can be given by

MME = f $\times$ (min[mx, my]/max[mx, my]) $\times$(min[sx, sy]/max[sx, sy]) ....(1)

where f = absolute value of the coefficient of correlation between X and Y,

mx = squared mean of the x's,

my = squared mean of the y's,

sx = standard deviation of the x's,

sy = standard deviation of the y's,

min = minimum of two or more values, and

max = maximum of two or more values.

Computation of the MME is as simple as other existing "goodness-of-fit" metrics. The values of MME range from 0 to 1. MME=0 means the model indicates that the model is not better than the comparison baseline (or mean of observed series). MME=1 indicates that there are no model errors. In Eq. (1), MME takes into consideration co-variation of X and Y, difference in variances of X and Y, and how biased the mean of Y is from the mean of Xans of X and Y. However, a typical shortcoming of R2 which MME does not address it that when MME yields the same value in the two cases including (i) regressing Y on X, and (ii) when X is regressed on Y.

Based on the need to derive a metric (hereinafter referred to as Coefficient of Model Accuracy or Agreement CMA) such that it offers an all-encompassing solution to the shortcomings of Rˆ2, we modify the term expressing ratio of standard deviations in terms of the sum of squared deviations of x's and h's from a common cb baseline herein taken as cb=3×(mx)ˆ0.5. In other words, CMA=0 if mx=my=0, otherwise,

CMA= f × (min[mx, my]/max[mx, my]) × w1/w2 ....(2)

where w1= sum of the minimum values of (x(i)-cb)ˆ2 and (y(i)-cb)ˆ2,

w2= sum of the maximum values of (x(i)-cb)ˆ2 and (y(i)-cb)ˆ2

Computation of CMA is as simple as other existing "goodness-of-fit" metrics. Given that the formula for correlation coefficient is already in-built in many of the computing software packages; it is very easy to automate MME and CMA. To allow applications of the new metrics, MATLAB and R codes as well as an illustrative MS Excel file to compute the MME and CMA will be provided for readers (in the form of supplementary materials to the revised manuscript).

COMMENT 5

I am unimpressed with the choice of the dataset used to illustrate application of the model. Data that are widely and readily accessible would be better, for example, data from a government website would be better. That original data set could be manipulated in various simple ways to compare the manipulation, now representing the "model", with the real data, and different metrics applied and compared.

REPLY TO COMMENT 5

Datasets used to compare the various "goodness-of-fit" metrics as presented in the discussion paper were selected over a catchment from a data scarce region. Indeed, hydro-meteorological datasets from such a region tend to generally have quality issues and possible attempts to deal with the data limitation problems by in-filling of missing data values can lower the accuracy of model predictions. In this line, the author agrees with the reviewer on the need to select and use another dataset which can easily be accessed by readers.

To take the reviewer's comment into consideration, hydro-meteorological datasets including daily catchment runoff, catchment-wide rainfall and evapotranspiration over Jardine River catchment in North Queensland, Australia were obtained from the website of "eWater toolkit" via https://toolkit.ewater.org.au/ (September 9, 2020). The streamflow data was for gauging station no. 927001 and the catchment area was 2500 kmˆ2. The datasets can be found in a folder named "Data" under Rainfall Runoff Library (RRL) which can be downloaded upon online registration.

Outputs of hydrological models used for comparison of the various "goodness-of-fit" metrics were generated using automatic calibration strategy in terms of the Generalized Likelihood Uncertainty Estimation framework GLUE (Beven and Binley 1992). In this calibration strategy, GLUE technique is a Bayesian approach in which several parameters' sets are randomized from the prior distribution to infer the output (posterior) distribution based on the simulations. In short, there was no issue subjectivity in obtaining model outputs for comparison of "goodness-of-fit" metrics.

Beven, K. J., and Binley, A. M.: The future role of distributed models: Model calibration and predictive uncertainty, Hydrol. Process., 6, 279–298, doi:10.1002/hyp.3360060305, 1992.

COMMENT 6

The extended narrative of the performance of different metrics is ineffective. A table whose first column lists the various properties, with additional columns to the right providing values for the particular metric that heads that column, would be far better and would allow direct and simple comparisons of the properties of each metric with all others. Is the metric constrained to range from 0 to 1? Are effects of bias easily excluded? How many different formulae are needed to compute the metric? Ease of computation: if not a trivial calculation, are automated programs readily and widely available, as they are for R2? Does the metric have real physical significance? Is it widely used? Etc.

REPLY TO COMMENT 6

The author is grateful to the reviewer for the constructive set of comments. Eventually, a table as shown in Fig.1 and Fig. 2 of this document will be provided in the revised manuscript. From Fig. 1 and Fig.2, it can be seen that CMA was compared with other metrics with respect to more than ten properties.

COMMENT 7

My conclusion is that the author has more work to do. A shorter, compact paper would be more effective.

REPLY TO COMMENT 7

Based on the reviewer's comments, a lot of work will be done during revision of the manuscript such as

(i) presenting a more considered CMA expression into a single formula,

(ii) tabulating comparison of CMA and other existing metrics,

(iii) eliminating extended narrative of the performance of various "goodness-of-fit" metrics,

(iv) revising the choice of datasets,

(v) revising the entire abstract,

(vi) removing possible misinterpretations regarding the shortcomings of R2.

To ensure the manuscript is short and straightforward, if comparison of performance of various metrics is to be done, it will be provided as supplementary material to the revised manuscript.

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

REPLY TO COMMENT No.6

SNo	Property	“Goodness-of-fit” metric							
		R^2	NSE	CMA	MME	IOA	TSS	RMSE	MAB
1	Range	0 – 1	-∞ – 1	0 – 1	0 – 1	0 – 1	0 – 1	RMSE ≥ 0	-∞ – +∞
2	Error quantification*	Relative error measure	Relative error measure	Relative error measure	Relative error measure	Absolute error measure	Relative error measure	Absolute error measure	Relative error measure
3	Computation difficulty	Low	Low	Moderate	Low	Low	Moderate	Low	Low
4	Number of extra columns in Ms Excel (apart from the two columns of where X and Y are put) that can be used to compute the metric.	No column is required if an in-built formula is used, otherwise, 3 columns required.	2	2	None	2	1	1	1
5	Are automated programs available for its computation?	Yes	Can easily be automated	MATLAB, & RStudio codes available along with this paper	Can easily be automated	Can easily be automated	Can easily be automated	Yes	Yes
6	Does the metric measure co-variation of observed and modeled series.	Yes	No	Yes	Yes	No	Yes	No	No
7	How many sub-formulae required to compute the metric	1	1	2	2	1	2	1	1

* For an absolute error measure, the difference between observed and modeled data is obtained in terms of the unit of the observed variable. In relative error measure, the mismatch between observed and modeled data is quantitatively evaluated from zero to one. Here, values of zero and one indicate no relationship and perfect agreement, respectively (Legates and Davis, 1997).

Fig. 1. Comparison of CMA and other metrics (part 1)

SNo	Property	“Goodness-of-fit” metric							
		R^2	NSE	CMA	MME	IOA	TSS	RMSE	MAB
8	Application or use of the metric	Widely used	Widely used	New	New	Widely used	Widely used	Widely used	Widely used
9	Sensitivity of squaring differences between observed and model output on large values	No squared differences between observed and modeled values	High	No squared differences between observed and modeled values	No squared differences	High	No squared differences between observed and modeled values	High	No squared differences between observed and modeled values
10	Increase towards its maximum	Fast	Moderate	Moderate	Moderate	Fast	Fast	Moderate	Moderate
11	Physical relevance of the metric. Does it have unit?	No unit	No unit	No unit	No unit	No unit	No unit	Has unit of the observed variable	No unit
12	Metric’s value in regressing X on Y compared to the case when Y is regressed on X	Same	Different	Different	Same	Same	Same	Same	Different
13	Sensitivity of the metric to possible outliers	High	High	Low	Low	High	High	High	High

Reference

Legates, D. R. and Davis, R. E.: The continuing search for an anthropogenic climate change signal: limitations of correlation-based approaches, Geophys. Res. Lett., 24, 2319–2322, doi: 10.1029/97GL02207, 1997.

Fig. 2. Comparison of CMA and other metrics (part 2)