Ecological Modelling 411 (2019) 108802

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Letter to the Editor

Evaluating model results in scatter plots: A critique

Through a synthetic experiment, Piñeiro et al. (2008) raised the argument that observed data should be regressed against model results when evaluating a model's performance in a scatter plot. This paper demonstrates that the conclusions from their study are an artefact of the experiment setup, the results of which can be proven analytically. Through a simple example, it is shown that regressing observations against model results can artificially make the results of a model look better than regressing the results against the observations. The overall conclusion from this paper is that assessing models in a scatter plot with the observations in abscissa (X-axis) and the corresponding simulations in ordinate (Y-axis) will lead to the correct conclusions regarding the model performance.

The experiment in Piñeiro et al. (2008) can be summarized as follows. First, a vector X with discrete values 1, 2, 3, … to 60 was constructed. Then, a vector Y was constructed by adding a Gaussian random number ϵ with mean zero and standard deviation 15 to each of these numbers. Piñeiro et al. (2008) assumed X and Y to be observation vectors. A vector $\hat{Y}$ was then generated through a linear regression with X as abscissa and Y as ordinate. This vector was then referred to as the model results.

A linear regression with Y as abscissa and $\hat{Y}$ as ordinate resulted in slope and intercept values different from one and zero, respectively. A regression with $\hat{Y}$ as abscissa and Y as ordinate resulted in slope and intercept values of one and zero, respectively. Based on these results, Piñeiro et al. (2008) argued that in scatter plots between observations and model results, the model results should be placed in abscissa and the corresponding observations in ordinate.

Piñeiro et al. (2008) also performed this analysis with a quadratic and a logarithmic relationship between Y and X, but for reasons of clarity and brevity this commentary focuses only on the linear case.

Two points of criticism need to be made in relation to Piñeiro et al. (2008). The first is that the objective of a regression is to model the linear relationship between a dependent and independent variable. The approach of adding noise to a variable, performing a regression through the resulting data set, and then making a scatter plot between the data with noise and the results of the regression is very different from the way model results are compared to observations. The second is that, because Y equals X plus a Gaussianly distributed perturbation, when the number of entries in the observation vector is increased, $\hat{Y}$ simply becomes equal to X. In other words, if the sample size is large enough, the analysis is simply equal to regressing Y as a function of X (because $\hat{Y}$ becomes equal to X), and the other way around. $\hat{Y}$ should thus not be interpreted as model results corresponding to observations Y.

The following shows that the numerical results of the regressions in Piñeiro et al. (2008), based upon which they draw their conclusions, can be calculated analytically, demonstrating that their conclusions are

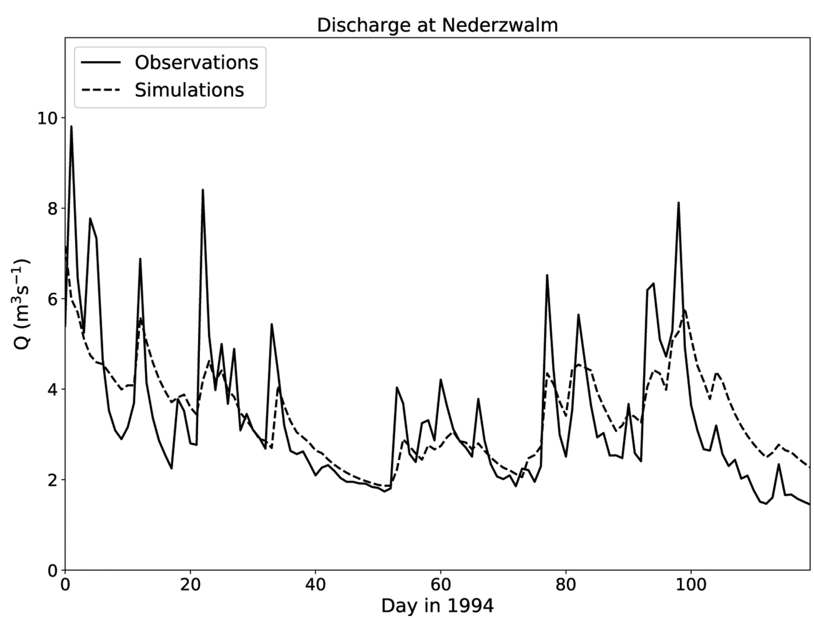

Fig. 1. Time-series of daily observed and simulated discharge at Nederzwalm.

https://doi.org/10.1016/j.ecolmodel.2019.108802
Received 27 June 2019

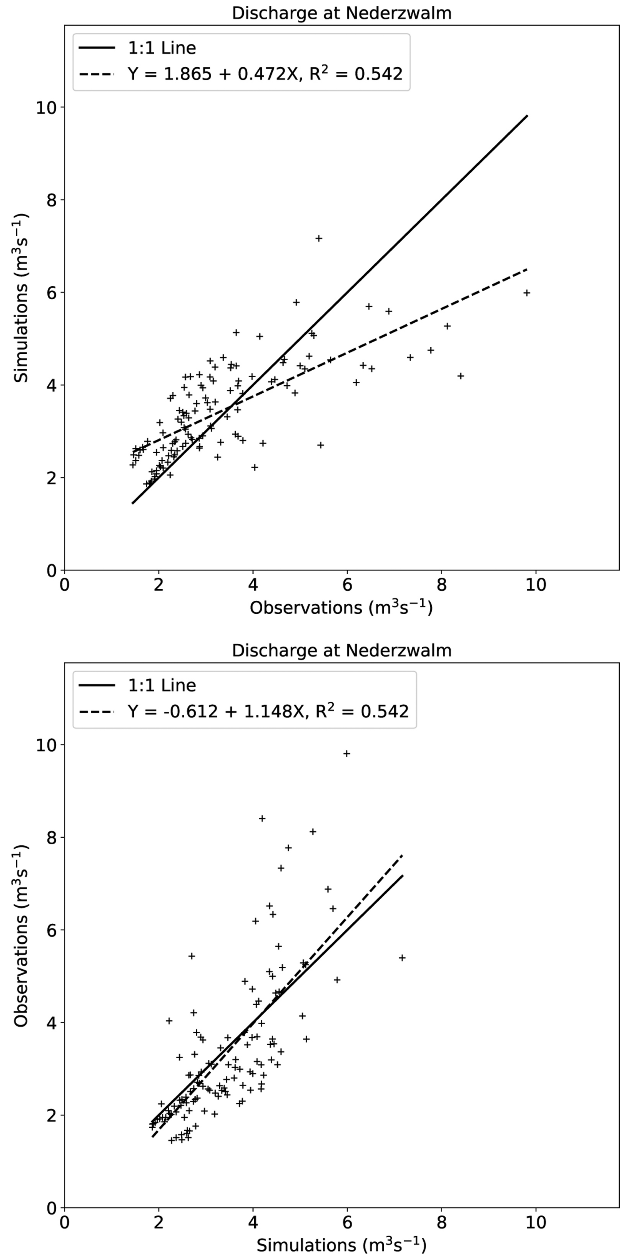

Fig. 2. Top panel: regression of the simulations to the observations. Bottom panel: regression of the observations to the simulations.

an artefact of the experiment setup. It should be noted that the principles of this analysis are well known (see for example Buonaccorsi, 2010). However, for reasons of clarity to the readers, the basics of the analysis are provided here.

The slope and intercept of a linear regression between a dependent variable Y (ordinate) and an independent variable X (abscissa) are calculated as:

$$\begin{cases} b_1 = \dfrac{S_{\mathrm{XY}}}{S_X^2} \\ b_0 = \bar{Y} - b_1\bar{X}, \end{cases} \tag{1}$$

where b_0 and b_1 are the regression slope and interval, respectively, S_{XY} and S_X^2 are the covariance between X and Y and the variance of X, respectively, and $\bar{X}$ and $\bar{Y}$ are the mean of X and Y, respectively.

As Y is written as $X + \epsilon$, it is straightforward to show that:

$$\begin{cases} S_Y^2 = S_X^2 + S_\epsilon^2 \\ S_{\mathrm{XY}} = S_X^2, \end{cases} \tag{2}$$

where S_Y^2 and S_ϵ^2 are the variance of Y and the Gaussian random number ϵ, respectively. Eq. (2) can therefore be applied to calculate the linear regression coefficients following Eq. (1) and the data from Piñeiro et al. (2008). If a regression is performed using X as abscissa and Y as ordinate, this leads to:

$$\begin{cases} b_1 = \dfrac{S_{\mathrm{XY}}}{S_X^2} = \dfrac{S_X^2}{S_X^2} = 1 \\ b_0 = \bar{Y} - b_1\bar{X} = 0. \end{cases} \tag{3}$$

However, if Y is the regression abscissa and X the ordinate, the result is:

$$\begin{cases} b_1 = \dfrac{S_{\mathrm{XY}}}{S_Y^2} = \dfrac{S_X^2}{S_X^2 + S_\epsilon^2} = \dfrac{1}{1 + \frac{S_\epsilon^2}{S_X^2}} \\ b_0 = (1 - b_1)\bar{X}. \end{cases} \tag{4}$$

Calculating the mean and standard deviation of the vector X can be interpreted as calculating these statistics for a uniform distribution

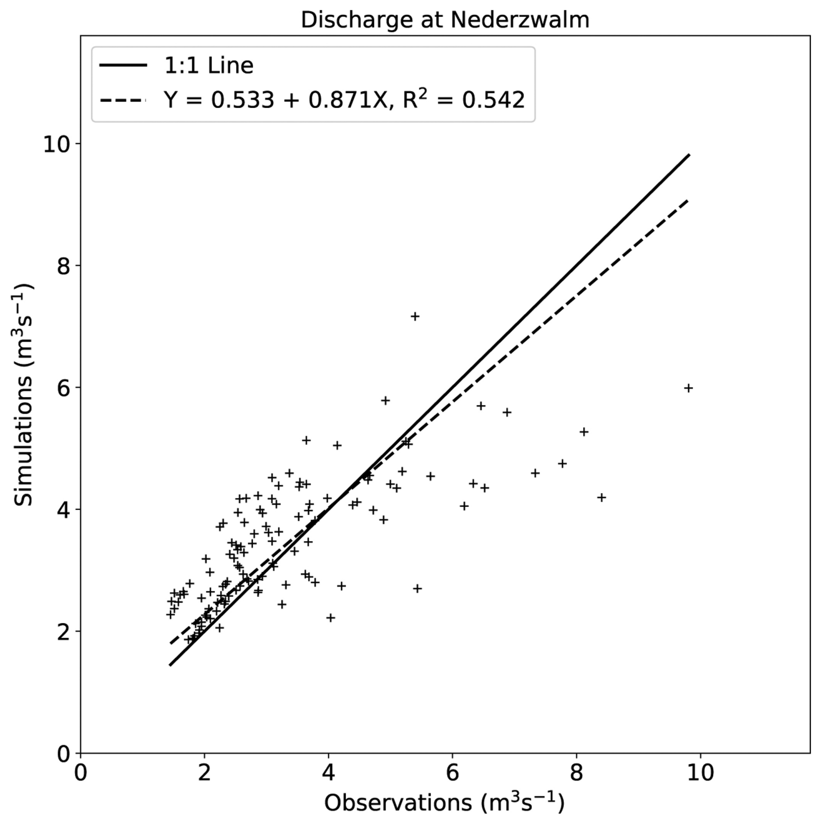

Fig. 3. Inverted regression of the observations to the simulations.

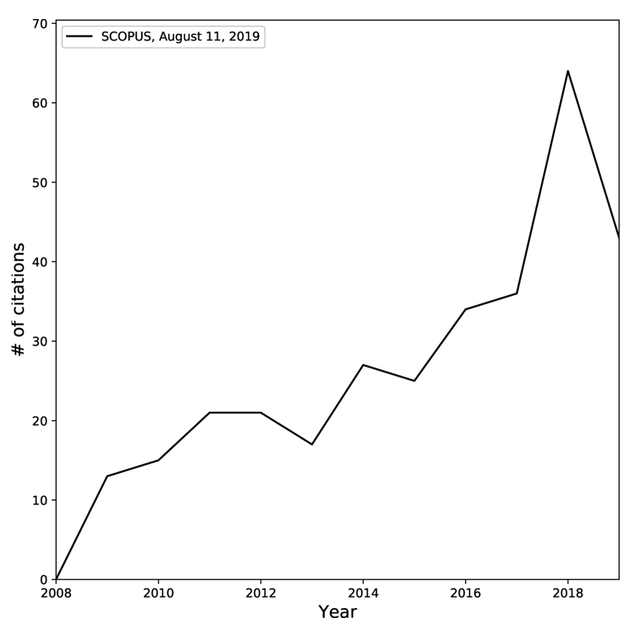

Fig. 4. Number of citations to Piñeiro et al. (2008) as of August 11, 2019.

Table 1
Journals with 3 or more papers referring to Piñeiro et al. (2008) ranked according to the number of papers. SNIP is the Source Normalized Impact per Paper. *Scientific Reports* has currently not been assigned a category and has not yet a SNIP.

	Journal	Number of papers	Cumulative number of papers	Rank in category	Number of journals in category	SNIP
1	Remote Sensing	11	11	7	182	1.768
2	Ecological Modelling	10	21	8	30	1.115
3	Remote Sensing of Environment	9	30	1	116	2.961
4	Agricultural and Forest Meteorology	8	38	2	139	1.738
5	Forest Ecology and Management	7	45	11	139	1.478
6	PLoS ONE	6	51	20	185	1.123
7	Field Crops Research	6	57	12	320	1.871
8	Water Resources Research	5	62	8	203	1.676
9	Environmental Modelling and Software	4	66	10	117	1.999
10	Ecological Applications	4	70	28	336	1.557
11	Journal of Hydrology	4	74	7	203	1.917
12	Estuarine, Coastal and Shelf Science	4	78	25	204	1.164
13	New Forests	3	81	17	139	1.281
14	Canadian Journal of Fisheries and Aquatic Sciences	3	84	36	204	1.205
15	New Zealand Journal of Marine and Freshwater Research	3	87	153	336	0.637
16	Journal of Applied Ecology	3	90	12	336	1.939
17	Freshwater Biology	3	93	11	204	1.333
18	Scientific Reports	3	96			
19	Ecological Engineering	3	99	13	140	1.451

Table 2
Top 20 journals referring to Piñeiro et al. (2008) ranked according to their Source Normalized Impact per Paper (SNIP) value.

	Journal	Number of papers	Cumulative number of papers	Rank in category	Number of journals in category	SNIP
1	Nature	1	1	1	90	9.199
2	Remote Sensing of Environment	9	10	1	116	2.961
3	ISPRS Journal of Photogrammetry and Remote Sensing	1	11	14	569	2.862
4	Nature Communications	1	12	5	189	2.805
5	IEEE Transactions on Geoscience and Remote Sensing	1	13	5	182	2.743
6	Global Change Biology	2	15	2	75	2.614
7	Methods in Ecology and Evolution	1	16	12	591	2.587
8	Landscape and Urban Planning	1	17	2	140	2.470
9	Water Research	1	18	1	203	2.426
10	International Journal of Applied Earth Observation and Geoinformation	1	19	1	136	2.209
11	Soil and Tillage Research	2	21	2	320	2.178
12	Energy Conversion and Management	1	22	2	61	2.151
13	Environmental Modelling and Software	4	26	10	117	1.999
14	Photogrammetric Engineering and Remote Sensing	1	27	10	34	1.994
15	Remote Sensing in Ecology and Conservation	1	28	6	140	1.944
16	Journal of Applied Ecology	3	31	12	336	1.939
17	Agricultural Water Management	2	33	11	203	1.933
18	Journal of Ecology	1	34	12	404	1.924
19	Journal of Hydrology	4	38	7	203	1.917
20	Soil Biology and Biochemistry	2	40	2	116	1.879

between 1 and 60 (see the experiment setup explained above), where the mean and standard deviation are:

$$\begin{cases} \bar{X} = \dfrac{60+1}{2} = 30.5 \\ S_X = \sqrt{\dfrac{(60-1)^2}{12}} = \sqrt{290.08} = 17.03 \end{cases} \quad (5)$$

Thus the regression coefficients will become (Eq. (4)):

$$\begin{cases} b_1 = \dfrac{1}{1+\frac{15^2}{290.08}} = 0.56 \\ b_0 = (1-0.56)\cdot 30.5 = 13.32 \end{cases} \quad (6)$$

Eq. (6) explains the slope values between 0.4 and 0.8, and the intercepts between 9 and 19 in Figure 2 of Piñeiro et al. (2008). The large spread in these values can be explained by the limited sample size of 60. Increasing this sample size leads to narrower distributions of the slope and intercept values around their theoretical values.

The implications of this analysis are illustrated through a numerical example with a relatively simple rainfall-runoff model. Fig. 1 shows the results of the application of a poorly calibrated Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Linström et al., 1997) to the Zwalm river in Belgium, for the first four months of 1994. The model clearly underestimates the peak flows and tends to overestimate low flows. Fig. 2 shows the comparison of the model simulations to the observations, and vice versa. The regression line calulated with the observations in abscissa is clearly below the 1:1 line for high observed values, and above the 1:1 line for low observations, which is consistent

Table 3
Categories into which the papers referring to Piñeiro et al. (2008) are classified, ranked according to the number of papers. 14 papers are not classified.

	Category	Number of papers	Cumulative number of papers
1	Forestry	33	33
2	Aquatic Science	24	57
3	Agronomy and Crop Science	20	77
4	Ecology, Evolution, Behavior and Systematics	19	96
5	General Earth and Planetary Sciences	19	115
6	Ecology	19	134
7	Water Science and Technology	14	148
8	Soil Science	14	162
9	General Agricultural and Biological Sciences	12	174
10	Animal Science and Zoology	12	186
11	Ecological Modelling	10	196
12	Nature and Landscape Conservation	9	205
13	Earth-Surface Processes	9	214
14	General Environmental Science	7	221
15	Insect Science	6	227
16	Horticulture	6	233
17	Plant Science	6	239
18	Multidisciplinary	5	244
19	Environmental Engineering	5	249
20	Geography, Planning and Development	3	252
21	General Veterinary	3	255
22	Safety, Risk, Reliability and Quality	3	258
23	Food Science	3	261
24	Condensed Matter Physics	2	263
25	Food Animals	2	265
26	Global and Planetary Change	2	267
27	Electrical and Electronic Engineering	2	269
28	Management, Monitoring, Policy and Law	2	271
29	Applied Mathematics	2	273
30	Nuclear Energy and Engineering	1	274
31	Health Policy	1	275
32	Atmospheric Science	1	276
33	Microbiology (medical)	1	277
34	Architecture	1	278
35	Computer Science Applications	1	279
36	Geology	1	280
37	Psychiatry and Mental Health	1	281
38	General Computer Science	1	282
39	General Physics and Astronomy	1	283
40	General Chemistry	1	284
41	Civil and Structural Engineering	1	285
42	Geochemistry and Petrology	1	286
43	History	1	287
44	Oceanography	1	288
45	Education	1	289
46	General Biochemistry, Genetics and Molecular Biology	1	290
47	Palaeontology	1	291
48	Industrial and Manufacturing Engineering	1	292
49	Engineering (miscellaneous)	1	293
50	Epidemiology	1	294
51	Agricultural and Biological Sciences (miscellaneous)	1	295
52	Clinical Biochemistry	1	296
53	Development	1	297
54	Geophysics	1	298
55	Organizational Behavior and Human Resource Management	1	299
56	Transplantation	1	300
57	Computers in Earth Sciences	1	301
58	Modelling and Simulation	1	302

with Fig. 1. However, the regression line calculated with the simulations in abscissa is located very close to the 1:1 line. This would lead to the conclusion that there is no consistent underestimation of high flows and underestimation of low flows, which is different from the results in Fig. 1.

Fig. 3 shows the inverted regression line obtained with the simulations in abscissa. Consistent with Fig. 2, the inverted regression line is located closer to the 1:1 line.

This artificially improved regression when comparing observations to simulations can be explained by examining the linear regression equations. More specifically, similar as in Eq. (1), the regression between Y (as abscissa) and X (as ordinate) is written as:

$$X = c_0 + \frac{S_{XY}}{S_Y^2} Y. \tag{7}$$

This relationship is inverted as:

$$Y = -\frac{S_Y^2}{S_{XY}} c_0 + \frac{S_Y^2}{S_{XY}} X. \tag{8}$$

The slope of this regression is divided by the original value (Eq. (1)) and this ratio is called R_S:

$$R_S = \frac{\frac{S_Y^2}{S_{XY}}}{\frac{S_{XY}}{S_X^2}} = \frac{S_X^2 S_Y^2}{S_{XY}^2} = \frac{1}{R^2}, \tag{9}$$

where R^2 is the coefficient of determination between X and Y. As R^2 ranges between 0 and 1, the implication of Eq. (9) is that the absolute value of the inverted slope obtained from a regression with the observations in ordinate will *always* be larger than the slope obtained from a regression with the observations in abscissa. If the latter slope is between 0 and 1, which suggests a model underestimating high values and overestimating low values, a regression with the observations in ordinate can lead to an artificially good match between the model results and the observations. A linear regression with the observations in abscissa will not be prone to this drawback.

Piñeiro et al. (2008) is receiving strong and increasing attention in the scientific literature. This is clearly demonstrated in Fig. 4, which shows the number of citations per year according to SCOPUS (August 11, 2019). Table 1 lists the journals with three or more citing papers, ranked according to the number of papers referring to Piñeiro et al. (2008). This indicates that papers using these recommendations have been published in highly ranked journals. Table 2 lists the top 20 of these journals ranked by their Source Normalized Impact per Paper (SNIP), which is an indication of the amount of citations the journal receives compared to its entire field. Of the 316 papers referring to Piñeiro et al. (2008), 54 and 117 are ranked in the top three and top ten of their categories, respectively. This further demonstrates that a number of very highly ranked journals have published papers that follow the recommendations of Piñeiro et al. (2008). Table 3 lists the categories into which the referring papers are classified, ranked according to the number of papers citing Piñeiro et al. (2008). In general the vast majority of these categories are in the field of Earth sciences, with this term being interpreted very broadly. Only two papers have been published in a category related to mathematics or statistics (the category *Applied Mathematics*). These papers simply applied the recommendations from Piñeiro et al. (2008) without scrutiny. A critical analysis of Piñeiro et al. (2008) has not yet been published.

In summary, an erroneous reasoning of Piñeiro et al. (2008) has led to the conclusion that a scatter plot between model results and observations needs to be made with the model results in the X-axis. The analysis here has demonstrated that presenting results in this manner can lead to wrong conclusions regarding the model performance.

Rather, regressing model simulations against the corresponding observations will lead to the correct conclusions regarding the quality of the model results. However, when assessing model results, a scatter plot will usually not be sufficient. This will need to be accompanied by an analysis of statistical indicators such as the bias, Root Mean Square Error, correlation coefficient, or other measures.

Acknowledgements

Adrien Guyot was funded through the Australian Research Council grant DP160101377. We thank the Flemish Environmental Agency for the discharge data for the Zwalm catchment.

References

Buonaccorsi, J.P., 2010. Measurement Error: Models, Methods, and Applications. Chapman and Hall/CRC, Boca Raton, FL.

Linström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201 (1–4), 272–288.

Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? Ecol. Model. 216 (3–4), 316–322. https://doi.org/10.1016/j.ecolmodel.2008.05.006.

Valentijn R.N. Pauwels*, Adrien Guyot, Jeffrey P. Walker
Monash University, Department of Civil Engineering, Clayton, Victoria, Australia
E-mail address: Valentijn.Pauwels@monash.edu (V.R.N. Pauwels).

* Corresponding author.