2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
DOI: https://doi.org/10.13031/aim.201901248

Paper Number: 1901248

Relationships Between Coefficients of Determination and Correlation, Bias, and Regression

Hubert Montas[1], Adel Shirmohammadi[2], Masoud Negahban-Azar[2] and Paul Leisnham[2]

[1]Fischell Department of Bioengineering, University of Maryland at College Park, College Park, MD, 20742

[2]Environmental Science and Technology Department, University of Maryland at College Park, College Park, MD, 20742

Written for presentation at the
2019 ASABE Annual International Meeting
Sponsored by ASABE
Boston, Massachusetts
July 7–10, 2019

ABSTRACT. ***Mathematical relationships between the coefficient of determination, the Nash-Sutcliffe coefficient of efficiency and the coefficient of linear correlation, are derived and presented for the general case, applicable to models calibrated by nonlinear regression processes. These coefficients are the most used for the calibration and validation of hydrologic models and are frequently assumed independent measures of goodness-of-fit. This contribution clarifies how each relates to the others, and how linear regression can be used to define transformed hydrologic models whose predictions are characterized by coefficients of determination that equal the square of their correlation coefficients (a situation which is not guaranteed to arise from nonlinear parameter-identification processes). The mathematical relationships presented in this work are expected to help model developers better identify the type of information provided by each test statistic during model calibration and validation, and to help guide both model development and parameter adjustment efforts to counteract the specific types of errors highlighted by each statistic. A synthetic example is presented and discussed to illustrate the derived relationships and related model adjustment approach.***

Keywords. Hydrology, Modeling, Model Calibration, Model Validation, Diagnostic Statistics

Introduction

The coefficient of determination, R^2, and the coefficient of linear correlation, r, are two of the most used indicator statistics for identifying the degree to which a set of model predictions matches a set of observations. Unfortunately, in the hydrologic literature at least, there is currently a significant degree of confusion as to the relationship between these two indicators, and how they relate to the Nash-Sutcliffe coefficient of Efficiency, NSE. The relationships between these three coefficients should be well known and clearly described in standard textbooks, article and model user's manuals, but this is not the case. The objective of this manuscript is to rectify this situation.

We start by deriving explicit formulas for R^2 and r in terms of statistics of predicted and observed values and demonstrate that R^2 is not generally equal to r^2 (in the general case, the coefficient of determination is not the square of the coefficient of linear correlation). We derive the explicit relation between R^2 and r^2 and, discuss its form and proceed to identify the special, restricted, case in which the two are equal. We also use the literature to show that the Nash-Sutcliffe coefficient of Efficiency is exactly the same as the coefficient of determination (NSE is exactly the same as R^2).

We continue with an analysis of linear least-square-error regression and demonstrate that if the parameters of a predictive model are identified using this specific process, then the restricted conditions for the special case where R^2 equals r^2 do apply. However, linear least-square-error regression is generally not applicable to parameter estimation in highly nonlinear hydrologic models and therefore one may expect, as observed in practice, that R^2 and r^2 can differ substantially in this type of work.

We show that by performing linear least-square-error regression of the predictions of a calibrated model, onto observations, we can obtain a linearly-modified model for which the coefficient of determination (and NSE) equals the square of the coefficient of correlation. This type of adjustment may be considered in situations where the original model of a phenomenon is believed to be missing the description of a process that otherwise affects observations. This may be diagnosed by values of R^2 for the nonlinearly calibrated model that are substantially lower than r^2. Clearly, the application of such linear adjustment of model predictions has to be justified based on characteristics of the observed phenomenon and specific model used to predict it.

Finally, we provide a synthetic example of the formulas and considerations developed in the manuscript.

Coefficients of Determination and Correlation

We consider a vector O of n observations of a phenomenon of interest (for example daily streamflow in a hydrologic context) and a vector P of n corresponding predictions performed by an arbitrary model of the phenomenon:

$$P=\begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix} \quad , \quad O=\begin{bmatrix} O_1 \\ O_2 \\ \vdots \\ O_n \end{bmatrix}$$

The above vectors are written in column form (as though they are one column of a matrix) and can be transposed to row form (one row of a matrix) by using the standard matrix transpose operator which will be depicted as a superscript T in this manuscript. This will make it possible to use standard linear algebra techniques (Strang, 1986) to simplify some aspects of the presentation (see also Neter et al., 1990, where the forward quote is used to indicate transposition).

The differences between model predictions and observations form a vector E of n error values, computed as:

$$E = P - O = \begin{bmatrix} P_1 - O_1 \\ P_2 - O_2 \\ \vdots \\ P_n - O_n \end{bmatrix}$$

The statistics of P, O and E are important values for the analysis of how well model predictions might fit observations. In this work, the mean of a variable is indicated by an overbar, and the means of predictions, observations and errors are calculated as:

$$\overline{P} = \frac{1}{n}\sum_{i=1}^{n} P_i \quad , \quad \overline{O} = \frac{1}{n}\sum_{i=1}^{n} O_i \quad , \quad \overline{E} = \frac{1}{n}\sum_{i=1}^{n} E_i = \overline{P} - \overline{O}$$

Using the means enables the definition of mean-removed variables which help to make several upcoming formulas more succint. Mean-removed variables are indicated using lower-case letters here and each consists of a vector of n values. For model predictions, observations and errors, they are defined as:

$$p = \begin{bmatrix} P_1 - \overline{P} \\ P_2 - \overline{P} \\ \vdots \\ P_n - \overline{P} \end{bmatrix} \; , \; o = \begin{bmatrix} O_1 - \overline{O} \\ O_2 - \overline{O} \\ \vdots \\ O_n - \overline{O} \end{bmatrix} \; , \; e = \begin{bmatrix} E_1 - \overline{E} \\ E_2 - \overline{E} \\ \vdots \\ E_n - \overline{E} \end{bmatrix} = \begin{bmatrix} (P_1 - O_1) - (\overline{P} - \overline{O}) \\ (P_2 - O_2) - (\overline{P} - \overline{O}) \\ \vdots \\ (P_n - O_n) - (\overline{P} - \overline{O}) \end{bmatrix} = \begin{bmatrix} (P_1 - \overline{P}) - (O_1 - \overline{O}) \\ (P_2 - \overline{P}) - (O_2 - \overline{O}) \\ \vdots \\ (P_n - \overline{P}) - (O_n - \overline{O}) \end{bmatrix} = p - o$$

The variances of predictions and observations, as well as their covariance, form a second group of important statistics. They are defined below on a sample basis (normalizing by n-1) and expressed equivalently as sums of vector values and as linear-algebraic products of an appropriate row vector by a column vector (which is more concise):

$$\sigma_p^2 = \frac{1}{n-1}\sum_{i=1}^{n} (P_i - \overline{P})^2 = \frac{1}{n-1}\sum_{i=1}^{n} p_i^2 = \frac{p^T p}{n-1}$$

$$\sigma_o^2 = \frac{1}{n-1}\sum_{i=1}^{n} (O_i - \overline{O})^2 = \frac{1}{n-1}\sum_{i=1}^{n} o_i^2 = \frac{o^T o}{n-1}$$

$$\sigma_{po} = \frac{1}{n-1}\sum_{i=1}^{n} (P_i - \overline{P})(O_i - \overline{O}) = \frac{1}{n-1}\sum_{i=1}^{n} p_i o_i = \frac{p^T o}{n-1}$$

The third group of statistics are those related to error variance. The specific statistics used for computing common measures of model fit are the variance of the error and the expected value of the squared error (normalized by n-1), respectively:

$$\sigma_e^2 = \frac{1}{n-1}\sum_{i=1}^{n} (E_i - \overline{E})^2 = \frac{1}{n-1}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-1}\sum_{i=1}^{n} (p_i - o_i)^2 = \sigma_p^2 - 2\sigma_{po} + \sigma_o^2$$

$$S_e^2 = \frac{1}{n-1}\sum_{i=1}^{n} E_i^2 = \frac{1}{n-1}\sum_{i=1}^{n} (e_i + \overline{E})^2 = \sigma_e^2 + \frac{n}{n-1}\overline{E}^2$$

It is notable that these error statistics are not independent values. As shown above, they are directly computable from the means, variances and covariance of predictions and observations.

The coefficient of determination, R^2, is one of the most frequently used goodness-of-fit indicator statistic for evaluating how well a set of model predictions match a set of observations. It is defined by equation 3.71 in the standard statistics textbook of Neter et al. (1990). Unfortunately, these authors confusingly denote the coefficient of determination using the lower-case r^2 as they reserve the upper-case R^2 (equation 7.35) for the coefficient of multiple determination that applies to multilinear models. Adding to the confusion, they state in their equation 3.73 that the coefficient of correlation, r, is the square root of the coefficient of determination (their r^2, our R^2) without stating the specific case that they are considering (but a major hint is in the title of their book: Applied Linear Statistical Models). We will see shortly that the relationship between coefficient of determination and correlation is not so simple, except in a special case. In their notation (equation 3.71) and ours, the coefficient of determination is given by:

$$R^2 = 1 - \frac{SSE}{SSTO} = 1 - \frac{S_e^2}{\sigma_o^2} \quad , \quad \text{where:} \quad SSE = \sum_{i=1}^{n} E_i^2 \; , \; SSTO = \sum_{i=1}^{n} (O_i - \overline{O})^2$$

The coefficient of determination indicates the amount of the variations in observed values that is explained by model predictions. It has a value of 1 for a perfect model that predicts all observed variations of a phenomenon, a value of 0 for a model that has no predictive value (one could just use the mean of observations as predictor, with the same result) and a

negative value when model predictions are less informative than the mean observation (the model just adds noise as its predictions deviate further from observations than observations deviate from their mean).

The Nash-Sutcliffe coefficient of Efficiency, NSE, is also frequently used to evaluate model predictions against observations. However, it is not distinct from the coefficient of determination but, rather, it is exactly the same. Nash and Sutcliffe themselves state that their coefficient is "analogous to the coefficient of determination" (Nash and Sutcliffe, 1970, p.288). The only difference is that it was derived within a hydrologic context, but the development resulted in the same formula, with the same interpretation. The formula for the NSE is found in equation 3 of Nash and Sutcliffe (1970), equation 11.30 of James and Burges (1982) or the third equation in Table 5 of Moriasi et al. (2015). Nash and Sutcliffe use the variable q for observations (our O), q' for predictions (our P), and use an overbar to indicate an average. Accordingly, their coefficient, which we denote NSE (they denote it by R^2, but we keep this letter for the coefficient of determination) is (using their equations 1-3):

$$NSE = \frac{F_o^2 - F^2}{F_o^2} = 1 - \frac{F^2}{F_o^2} = 1 - \frac{\sum_{i=1}^{n} \left(q_i' - q_i\right)^2}{\sum_{i=1}^{n} \left(q_i - \overline{q}\right)^2} = 1 - \frac{\sum_{i=1}^{n} \left(P_i - O_i\right)^2}{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2} = 1 - \frac{SSE}{SSTO} = R^2$$

In other words, the NSE is exactly the same as the coefficient of determination:

$$NSE = R^2 \qquad : always$$

The coefficient of linear correlation, r, is a second diagnostic statistic that is frequently used to evaluate how model predictions fit data. The formula for r is found in equation 3.75 of Neter et al. (1990), in equation 11.13 of James and Burges (1982) with S_R and S_S normalized by n, and in the first equation in Table 5 of Moriasi et al. (2015). The coefficient is calculated as:

$$r = \frac{\sum_{i=1}^{n} \left(P_i - \overline{P}\right)\left(O_i - \overline{O}\right)}{\sqrt{\left[\sum_{i=1}^{n} \left(P_i - \overline{P}\right)^2\right]\left[\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2\right]}} = \frac{\sigma_{po}}{\sqrt{\sigma_p^2\, \sigma_o^2}}$$

its value is positive in the case where model predictions go up when observations go up and model predictions go down when observations go down (predictions and observations are positively co-variated) as one would hope, and its value is negative if model predictions go up when observations go down, and predictions go down when observations go up (the model is inverted relative to the data). When model predictions are a perfect fit to the data (in a mean-removed sense since the coefficient explicitly excludes mean values), one can freely exchange O and P in variance and covariance calculations, the covariance of O and P becomes equal to both variances such that the coefficient takes on a value of 1. Similarly, if the model provides a perfectly inverted prediction of the data, we can exchange P and -O (minus O) such that the covariance equals minus both variances and the coefficient has a value of -1.

The coefficient of determination and the coefficient of linear correlation can be expressed in terms of the same statistics of predictions, observations and errors to directly see that, in the general case, one is not the square of the other:

$$R^2 = NSE = 1 - \frac{S_e^2}{\sigma_o^2} = 1 - \frac{\sigma_e^2}{\sigma_o^2} - \frac{n}{n-1}\frac{\overline{E}^2}{\sigma_o^2}$$

$$r = \frac{\sigma_{po}}{\sqrt{\sigma_p^2\, \sigma_o^2}} = \frac{1}{2}\frac{\sigma_p^2 + \sigma_o^2 - \sigma_e^2}{\sqrt{\sigma_p^2\, \sigma_o^2}}$$

The general relationship between these two coefficients may however be obtained by solving the coefficient of linear correlation equation for error variance (expressing it as a function of r) and substituting the resulting expression into the equation for the coefficient of determination. This yields:

$$R^2 = NSE = 2\sqrt{\frac{\sigma_p^2}{\sigma_o^2}}\, r - \frac{\sigma_p^2}{\sigma_o^2} - \frac{n}{n-1}\frac{\overline{E}^2}{\sigma_o^2}$$
$$= \frac{\sigma_p^2}{\sigma_{po}}\left[2 - \frac{\sigma_p^2}{\sigma_{po}}\right] r^2 - \frac{n}{n-1}\frac{\overline{E}^2}{\sigma_o^2}$$

Before discussing this general relationship, we note that it may be expressed in terms of other frequently used statistics, namely model prediction bias, B, and coefficient of variation of the observations, C_o, as follows:

$$R^2 = NSE = \frac{\sigma_p^2}{\sigma_{po}}\left[2 - \frac{\sigma_p^2}{\sigma_{po}}\right] r^2 - \frac{n}{n-1}\frac{B^2}{C_o^2} \quad , \quad \text{with:} \quad B = \frac{\overline{E}}{\overline{O}} \quad , \quad C_o = \frac{\sqrt{\sigma_o^2}}{\overline{O}}$$

The general relationship between the coefficient of determination and the coefficient of linear correlation clearly demonstrates that, in the general case, the former is not simply the square of the latter. In particular, one notes that the mean error between predictions and observations (or the bias) is always a negative contributor to the coefficient of determination. Such error always pulls R^2 down and one may expect that in many nonlinear calibration cases, low values of the coefficient of determination (especially relative to r^2) are caused by this mean error (nonlinear optimization does not guarantee equality of predicted and observed means). Nonlinear optimization is necessary, for example, to calibrating state-of-the-art hydrologic models, such as SWAT (Neitsch et al, 2011) and HSPF (Bicknell et al., 1996), which are highly nonlinear. Equally important, the factor multiplying r^2 in the above expression is a parabola of the form: $f(x) = x(2-x)$ which has negative curvature and a maximum value of 1 (obtained at $x = 1$, which means that the covariance of P and O equals the variance of P). The nonlinear optimization methods frequently used to identify model parameters do not guarantee equality of the prediction-observation covariance with the variance of predictions and will therefore generally result in factors of r^2 that are lower than 1. Taking this fact together with the negative contribution of the mean error implies that R^2 is always either smaller than r^2, or at best equal to it, in the general case. In other words, the square of the coefficient of linear correlation is the upper limit of the coefficient of determination:

$$R^2 \le r^2 \qquad : always$$

The special, best-case, of equality is obtained when:

$$\sigma_{po} = \sigma_p^2 \quad \left(\equiv \quad \sigma_e^2 = \sigma_o^2 - \sigma_p^2\right) \qquad \text{and} \quad \overline{E} = 0 \quad \left(\equiv \quad \overline{P} = \overline{O}\right) \qquad : special \ \text{case}$$

for which one obtains the best-case result:

$$R^2 = NSE = r^2 = \frac{\sigma_p^2}{\sigma_o^2} = \frac{\sigma_{po}}{\sigma_o^2} \qquad : special \ \text{case}$$

In addition to the coefficients of determination and correlation, a regression line is frequently used to appraise the goodness of model predictions, visually. This line is obtained by performing a linear least-square-error regression process between observations, O, as independent variable, and model predictions, P, as reference (target) variable. The process is detailed in Appendix (albeit with X as independent variable and O as target) and produces:

$$L = \left[\overline{P} - \frac{\sigma_{po}}{\sigma_o^2}\,\overline{O}\right] + \frac{\sigma_{po}}{\sigma_o^2}\,O = \left[\overline{P} - r\sqrt{\frac{\sigma_p^2}{\sigma_o^2}}\,\overline{O}\right] + r\sqrt{\frac{\sigma_p^2}{\sigma_o^2}}\,O$$

Models that provide a better fit to data are those where the slope of this line is closest to 1 and its abscissa at the origin is closest to zero. For the special situation described above (which leads to $R^2 = r^2$) the regression line simplifies to:

$$L = \left(1 - r^2\right)\overline{O} + r^2\,O \qquad : special \ \text{case}$$

which clearly illustrates how model fit gets better as the correlation coefficient increases towards 1.

Special Case of Model Parameters Derived by Linear Regression

A predictive model commonly consists of one or more functions (eg. G) that take vectors of variables as input (eg. T, X, Y, ...) and produce a corresponding vector of predictions, P, as output. The model functions involve a set of parameters (eg. U_1, U_2, U_3,...) that the modeler adjusts, during model calibration, such that the model produces good predictions of a vector of reference observed values (eg. O). Such a model, and its output (vector of predictions) may be expressed as:

$$P = G\left(T, X, Y, Z, \ldots; U_{1,} U_{2,} U_{3,} \ldots\right)$$

where the vectors of input variables are:

$$T = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{bmatrix} \quad , \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \quad , \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad , \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}$$

A linear or nonlinear regression process, aimed at minimizing the deviations between model predictions and observations, is frequently used to obtain model parameter values. The linear regression process that minimizes the sum of squared prediction errors is detailed in the appendix and produces the following explicit results for the case where the predictive model is a line:

$$P = \left[\overline{O} - \frac{\sigma_{xo}}{\sigma_x^2}\,\overline{X}\right] + \frac{\sigma_{xo}}{\sigma_x^2}\,X$$

and for the case where the model is a plane (and input variables X and Y are uncorrelated):

$$P = \left[\overline{O} - \frac{\sigma_{xo}}{\sigma_x^2}\,\overline{X} - \frac{\sigma_{yo}}{\sigma_y^2}\,\overline{Y}\right] + \frac{\sigma_{xo}}{\sigma_x^2}\,X + \frac{\sigma_{yo}}{\sigma_y^2}\,Y$$

For this particular model-parameter identification approach, one readily verifies from the above equations (by taking averages on both sides, and simplifying) that the mean of predicted values equals the mean of observed values:

$$\overline{P} = \overline{O}$$

Additionally, computing the variance of model predictions (done here for the plane model as an example):

$$\sigma_p^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{\sigma_{xo}}{\sigma_x^2}\left(X_i - \overline{X}\right) + \frac{\sigma_{yo}}{\sigma_y^2}\left(Y_i - \overline{Y}\right)\right]^2$$
$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\frac{\sigma_{xo}}{\sigma_x^2}\right)^2 x_i^2 + 2\frac{\sigma_{xo}\sigma_{yo}}{\sigma_x^2\sigma_y^2}\,x_i y_i + \left(\frac{\sigma_{yo}}{\sigma_y^2}\right)^2 y_i^2\right] = \frac{\sigma_{xo}^2}{\sigma_x^2} + \frac{\sigma_{yo}^2}{\sigma_y^2}$$

and the covariance of model predictions and observations:

$$\sigma_{po} = \frac{1}{n-1}\sum_{i=1}^{n}\left(P_i - \overline{P}\right)\left(O_i - \overline{O}\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{\sigma_{xo}}{\sigma_x^2}\left(X_i - \overline{X}\right) + \frac{\sigma_{yo}}{\sigma_y^2}\left(Y_i - \overline{Y}\right)\right]\left(O_i - \overline{O}\right)$$
$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{\sigma_{xo}}{\sigma_x^2}\,x_i + \frac{\sigma_{yo}}{\sigma_y^2}\,y_i\right]o_i \;= \frac{\sigma_{xo}^2}{\sigma_x^2} + \frac{\sigma_{yo}^2}{\sigma_y^2}$$

demonstrates that they are equal in this special case of linear least-square-error regression, and therefore, the

relationship:

$$R^2 = NSE = r^2 = \frac{\sigma_p^2}{\sigma_o^2} \quad : least \text{ square error linear regression}$$

applies in this particular case. However, in the general case of nonlinear regression, or of linear regression that is not aimed at minimizing the sum of squared errors, or in cases where input variables are covariated, or when parameters are optimized using a genetic algorithm or a neural network, the above best-case relationship is not likely to apply.

In light of the above result, some of the confusion between the coefficients of determination and linear correlation is illuminated. When Neter et al. (1990) state in their equation 3.73 that the coefficient of correlation, r, is the square root of the coefficient of determination, for example, they mean that this is so for the specific case of linear least-square-error regression, which is the topic of their textbook, aptly named: "Applied Linear Statistical Models". It is a special case to be sure, but within that limited context, their statement is correct.

Linear Model Adjustment and Linear Regression

The reasons for which nonlinear optimization processes may not provide parameter estimates that eliminate mean error and meet the equal variance-covariance conditions include the possibility of being trapped in a local minimum, within a limited search space, and the possibility that the model does not include the description of a particular process that affects the observed phenomenon. In the latter case, a modeler may be justified in adjusting model outputs in a way that (partially at least) accounts for the missing phenomenon, producing an adjusted model with better performance than the original. To be sure, a better approach would be to modify the original model such that it describes the missing process, with appropriate parameters and functions, and calibrate that, which should increase both r^2 and R^2, but this is not always possible due to model complexity and source code availability. Examples of model output adjustment for missing processes can be found in Nejadhashemi (2006) and Nejadhashemi et al. (2008) where various formulations were evaluated to account for baseflow contributions coming from outside of a watershed modeled using SWAT.

Where justified, a simple model output adjustment can be performed using a linear function, leading to an adjusted model with predictions, M, of the form:

$$M = a + b\,P$$

The statistics of the predictions of the adjusted model are then related to those of the original model by:

$$\overline{M} = a + b\,\overline{P} \quad , \quad \sigma_m^2 = b^2\,\sigma_p^2 \quad , \quad \sigma_{mo} = b\,\sigma_{po}$$

It is readily verified that the correlation coefficient between M and O is the same as that between P and O, and the adjustment parameters, a and b, can be chosen such that, for the predictions of the adjusted model, the coefficient of determination (between M and O) equals the square of the correlation coefficient (best case). This results in the specific form:

$$M = \left[\overline{O} - \frac{\sigma_{po}}{\sigma_p^2}\,\overline{P}\right] + \frac{\sigma_{po}}{\sigma_p^2}\,P = \left[\overline{O} - r\,\sqrt{\frac{\sigma_o^2}{\sigma_p^2}}\,\overline{P}\right] + r\,\sqrt{\frac{\sigma_o^2}{\sigma_p^2}}\,P$$

Remarkably, comparison with results of the previous section immediately shows that this linear model adjustment would also result from using a line model and performing linear least-square-error regression with the predictions of the original model, P, as independent variable and, as before, O, as the vector of observations (target). This is exactly the opposite of the process used earlier to identify the regression line, L, between the predictions of an arbitrary model (P as target variable) and observations (O as independent variable). For the present case of an adjusted model, the regression line is given by the special, best-case formula (which is a consequence of the model adjustment):

$$L = \left(1 - r^2\right)\,\overline{O} + r^2\,O \qquad : adjusted \text{ model}$$

Example

The relationship between coefficient of determination, coefficient of linear correlation and model parameter adjustment approach is illustrated with a hydrologic example. The observed data for this example is a single event streamflow hydrograph from USGS gauging station 01491000 for the period of October 12 to October 15, 2018. This station is located at the outlet of the Greensboro watershed, on the Eastern Shore of Maryland, which was recently studied by Renkenberger et al. (2016). Four models are used to provide approximations of the observed hydrograph: 1) a line; 2) the gamma-function Dimensionless Unit Hydrograph (DUH) described by equation 16-1 in the National Handbook of Engineering (NRCS, 2007); 3) a linear adjustment of the predictions of the DUH, and; 4) the DUH modified by the addition of a baseflow term:

$$\text{Model 1:} \quad P = a + b\,T$$

$$\text{Model 2:} \quad P = Q_p \left[\frac{T}{T_p}\, e^{\left(1 - \frac{T}{T_p}\right)} \right]^m$$

$$\text{Model 3:} \quad M = \left[\overline{O} - \frac{\sigma_{po}}{\sigma_p^2}\, \overline{P} \right] + \frac{\sigma_{po}}{\sigma_p^2}\, P$$

$$\text{Model 4:} \quad P = Q_p \left[\frac{T}{T_p}\, e^{\left(1 - \frac{T}{T_p}\right)} \right]^m + c + d\,T$$

where the input to Model 3 is the output of Model 2. The two parameters of Model 1 are identified by linear least-square-error regression while those of Models 2 and 4 are obtained using the nonlinear Levenberg-Marquardt optimization process.

The observed hydrograph and its statistics are presented in Figure 1, along with the predictions of the calibrated line model for the duration of the event. The linear function of time embodied by the model is clearly not a good representation of the observed hydrologic phenomenon. This is visually evident from the figure and is confirmed by the low value of the diagnostic statistics. However, and this is what is important in the present analysis, because the model parameters were derived using the linear least-square-error process, the coefficient of determination of model predictions, R^2, is equal to the square of the coefficient of linear correlation between predictions and observations, r^2. This is the highest value that R^2 can have, and, as discussed earlier, is attained in this special case of linear least-square-error regression. The model is a poor fit to the data, but it has the same mean value (such that the mean error is zero) and the variance of predictions equals their covariance with observations, both of which result specifically from the linear regression process used to identify the model's parameters ($1.0^{e\text{-}16}$ is essentially zero in IEEE-754 double-precision floating-point number representation).

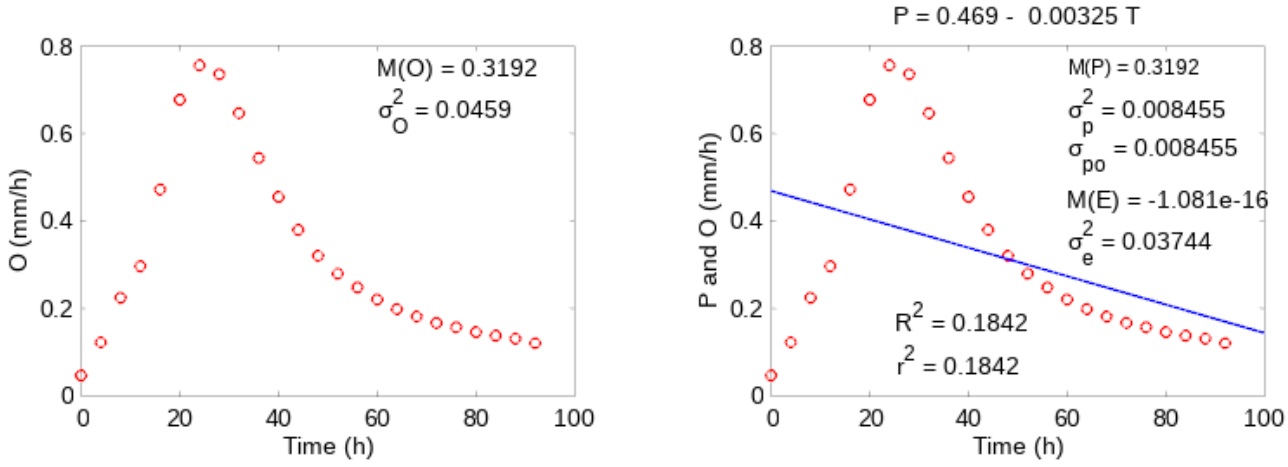

Figure 1. Event-Hydrograph Observation Data and Statistics (left) and Predictions of the Calibrated Model 1.

Figure 2 presents calibration results for the nonlinear NRCS DUH model. The DUH is a much more appropriate model of the observed event hydrograph than the line used by Model 1. This is evidenced by values of the coefficients of determination and correlation higher than 0.9 and a regression line, between predictions and observations, that has a small value of intercept and a slope close to 1. The nonlinear process used to parameterize the DUH guarantees neither a mean error of zero, nor that the covariance between predictions and observations equals the variance of P, and, consequently, the coefficient of determination is lower than the square of the coefficient of correlation, as shown on the figure. This is the general case encountered when calibrating nonlinear hydrologic models. It is notable, here, that the DUH does not include a representation of all processes which we expect to affect the observed hydrograph, and, in particular, it does explicitly excludes baseflow, which is commonly removed from an event hydrograph before fitting the DUH to the remaining direct runoff. The effect of baseflow is noticeable at early and large times, where the DUH tends to zero while the observed hydrograph has non-zero, positive, values.

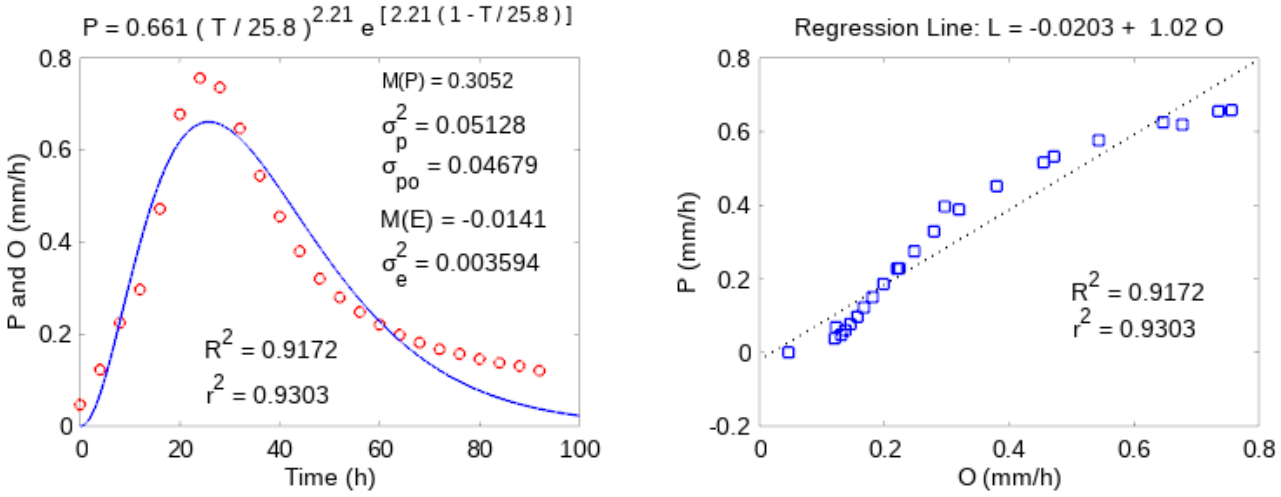

Figure 2. Predictions of the Calibrated NRCS DUH Model 2, versus Time (left) and versus Observations (right)

Model 3 represents a least-square-error regression adjustment of the predictions of Model 2, which may be thought of as a post-calibration correction for Model 2's missing baseflow component. Its predictions are compared to observations in Figure 3 where one observes that, at early and large times during the event, which is when baseflow dominates the hydrograph, Model 3 is closer to observations than Model 2. As described earlier, the adjustment embodied by Model 3 results in a mean error of zero between predictions and observations, and in the equality of the variance of the adjusted model's output with the covariance between this output and observations such that the special case of equality between coefficients of determination and correlation (squared) is reached. The correlation with observations remained the same as that of Model 2, only R^2 improved, but, simultaneously, the regression line moved slightly farther from the ideal one.

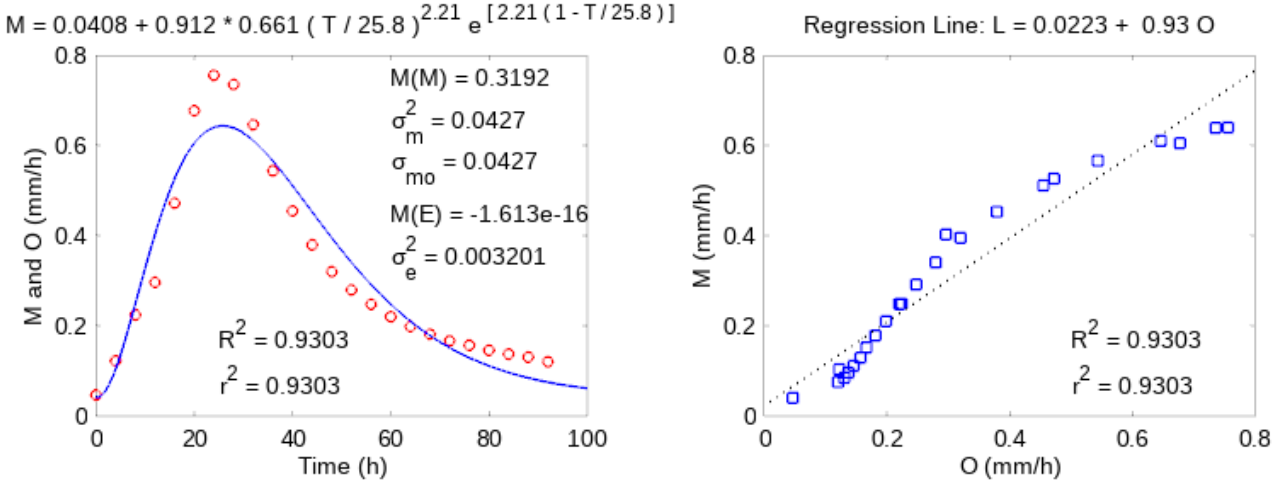

Figure 3. Predictions of the Adjusted NRCS DUH Model 3, versus Time (left) and versus Observations (right)

Model 4 adds a simple description of baseflow (as a linear trend in time) to the direct runoff DUH Model, resulting in a modified version of Model 2 and an alternative to post-calibration adjustments for missing processes. The nonlinear Levenberg-Marquardt process was used to adjust the 5 parameters of this model simultaneously, with results presented in Figure 4. The mean error of the optimized model is quite close to zero (1.0^{e-12} is significantly different from zero in IEEE-754 double-precision floating-point number representation) and the variance of predictions and their covariance with observations are nearly equal such that, for this specific model and dataset, the coefficient of determination is essentially not distinguishable from the square of the coefficient of correlation. These coefficients exceed 0.98, indicating a very good fit of the model to the data, better than with the post calibration adjustment implemented in Model 3, and this is confirmed by the regression line which has the smallest value of intercept and slope closest to 1.0 (ideal) of the 4 models evaluated in this example. This, obviously, matches one's expectation that a model which incorporates (accurate) descriptions of all key processes affecting a given phenomenon should provide better descriptions of said phenomenon than models that don't. There remains small deviations between predictions and observations at the earliest and largest times which would likely be eliminated with a more accurate baseflow submodel.

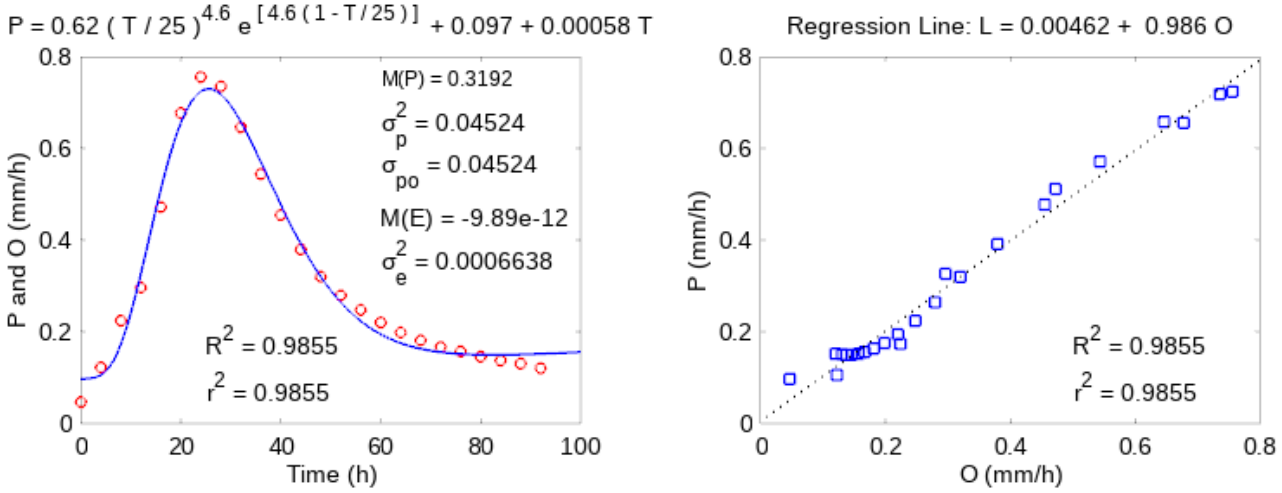

Figure 4. Predictions of the Modified NRCS DUH Model 4, versus Time (left) and versus Observations (right)

Conclusions

Nonlinear predictive models are important tools for engineering design and analysis. In hydrology in particular, accurate nonlinear models of watershed and field responses to meteorological inputs are expected to become increasingly important as climate change shifts the historical weather statistics on which past designs of canals, dams, storm drains, sewers, irrigation and drainage systems (among others) were based, towards location-specific new normals at which these designs are no longer effective, or fail, leading to increased flooding, decreased agricultural production, preventable accidents and major catastrophes. Unfortunately, there has been long-standing confusion about the meaning, names and formulas used to refer to some of the goodness-of-fit diagnostic statistics used during the calibration and validation of nonlinear hydrologic models, resulting potentially in sub-optimal parameterizations and sub-accurate models. The objective of this manuscript was to dissolve the components of this confusion related to three particular statistics: the coefficient of determination, R^2, the Nash-Sutcliffe coefficient of Efficiency, NSE, and the coefficient of linear correlation, r. To this effect, it was demonstrated, unequivocally, mathematically, using standard literature on statistics, and verified by example, that:

$$NSE = R^2 \qquad : always$$

and:

$$R^2 \leq r^2 \qquad : always$$

Furthermore, in the second formula, equality is not a most likely outcome of hydrologic model calibration as it requires zero mean error and equality of a specific variance and covariance, neither of which is guaranteed by nonlinear optimization. It is the exception, rather than the rule. It is hoped that these results, along with explicit formulas presented in the manuscript, will help modelers choose the most useful, least cross-correlated, most independent, and orthogonal, goodness-of-fit statistics, to efficiently calibrate their predictive models, towards the best possible accuracy.

References

Bicknell, B. R., J.C. Imhoff, J.L. Kittle Jr, A.S. Donigian Jr and R.C. Johanson, 1996. Hydrological simulation program-FORTRAN. User's manual for release 11. US EPA.

James, L.D. and S.J. Burges, 1982. Selection, Calibration, and Testing of Hydrologic Models, Chapter 11 in Haan, C.T., Johnson, H.P. and Brakensiek, D.L. (Eds) Hydrologic modeling of small watersheds, ASAE Monograph No. 5, American Society of Agricultural Engineers, Michigan, 1982, 533 pp.

Moriasi, D.N, M. W. Gitau and P. Daggupati, 2015. Hydrologic and Water Quality Models: performance measures and evaluation criteria. Transactions of the ASABE, 58(6):1763-1785.

Nash, J.E. and J.V. Sutcliffe, 1970. River flow forecasting through conceptual models Part I, a discussion of principles. Journal of Hydrology, 10(1970):282-290.

Neitsch, S.L., J.G. Arnold, J.R. Kiniry, J.R. Williams, 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute Technical Report No. 406, Texas A&M University System, College Station, Texas 77843-2118, 647pp.

Nejadhashemi, A.P., A. Shirmohammadi, H.J. Montas, J.M. Sheridan and D.D. Bosch, 2008. Watershed physical and hydrological effects on baseflow separation. Journal of Hydrologic Engineering, 13(10):971-980

Nejadhashemi, A.P., 2006. Improvement in Estimating Pollution Transport by Developing Streamflow Components Assessment in the GIS Environment. Ph.D. Thesis. University of Maryland at College Park, College Park, MD, 20742.

Neter, J., W. Wasserman and M.H. Kutner, 1990. Applied Linear Statistical Models, 3rd edition. Irwin pub., Boston, MA, 1990, 1181pp.

NRCS, 2007. National Engineering Handbook Part 630: Hydrology, Chapter 16: Hydrographs. USDA NRCS, Washington DC, 50pp.

Renkenberger, J., H. Montas, P.T. Leisnham, V. Chanse, A. Shirmohammadi, A. Sadeghi, K. Brubaker, A. Rockler, T. Hutson and D. Lansing, 2016. Climate Change Impact on Critical Source Area Identification in a Maryland Watershed. Transactions of the ASABE. 59(6): 1803-1819. doi:10.13031/trans.59.11677 .

Strang, G., 1988. Linear Algebra and its Applications, third edition. Harcourt Brace Jovanovich Inc., New York, NY, 505pp.

Appendix: Linear Regression (least square error)

Consider a linear model with m coefficients (U_1 ... U_m) used to predict a vector of n values, P, from the values produced by m functions applied to vectors of n values of a set of independent variables (T, X, Y, Z, ...):

$$P = U_1 F_1(T, X, Y, Z, \ldots) + U_2 F_2(T, X, Y, Z, \ldots) + \cdots + U_m F_m(T, X, Y, Z, \ldots)$$

The model is linear in the sense that its coefficients appear only at power 1, and are not introduced through nonlinear functions (eg. squaring, logarithms or trigonometric functions). For a model of this type, the vector of predictions can be computed concisely as:

$$P = F(T, X, Y, Z, \ldots) U = A U$$

where the vector of model parameters (or coefficients) is:

$$U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_m \end{bmatrix}$$

and the n-by-m matrix A is:

$$A = F(T, X, Y, Z, \ldots) = \begin{bmatrix} F_1(T, X, Y, Z, \ldots) & F_2(T, X, Y, Z, \ldots) & \cdots & F_m(T, X, Y, Z, \ldots) \end{bmatrix}$$

$$= \begin{bmatrix} F_1(T_1, X_1, Y_1, Z_1, \ldots) & F_2(T_1, X_1, Y_1, Z_1, \ldots) & \cdots & F_m(T_1, X_1, Y_1, Z_1, \ldots) \\ F_1(T_2, X_2, Y_2, Z_2, \ldots) & F_2(T_2, X_2, Y_2, Z_2, \ldots) & \cdots & F_m(T_2, X_2, Y_2, Z_2, \ldots) \\ \vdots & \vdots & & \vdots \\ F_1(T_n, X_n, Y_n, Z_n, \ldots) & F_2(T_n, X_n, Y_n, Z_n, \ldots) & \cdots & F_m(T_n, X_n, Y_n, Z_n, \ldots) \end{bmatrix}$$

Simple examples of linear models include the line, the parabola and the plane. The line and parabola depend on a single independent variable, say X, while the plane depends on two (X and Y for example). If the model is a line, then:

$$F_1(X) = 1 \quad , \quad F_2(X) = X \quad , \quad A = \begin{bmatrix} 1 & X \end{bmatrix} \quad , \quad P = U_1 + U_2 X$$

similarly, for a parabola:

$$F_1(X) = 1 \quad , \quad F_2(X) = X \quad , \quad F_3(X) = X^2 \quad , \quad A = \begin{bmatrix} 1 & X & X^2 \end{bmatrix} \quad , \quad P = U_1 + U_2 X + U_3 X^2$$

and, for a plane:

$$F_1(X, Y) = 1 \quad , \quad F_2(X, Y) = X \quad , \quad F_3(X, Y) = Y \quad , \quad A = \begin{bmatrix} 1 & X & Y \end{bmatrix} \quad , \quad P = U_1 + U_2 X + U_3 Y$$

The least square error regression process seeks to identify values of the model coefficients such that the sum of squared errors, SSE, between model predictions and a vector of n observations, O, is minimized. The vector of errors, E, and the SSE are given by (using the subscript T to indicate a transpose):

$$E = P - O = A U - O \quad , \quad SSE = E^T E$$

The coefficients that minimize the SSE are found where the gradient of the SSE, with respect to model coefficients, is zero. This gradient is computed as:

$$\frac{d\,SSE}{d\,U} = \frac{d}{d\,U} E^T E = 2\frac{d\,E^T}{d\,U} E = 2\frac{d\,(A\,U - O)^T}{d\,U}(A\,U - O) = 2\,A^T(A\,U - O)$$

and setting it to zero produces the linear algebraic equation to be solved in order to find the coefficients that minimize the sum of square deviations between model predictions and observations:

$$2\,A^T(A\,U - O) = 0 \quad \Rightarrow \quad A^T A\,U = A^T O \quad \Rightarrow \quad U = (A^T A)^{-1} A^T O$$

The solution can be obtained explicitly when the model is a line (values with overbars are averages):

$$P = A\,U = U_1 + U_2 X$$

$$U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \quad , \quad A = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad , \quad \frac{1}{n} A^T A = \begin{bmatrix} 1 & \overline{X} \\ \overline{X} & \dfrac{X^T X}{n} \end{bmatrix} \quad , \quad \frac{1}{n} A^T O = \begin{bmatrix} \overline{O} \\ \dfrac{X^T O}{n} \end{bmatrix}$$

for which (after mild rearrangements):

$$U_1 = \overline{O} - \frac{\sigma_{xo}}{\sigma_x^2}\,\overline{X} \quad , \quad U_2 = \frac{\sigma_{xo}}{\sigma_x^2}$$

$$P = \left[\overline{O} - \frac{\sigma_{xo}}{\sigma_x^2}\,\overline{X}\right] + \frac{\sigma_{xo}}{\sigma_x^2}\,X$$

where the variance of X and covariance between X and O are defined conventionally:

$$\sigma_x^2 = \frac{1}{n-1}\;\; x^T x \quad , \quad \sigma_{xo} = \frac{1}{n-1}\;\; x^T o$$

using mean-removed vectors (in lower-case):

$$x = X - \overline{X} \quad , \quad o = O - \overline{O}$$

For more complex models, a general results can be obtained in the specific case where the functions F1 ... Fm produce values that are not cross-correlated when applied to input vector T, X, Y, ... (which are themselves not cross-correlated). If statistics of the outputs of these functions are defined as:

$$\overline{F_i} = \frac{1}{n}\sum_{k=1}^{n} F_i(T_k, X_k, Y_k, Z_k, \ldots)$$

$$\sigma_{f_i}^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left[F_i(T_k, X_k, Y_k, Z_k, \ldots) - \overline{F_i}\right]^2$$

$$\sigma_{f_i f_j} = \frac{1}{n-1}\sum_{k=1}^{n}\left(F_i(T_k, X_k, Y_k, Z_k, \ldots) - \overline{F_i}\right)\left(F_j(T_k, X_k, Y_k, Z_k, \ldots) - \overline{F_j}\right)$$

then the no cross-correlation condition means that:

$$r_{f_i f_j} = \frac{\sigma_{f_i f_j}}{\sqrt{\sigma_{f_i}^2 \sigma_{f_j}^2}} = \begin{cases} 1 & , \quad i = j \\ 0 & , \quad i \neq j \end{cases}$$

Ortho-normal polynomials form a convenient set of functions with these characteristics, with the additional feature that their mean is 0 and their 2nd central moment (variance scaled by n rather than n-1) is 1, except for F_1 which has a mean of 1

and variance of 0. A paraboloid model provides a non-trivial illustrative example:

$$F_1(X,Y)=1 \quad , \quad F_2(X,Y)=\frac{X-\overline{X}}{\mu_{2[x]}^{1/2}}=\frac{x}{\mu_{2[x]}^{1/2}}=\lambda_x \quad , \quad F_3(X,Y)=\frac{Y-\overline{Y}}{\mu_{2[y]}^{1/2}}=\frac{y}{\mu_{2[y]}^{1/2}}=\lambda_y$$

$$F_4(X,Y)=\frac{\lambda_x^2-\overline{\lambda_x^3}\,\lambda_x-1}{\sqrt{\overline{\lambda_x^4}-\left(\overline{\lambda_x^3}\right)^2-1}}=\gamma_{x^2} \ , \ F_5(X,Y)=\frac{\lambda_x\lambda_y}{\sqrt{\overline{\lambda_x^2\lambda_y^2}}}=\gamma_{xy} \ , \ F_6(X,Y)=\frac{\lambda_y^2-\overline{\lambda_y^3}\,\lambda_y-1}{\sqrt{\overline{\lambda_y^4}-\left(\overline{\lambda_y^3}\right)^2-1}}=\gamma_{y^2}$$

with relevant statistics:

$$\overline{\lambda_x}=0 \quad , \quad \overline{\lambda_x^2}=1 \quad , \quad \overline{\lambda_x^3}=\frac{\mu_{3[x]}}{\mu_{2[x]}^{3/2}} \quad , \quad \overline{\lambda_x^4}=\frac{\mu_{4[x]}}{\mu_{2[x]}^2}$$

$$\overline{\lambda_y}=0 \quad , \quad \overline{\lambda_y^2}=1 \quad , \quad \overline{\lambda_y^3}=\frac{\mu_{3[y]}}{\mu_{2[y]}^{3/2}} \quad , \quad \overline{\lambda_y^4}=\frac{\mu_{4[y]}}{\mu_{2[y]}^2}$$

$$\overline{\lambda_x\lambda_y}=\overline{\lambda_x^2\lambda_y}=\overline{\lambda_x\lambda_y^2}=\overline{\lambda_x^3\lambda_y}=\overline{\lambda_x\lambda_y^3}=0 \quad , \quad \overline{\lambda_x^2\lambda_y^2}=\frac{\mu_{2[xy]}}{\mu_{2[x]}\mu_{2[y]}}$$

$$\overline{\gamma_{x^2}}=\overline{\gamma_{y^2}}=\overline{\gamma_{xy}}=0 \quad , \quad \overline{\gamma_{x^2}^2}=\overline{\gamma_{y^2}^2}=\overline{\gamma_{xy}^2}=1 \quad , \quad \overline{\gamma_{x^2}\gamma_{y^2}}=\overline{\gamma_{x^2}\gamma_{xy}}=\overline{\gamma_{y^2}\gamma_{xy}}=0$$

$$\overline{\lambda_x\gamma_{x^2}}=\overline{\lambda_x\gamma_{y^2}}=\overline{\lambda_x\gamma_{xy}}=\overline{\lambda_y\gamma_{x^2}}=\overline{\lambda_y\gamma_{y^2}}=\overline{\lambda_y\gamma_{xy}}=0$$

and central moments given by:

$$\mu_{2[x]}=\overline{x^2}=\frac{1}{n}\sum_{i=1}^{n}x_i^2=\frac{n-1}{n}\sigma_x^2 \quad , \quad \mu_{3[x]}=\overline{x^3}=\frac{1}{n}\sum_{i=1}^{n}x_i^3 \quad , \quad \mu_{4[x]}=\overline{x^4}=\frac{1}{n}\sum_{i=1}^{n}x_i^4$$

$$\mu_{2[y]}=\overline{y^2}=\frac{1}{n}\sum_{i=1}^{n}y_i^2=\frac{n-1}{n}\sigma_y^2 \quad , \quad \mu_{3[y]}=\overline{y^3}=\frac{1}{n}\sum_{i=1}^{n}y_i^3 \quad , \quad \mu_{4[y]}=\overline{y^4}=\frac{1}{n}\sum_{i=1}^{n}y_i^4$$

$$\mu_{2[xy]}=\overline{x^2y^2}=\frac{1}{n}\sum_{i=1}^{n}x_i^2y_i^2$$

Applying linear least-square-error regression to the orthonormal paraboloid model produces:

$$P=AU=U_1+U_2\lambda_x+U_3\lambda_y+U_4\gamma_{x^2}+U_5\gamma_{xy}+U_6\gamma_{y^2}$$

$$U=\begin{bmatrix}U_1\\U_2\\\vdots\\U_6\end{bmatrix} \quad , \quad A=\begin{bmatrix}1 & \lambda_{x[1]} & \lambda_{y[1]} & \gamma_{x^2[1]} & \gamma_{xy[1]} & \gamma_{y^2[1]}\\1 & \lambda_{x[2]} & \lambda_{y[2]} & \gamma_{x^2[2]} & \gamma_{xy[2]} & \gamma_{y^2[2]}\\\vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\1 & \lambda_{x[n]} & \lambda_{y[n]} & \gamma_{x^2[n]} & \gamma_{xy[n]} & \gamma_{y^2[n]}\end{bmatrix}$$

$$\frac{1}{n}A^TA=\begin{bmatrix}1 & 0 & \cdots & 0\\0 & \ddots & \ddots & \vdots\\\vdots & \ddots & \ddots & 0\\0 & \cdots & 0 & 1\end{bmatrix} \quad , \quad \frac{1}{n}A^TO=\begin{bmatrix}\overline{O}\\\overline{\lambda_x o}\\\vdots\\\overline{\gamma_{y^2}o}\end{bmatrix}$$

and, since A^TA is the identity matrix (a consequence of using ortho-normal polynomials and of the assumption of no cross-correlation between model input variables), one obtains model parameters at once:

$$U_1 = \overline{O} \quad , \quad U_2 = \overline{\lambda_x o} = \sqrt{\frac{n-1}{n}}\ \frac{\sigma_{xo}}{\sqrt{\sigma_x^2}} \quad , \quad U_3 = \overline{\lambda_y o} = \sqrt{\frac{n-1}{n}}\ \frac{\sigma_{yo}}{\sqrt{\sigma_y^2}}$$

$$U_4 = \overline{y_{x^2} o} = \sqrt{\frac{\mu_{2[x]}^3}{\mu_{4[x]}\mu_{2[x]} - \mu_{3[x]}^2 - \mu_{2[x]}^3}} \left[\frac{\sigma_{x^2 o}}{\sigma_x^2} - \mu_{3[x]} \frac{\sigma_{xo}}{\sigma_x^2} \right] \quad , \quad U_5 = \overline{y_{xy} o} = \sqrt{\frac{n-1}{n}}\ \frac{\sigma_{xyo}}{\sqrt{\sigma_{x^2 y^2}}}$$

$$U_6 = \overline{y_{y^2} o} = \sqrt{\frac{\mu_{2[y]}^3}{\mu_{4[y]}\mu_{2[y]} - \mu_{3[y]}^2 - \mu_{2[y]}^3}} \left[\frac{\sigma_{y^2 o}}{\sigma_y^2} - \mu_{3[y]} \frac{\sigma_{yo}}{\sigma_y^2} \right]$$

The sequence of ortho-normal polynomials contributes incrementally to the model. Accordingly, the least-square-error line model obtained earlier also results from the above by considering only the first 2 terms:

$$P = U_1 + U_2 \lambda_x = \overline{O} + \left(\sqrt{\frac{n-1}{n}}\ \frac{\sigma_{xo}}{\sqrt{\sigma_x^2}} \right) \frac{X - \overline{X}}{\sqrt{\frac{n-1}{n}\sigma_x^2}} = \left[\overline{O} - \frac{\sigma_{xo}}{\sigma_x^2}\ \overline{X} \right] + \frac{\sigma_{xo}}{\sigma_x^2}\ X$$

and the least-square-error plane model results from the first 3 terms:

$$P = U_1 + U_2 \lambda_x + U_3 \lambda_y = \left[\overline{O} - \frac{\sigma_{xo}}{\sigma_x^2}\ \overline{X} - \frac{\sigma_{yo}}{\sigma_y^2}\ \overline{Y} \right] + \frac{\sigma_{xo}}{\sigma_x^2}\ X + \frac{\sigma_{yo}}{\sigma_y^2}\ Y$$