Environmental Geology (1993) 22: 121–128

T. J. Toy · W. R. Osterkamp · K. G. Renard

Prediction by regression and intrarange data scatter in surface-process studies

Received: 28 December 1992 / Accepted: 4 January 1993

Abstract Modeling is a major component of contemporary earth science, and regression analysis occupies a central position in the parameterization, calibration, and validation of geomorphic and hydrologic models. Although this methodology can be used in many ways, we are primarily concerned with the prediction of values for one variable from another variable. Examination of the literature reveals considerable inconsistency in the presentation of the results of regression analysis and the occurrence of patterns in the scatter of data points about the regression line. Both circumstances confound utilization and evaluation of the models. Statisticians are well aware of various problems associated with the use of regression analysis and offer improved practices; often, however, their guidelines are not followed. After a review of the aforementioned circumstances and until standard criteria for model evaluation become established, we recommend, as a minimum, inclusion of scatter diagrams, the standard error of the estimate, and sample size in reporting the results of regression analyses for most surface-process studies.

Key words Modeling — Regression analysis — Data scatter — Surface-process studies

Introduction

Relevant literature demonstrates that modeling has become a major component of contemporary earth science.

T. J. Toy
Department of Geography, University of Denver, Denver, Colorado 80208, USA

W. R. Osterkamp
US Geological Survey, Box 25046, MS 413, Denver Federal Center, Lakewood, Colorado 80225, USA

K. G. Renard
USDA Agricultural Research Service, Aridland Watershed Management Research Unit, 2000 E. Allen Road, Tucson Arizona 85719, USA

Correspondence to: T. J. Toy

Foster (1988) remarked that physically based research models are important tools used by scientists and research managers to enlarge their understanding of runoff and erosion processes, to identify research needs, and to set research priorities. Further, he notes that within the last decade, runoff and erosion modeling has become used increasingly to deal with "real-world" applied problems. Haan (1988) observed that virtually all hydrologic design is based on the results of applying a hydrologic model.

Knisel (1980) commented that the mathematical modeling of the processes in erosion and sediment transport has increased significantly since the passage of the Clean Water Act, Public Law 92-500, in 1972. Renard and Meyer (1986) submitted that the United States Congress recognized the potential of models when it instructed the Office of Technology Assessment to: (1) assess the nation's ability to use models efficiently and effectively in analyzing and solving water resource problems, and (2) provide recommendations for improving the use of available technologies.

However, modeling is not yet the panacea for solving all geomorphic and hydrologic problems. The American Society of Civil Engineers Task Committee on Evaluation Criteria for Watershed Models (ASCE 1990) identified three major limitations in the presentation of hydrologic models in today's literature: (1) developers typically do not provide statistical evaluation criteria to assist the user in determining how well the model reproduces the measured data or how well it compares to other models; (2) most models are not tested for a wide range of conditions, so the user cannot evaluate where and when the particular model should be applied and how different models work for different conditions; and (3) the models are usually not documented sufficiently for a user to select parameters effectively, apply the model, and then evaluate the results. In addition, there are often inadequate data with which to compare simulations by various models. The ASCE Task Committee (ASCE 1990, p. 393) concluded that "currently, the evaluation of watershed models is a subjective business." Wagenet (1988) and Foster (1988) provided further discussion of limitations associated with various types of models.

Regression analysis is a statistical tool that plays a key role, both directly and indirectly, in a substantial proportion of earth science modeling. It is a versatile and powerful technique for predicting one entity from another and then comparing the predicted and measured values. In this report, we focus upon the use of regression analysis for prediction. Although criticized as an empirical method, Foster (1988) acknowledged that our lack of complete understanding (*of physical processes*) requires a degree of empiricism even in the most process-based model.

As with any tool, regression analysis has the potential for abuse. In fact, regression or structural analyses are almost always, to some degree, inappropriate methods for earth science studies, usually as a consequence of the assumptions upon which they are predicated and also because there is often no clear distinction between the dependent and independent variables. Diskin (1970) discussed problems associated with evaluating the parameters of linear regression models used for rainfall runoff modeling. Mark and Church (1977) described conditions under which least-squares regression is properly utilized and circumstances when structural analysis is a more appropriate alternative. Subsequently, Osterkamp and others (1978) provided guidelines for choosing between regression and structural analysis in geomorphic studies. Williams (1983) cautioned against the improper use of regression equations to estimate the independent variable, and Wagenet (1988, p. 7) commented that "the application of models beyond the experimental experience used in their development is in many respects an extrapolation exercise; the perils of ill-considered extrapolation are well-covered in many elementary texts." Indeed, regression equations as utilized in the earth sciences are infrequently robust.

Perusal of the contemporary geomorphic and hydrologic literature reveals numerous examples of regression analyses, demonstrating that it is the single most widely used technique for defining presumed cause-and-effect relationships. This examination further reveals considerable inconsistency in the presentation of results from these analyses. Commonly, a graphic display includes the scatter of data upon which the regression is based. However, sometimes only the derived regression line is provided, making it difficult to evaluate the utility of the model even across the range of values depicted on the axes. Commonly, the correlation coefficient (r) or coefficient of determination (r^2) is provided from which one is apparently expected to infer error (or absence thereof) associated with predictions from the model. In addition, for data sets of any size, there are frequently patterns in the scatter of data points about the regression line; in other words, the data are in proximity to the regression line through a part of the distribution, but scattered through another part of the distribution. This can occur despite the transformation of data. Each of these three situations—provision of the regression line alone, utilization of the correlation coefficient, and pattern in the distribution of data points about the regression line—confound evaluation of the model by the reader and potential user.

It is the purpose of this paper to remind earth scientists of various possibilities for misuse of regression analysis and to offer recommendations for standard reporting practices based upon consideration of the three aforementioned situations. The potential for improper use of regression techniques is well recognized by statisticians and many earth scientists, but sometimes overlooked by the occasional practitioner. Further, methods for handling problematic data are often more theoretical than practical, requiring knowledge about the data that is difficult to obtain or necessitating complex computations for which computer software may not be available. It is a thesis of this paper that scientific progress in the earth sciences can be placed upon a stronger statistical foundation through a complete presentation of the results from regression analysis.

Graphic display

There seem to be few cases that warrant the omission of scatter diagrams from the presentation of regression results. The regression line alone connotes a functional relation between independent and dependent variables. The real world is rarely that simple, even when measured accurately. Furthermore, the line alone deprives the reader and potential user of the opportunity to judge the value of the model through the range of data of particular interest. Lastly, outliers that produce large residuals are frequently "research hypotheses generators," as we search for their cause.

Errors in prediction

Regression is generally used for statistical prediction, whether predicting the dependent variable from the independent variable or predicting the errors that may be expected when utilizing a model. Obviously, accuracy is one important component in evaluating and choosing a particular model (Renard and Meyer 1986; Woolhiser and Brakensiek 1982; Dawdy and Lichty 1968; among others). There is a substantial body of literature concerning this topic under keywords such as "error analysis," "uncertainty analysis," and "influence analysis." Hardison (1971) provided a valuable comparison of the standard error of prediction and the standard error of the estimate for estimation of streamflow properties from drainage-basin characteristics. Troutman (1985a,b) presented an excellent treatise on errors associated with the estimation of runoff from rainfall. Haan (1988) discussed the subject of "uncertainty" in hydrologic modeling, and Gilroy (1991) described "uncertainty" in estimates of stream sediment loads. Tasker and Stedinger (1989) briefly discussed the "influence" or "leverage" that individual data points can exert on the prediction of streamflow properties from physiographic basin characteristics.

Various statistics are used to express the probable accuracy or error associated with a particular model. Perhaps the most commonly utilized are the correlation coefficient (r) and the coefficient of determination (r^2). The former is intended to express the strength of the relation between two variables without regard to their assumed dependence or independence. The latter expresses the percentage of the variance in the dependent variable that is "explained" by the independent variable, or, put otherwise, the coefficient of determination indicates the improvement in prediction that might be expected by using the independent variable for that purpose rather than merely the mean of the dependent variable. Neither statistic, however, is designed to assess the probable errors in prediction that are likely to result from the utilization of a particular regression model.

The standard error of the estimate (SE_{yx}) is used to indicate, with a given level of confidence, the range within which a predicted value should lie. It is, in practice, equivalent to the standard deviation of the residuals, except that SE_{yx} relates to an infinite sampling distribution while the standard deviation of the residuals relates to a finite distribution of actual observations (Hammond and McCullagh 1974). Of course, the standard error of the estimate, correlation coefficient, and coefficient of determination are related; as the standard error of the estimate increases, the correlation coefficient and coefficient of determination decrease. In addition, the standard error of the estimate is a well-known statistic and most regression analysis computer programs provide the statistic.

Haan (1988) acknowledged that there is no widely accepted criterion on which to base the evaluation of a model. As something of an expedient compromise, we suggest that the standard error of the estimate should be given as part of the regression analysis, in addition to the correlation coefficient or the coefficient of determination. Furthermore, the sample size for a particular analysis should be stated explicitly so that degrees of freedom for tests of significance can be readily ascertained.

Dispersion about regression line

Although several potential and actual problems inherent in the use of regression are treated in the literature, there is little discussion regarding patterns in the scatter of data about the regression line. Patterns are a likely consequence of an uneven distribution of data in a particular set. A result can be significant differences in the variance of the dependent variable through the range of values for the independent variable (referred to as "heteroscedasticity" of the data); these differences may persist despite data transformation. Sorooshian (1981) discussed this problem in relation to streamflow errors. The use of regression analysis in such circumstances violates a fundamental assumption upon which the technique is predicated, but there are plenty of such examples of misuse to be found in the literature. Patterns in the data itself can cause subsequent patterns in the residuals from regression analysis. We identify five basic patterns: (1) a balanced distribution of data points about the regression line, (2) greater scatter about the line at high values, (3) greater scatter about the line at low values, (4) greater scatter about the line in mid-range values, and (5) concentration of data points at low values. These patterns are illustrated in Figs. 1–5. Conceivably, a pattern could exist with scatter at both high and low values while approximating the regression line in the mid-range, but we did not find examples. Lastly, we will briefly discuss some consequences of reporting only the mean values of experimental repetitions, rather than the entire data set, as shown in Figs. 6–7.

Generic examples

In order to address this issue generically, data sets were constructed that possess the identified distribution patterns. For the first four patterns, the number of data points and the interval between the values of the independent variable remained the same. In addition, the sum of the squared deviations of the actual dependent variable from the predicted dependent variable, $\sum(y - y_p)^2$, remain the same. Hence, the standard error of the estimate is approximately the same for all four patterns (within rounding error). Subsequently, these data sets were divided into ranges of the independent variable for the purpose of comparison and discussion. The actual values of the standard errors of the estimate are a function of the constructed data sets, but it is the relative values within and between data sets that are germane here. Table 1 presents the summary of the regression analysis results for these data sets. Note that for the four basic patterns, the standard errors of the estimate, regression coefficients, correlation coefficients, and coefficients of determination are very similar, although the patterns of data-point distribution are very different.

Figure 1 depicts a pattern in which the data points are balanced about the regression line. This is the distribution inferred by the correlation coefficient, coefficient of determination, and standard error of the estimate. The errors that might be expected using this model for prediction are the same throughout the range of values shown.

Figure 2 depicts a common pattern in which the data points are closely associated with the regression line at low values, but scattered at high values. Although the descriptive statistics are about the same as for the balanced pattern, it is evident that the model predicts very well for the low values but poorly for the high values. The standard error of the estimate represents the average situation, suggesting that the model is less accurate than actually for low values but more accurate than actually for high values. The data included in this set can be partitioned into high and low ranges for comparison. The standard error of the estimate in the low range is very small while that for the high range is much larger. Likewise the correlation coefficient and coefficient of determination are greater for the low range than for the high range.

Figure 3 depicts a complementary pattern with scatter of the data at low values and close association with the regression line at high values. Again, the correlation coefficient, coefficient of determination, and standard error of

Table 1. Statistical summary

Distribution	Standard error	Regress. constant	Regress. coeff	Correl. coeff	Coeffic. deter. (%)	Sample size
Balanced	4.133	0.414	0.987	0.974	94.8	30
High-end scatter	4.137	0.276	0.991	0.974	94.9	30
Low range	0.531	0.100	0.994	0.998	99.7	16
High range	6.260	3.719	0.921	0.788	62.1	14
Low-end scatter	4.137	0.276	0.991	0.974	94.9	30
Low range	6.029	0.962	0.942	0.823	67.8	15
High range	0.536	−0.033	1.000	0.998	99.8	15
Middle scatter	4.140	0.083	0.997	0.974	94.9	30
Middle	6.296	0.545	0.982	0.806	64.9	14
High/low	0.534	0.031	0.999	1.000	100.0	16
Low-end concentration	4.094	−0.429	1.034	0.964	92.9	30
Low range	1.032	0.263	0.965	0.965	93.1	26
Data versus means						
Data set	4.043	1.200	0.966	0.973	94.7	30
Means	2.342	1.200	0.966	0.993	98.7	6
Santa Ana River	0.014	2.312	0.960	0.783	61.3	1461

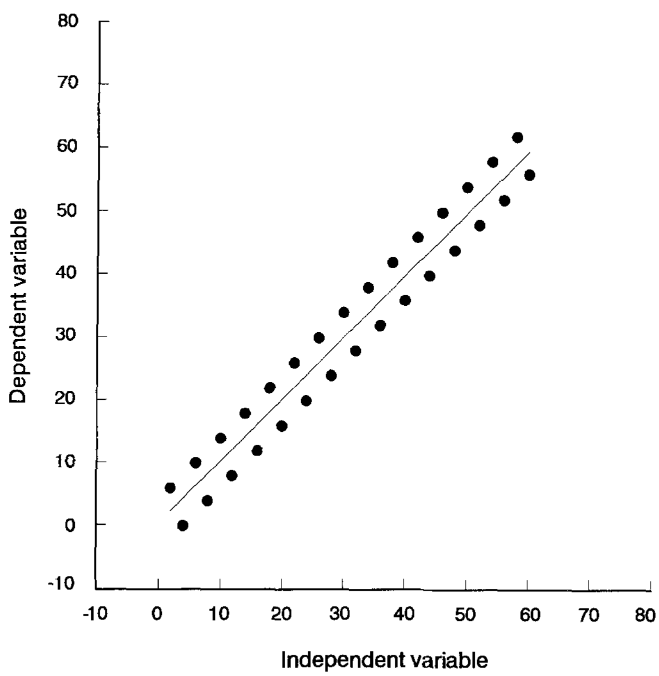

Fig. 1. Balanced distribution of data about the regression line

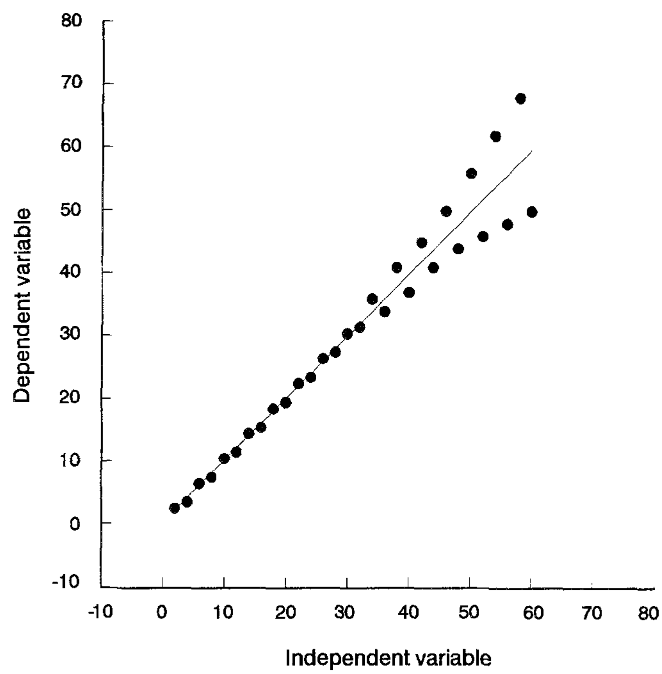

Fig. 2. Greater scatter of data at high values than at low values

the estimate are about the same as for the previous two patterns. Now, the standard error for the low range is several times larger than for the high range.

Figure 4 depicts a less common scenario with a pattern of scatter in the middle of the data set. As before, the correlation coefficient, coefficient of determination, and standard error of the estimate are about the same as for the previous three patterns. This model predicts well at high and low values but poorly in the middle range.

Often in geomorphic and hydrologic investigations systematic collection of data at regular time intervals yields many measurements of processes operating at low rates and a few measurements of processes operating at substantially higher rates. This produces a pattern in the distribution of data points as depicted in Fig. 5. For the purpose of comparison, this data set was constructed so that the standard error of the estimate is similar to those of the previous four cases. The standard error of the estimate for the entire distribution is much larger than for the concentration of data points at low values. It is evident that the standard error of the estimate is weighted by the large number of low value data points.

Lastly, mean values are sometimes used to represent data collected through several replications of an experiment and the subsequent regression analyses are based upon these mean values alone. This practice may produce a distribution of data similar to that depicted in Fig. 6, whereas the entire data set might look something like that

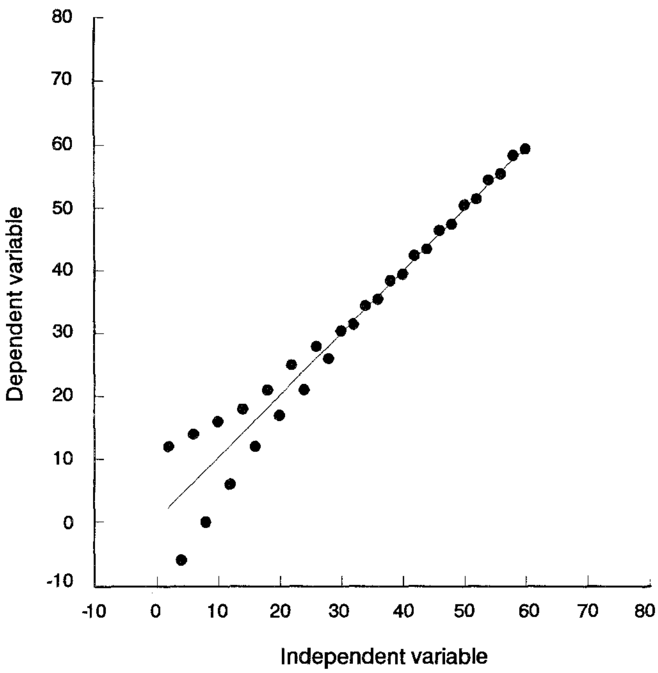

Fig. 3. Greater scatter of data at low values than at high values

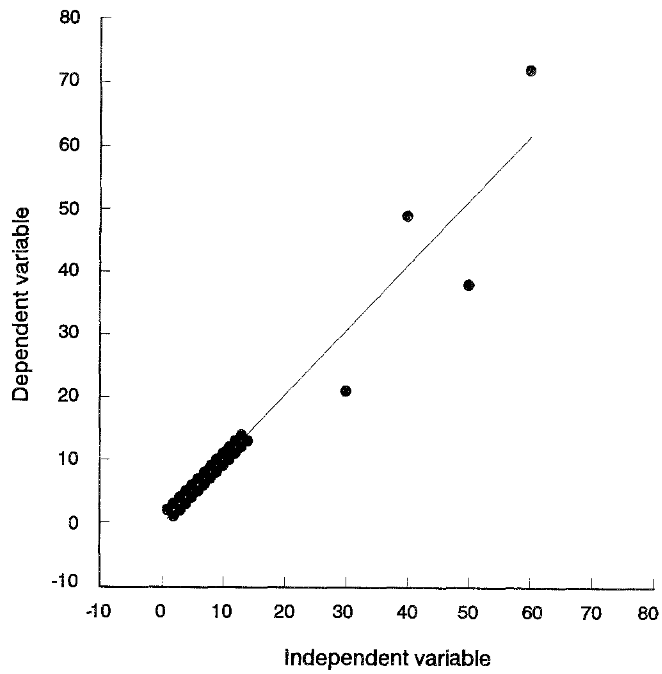

Fig. 5. Concentration of data at low values

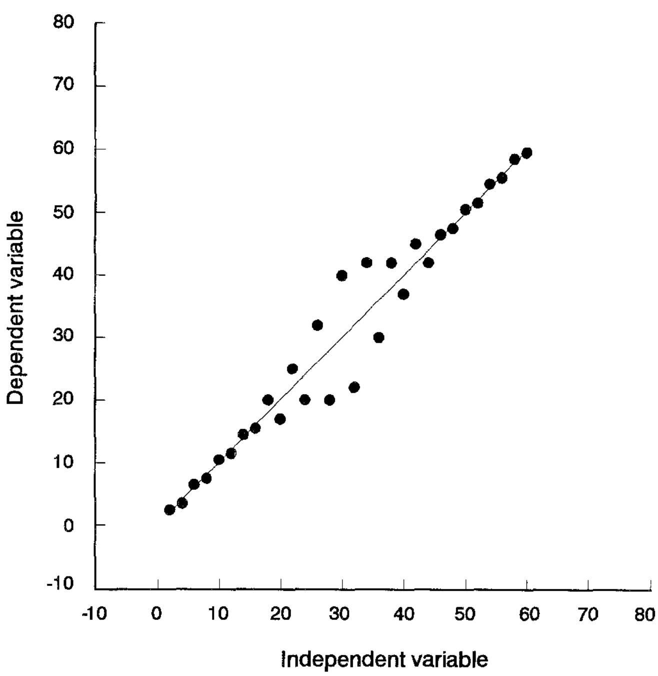

Fig. 4. Greater scatter of data at mid-range than at high or low values

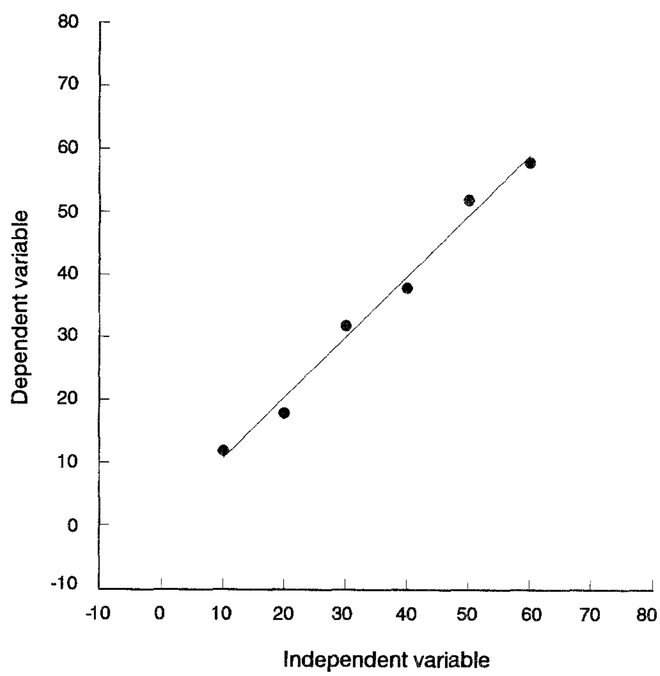

Fig. 6. Mean value of data collected through repeated experimentation

depicted in Fig. 7. The correlation coefficient, coefficient of determination, and standard error of the estimate for the entire data set are similar to those for the previous five patterns. However, the standard error of the estimate based upon the means alone is considerably less. Thus, this practice of utilizing only the means tends to substantially underestimate the amount of error that may be expected when using the model. Note also that the regression constant and coefficient remained the same, while the correlation coefficient and coefficient of determination changed slightly. The latter statistics tended to disguise the change in the standard error of the estimate.

Actual example

A regression between stream discharge and sediment concentration is frequently presented in the geomorphic and hydrologic literature. It is usually accepted that each vari-

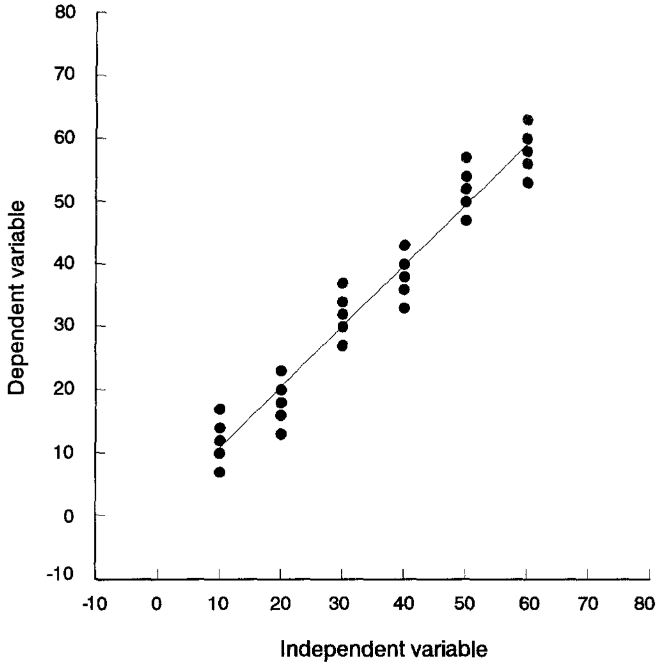

Fig. 7. Distribution of all data collected through repeated experimentation

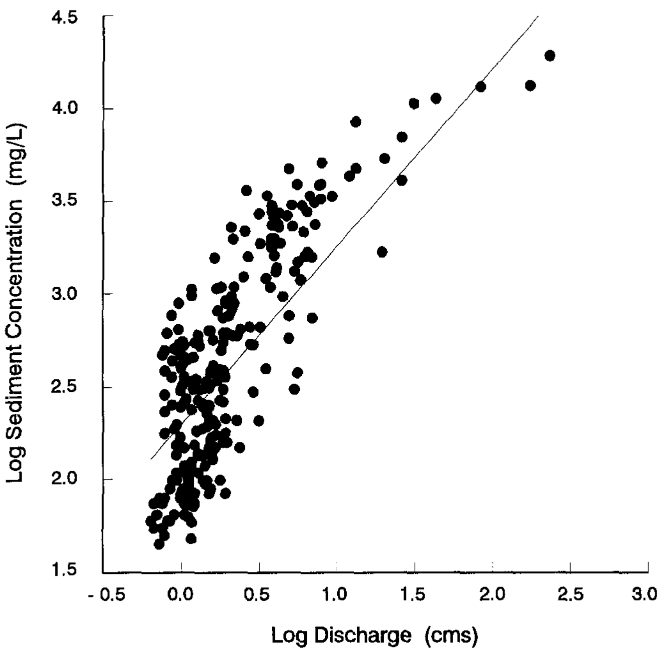

Fig. 8. Relation between discharge and sediment concentration for the Santa Ana River, Prado Park, California

able requires logarithmic$_{10}$ transformation. A scatter diagram of such data often resembles Fig. 5, with most of the data points concentrated toward lower values, representing many measurements during small magnitude events, and a few data points toward higher values representing occasional large magnitude events. Figure 8 shows such a regression for mean daily discharge and sediment concentration data from the US Geological Survey gaging station on the Santa Ana River at Prado Park near Corona, California (R. S. Parker, US Geological Survey, personal communication, 1992). From the time series of data through the period 1 October, 1976 to 30 September, 1980, the values for every fourth day are plotted.

First, it is evident that there is considerable scatter of data about the regression line. In fact, the distribution seems to possess features of both Figs. 3 and 5. Generally, the scatter appears widest for lower values and narrows some toward higher values. Further, it would be expected that the few data points at the higher values substantially influence, or leverage, the orientation of the regression line. Finally, it might be argued that, despite data transformation, the scatter indicates a curvilinear, convex-upward, relation, rather than a simple logarithmic relation. These features affect the accuracy of the prediction that are made from this regression analysis.

Commentary

Implicit in most models is the notion that they predict equally well throughout the range of data from which they were developed. The preceding indicates that this is often not the situation. The correlation coefficient, coefficient of determination, or standard error of the estimate do not reveal the existence or nature of intrarange data scatter. This is one good reason why the inclusion of scatter diagrams is of paramount importance in the presentation of the results from regression analyses.

There are various reasons why patterns such as those shown in Figs. 2–4 may occur in data sets. Perhaps a measurement techniques or instrumentation was utilized beyond its effective range. Perhaps a geomorphic or hydrologic threshold was surpassed during the recording period, triggering significant change within the system under investigation. For example, at some discharge along a rating curve, the flow may produce sufficient shear stress at the channel bed and banks to cause erosion, alteration of channel morphology, and modification of the stage/discharge relation. Weathering of bank material and saturation may contribute by reducing channel bank shear strength. Until a new equilibrium is attained along the channel reach, the relation might be expected to exhibit a measure of scatter. Perhaps the contribution of water and sediment from tributaries varies seasonally or with precipitation events of various magnitudes.

The place or places where a strong relation seems not to hold are precisely the places of greatest scientific interest. The data points on or very near the regression line generally confirm that which we already knew or suspected. The data points that depart significantly, as revealed by large residual values, direct our attention to the unknown or unsuspected. Such distributions confirm that the physical environment is more complicated than the linear model infers and perhaps a linear model itself is inappropriate.

The options for handling data sets that contain patterns in the distribution of points range from relatively simple to quite complicated. First, as suggested earlier, we are obliged to provide a graphic display of the data so that potential users can judge for themselves whether the model is useful through the range of data of interest to them. It may be prudent to divide the data into subsets, as suggested by Diskin (1970). The relation between the independent and dependent variables might be linear through one range of values but curvilinear through another range, perhaps following a threshold. Possibly, one independent variable is sufficient to predict a dependent variable through one range of values but other independent variables become important through another range. It may be that above or below certain ranges of values additional geomorphic or hydrologic processes or suite of processes become operational. Concerning precipitation runoff modeling, Troutman (1985a) remarked that: "we are at liberty to state that the model is correctly specified only for large events and simply discard the predictions for small events because we know that the model does not work very well for these events.... We may go even further if we can state that there is one set of parameters, say B_1, that gives good predictions for small events, and another set, say B_2, that gives good predictions for large events." The foregoing possibilities should encourage us to examine our measurement techniques and instrumentation to assure ourselves that the distributions are true reflections of reality.

From a statistical perspective, there are tools to assist in dealing with problematic data sets. These include data transformations, robust regression, weighted least-squares regression (WLS), and generalized least-squares regression (GLS). Further information concerning the application of these techniques is available in Sorooshian (1981), Hoaglin and others (1985), Troutman (1985a,b), and Tasker and Stedinger (1989).

Nevertheless, as noted by Woolhiser and Brakensiek (1982, p. 16), objective methods for choosing the best model have not yet been developed; "the final choice of the best model will depend upon the problem, the resources available to the analyst, the time frame available, the input resources available, and a number of other implicit criteria like experience, and maybe even 'horse sense'."

Discussion and conclusion

In discussing association and indeterminacy in geomorphology, Leopold and Langbein (1967) commented:

> The landscape, in other words, exhibits a variability which may be expected as a result of incomplete dynamic determinacy. General physical laws are necessary but not sufficient to determine the exact shape of each land form. Some scatter of points on graphs showing interrelations between factors is expected, although the mean or median condition is reproducible in different sets of samples.

The consequences of this indeterminacy is reflected in the observation of Wagenet (1988):

> It is now obvious that existing models do not predict exactly the field-measured values of any property or process. ... this indicates that model builders not only must indicate the confidence limits of their predictions as a part of future model building and application exercises, they also must educate the balance of the world that uncertainty is an inescapable component of the natural systems.

Accordingly, some dispersion of data about the regression line is a realistic expectation. It is incumbent upon geomorphologists and hydrologists to develop strategies to accommodate this eventuality. The ASCE Task Committee on Evaluation Criteria for Watershed Models (ASCE 1990) offers two general recommendations: (1) When model results are presented, always provide statistical criteria to assist in evaluating model performance; graphical plots should be presented together with the statistical criteria. (2) Attempt to keep the presentation of results simple by including only the minimum number of criteria necessary. The committee suggests that implementation of these and other recommendations be mandatory for publication in ASCE journals in the same way that the use of SI units is required.

Until criteria and standards for model evaluation are developed and adopted, we advocate, as a minimum, routine inclusion of scatter diagrams and the standard error of the estimate in addition to, if not in preference to, the correlation coefficient or coefficient of determination, in the presentation of the results from regression analyses. In addition, the sample size for the analysis should be explicitly stated. Eventually, somewhat less familiar statistics, such as the standard error of prediction, and somewhat less familiar techniques, such as uncertainty and influence analyses, should be utilized as geomorphologists, hydrologists, planners, and managers become more dependent upon models in their work.

Acknowledgement We would like to convey our gratitude to our colleagues who provided thorough and thoughtful review of this work: R. F. Hadley, R. H. Hawkins, M. R. Karlinger, L. J. Lane, H. B. Osborn, and J. J. Stone. This report was enhanced by their expertise and diligence.

References

ASCE (American Society of Civil Engineers) Task Committee on Evaluation Criteria for Watershed Models, (1990) Summary report: Evaluation criteria for watershed models. Proceedings, watershed planning and analysis in action. NY: American Society of Civil Engineers. pp 386–394

Dawdy DR and Lichty RW (1968) Methodology of hydrologic model building. Proceedings, use of analog and digital computers in hydrology, vol. 2. Tucson Arizona: International Association of Science Hydrology. pp 347–355.

Diskin MH (1970) Definition and uses of the linear regression model. Water Resour Res 6:1668–1673

Foster GR (1988) Development and use of models for management of runoff and erosion processes. Proceedings, modeling agricultural, forest, and rangeland hydrology. St. Joseph, Michigan: American Society of Agricultural Engineers. pp 156–165.

Gilroy E (1991) Uncertainty in estimates of suspended sediment load. Proceedings, fifth interagency sedimentation conference. Washington, DC: US Government Printing Office. pp 9-17 to 9-23
Haan CT (1988) Parametric uncertainty in hydrologic modeling. Proceedings, modeling agricultural, forest, and rangeland hydrology. St. Joseph: American Society of Agricultural Engineers. pp 330–346
Hammond R and McCullagh PS (1974) Quantitative techniques in geography: An introduction. Oxford: Clarendon Press. 318 pp
Hardison CH (1971) Prediction error of regression estimates of streamflow characteristics at ungaged sites. Geological Survey Research, Chapter C. US Geological Survey Professional Paper 750-C. pp C228–C236
Hoaglin DC, Mosteller F and Tukey JW (1985) Exploring data tables, trends, and shapes. NY: John Wiley & Sons. 527 pp
Knisel WG (1980) Erosion and sediment yield models—an overview. Proceedings, watershed management '80. Boise, Idaho: American Society of Civil Engineers. pp 141–150
Leopold LB and Langbein WB (1967) Association and indeterminacy in geomorphology. In: The fabric of geology. Albritton, pp 184–192
Mark DM and Church M (1977) On the misuse of regression in earth sciences. Math Geol 9:63–75
Osterkamp WR, McNellis JM and Jordan PR (1978) Guidelines for the use of structural versus regression analysis in geomorphic studies. Water Resources Investigations, 78–135, US Geological Survey. 22 pp
Renard KG and Meyer LD (1986) The effect of erosion on productivity: Role of mathematical models. Forum on erosion productivity impact estimators. US Department of Agriculture, Soil Conservation Service, pp 4–13, 88–110
Sorooshian S (1981) Parameter estimation of rainfall-runoff models with heteroscedastic streamflow errors—the noninformative data case. J Hydrol 52:127–138
Tasker GD and Stedinger JR (1989) Operational GLS model for hydrologic regression. J Hydrol 111:361–375.
Troutman BM (1985a) Errors and parameter estimation in precipitation-runoff modeling: 1. Theory. Water Resour Res 21:1195–1213
Troutman BM (1985b) Errors in parameter estimation in precipitation-runoff modeling: 2. Case study. Water Resour Res 21:1214–1222
Wagenet RJ (1988) Modeling soil hydrology: Perspectives, perils, and directions. Proceedings, modeling agricultural, forest, and rangeland hydrology. St. Joseph, Michigan: American Society of Agricultural Engineers. pp 1–9
Williams GP (1983) Improper use of regression equations in earth sciences. Geology 11:195–197
Woolhiser DA and Brakensiek DL (1982) Hydrologic system synthesis. Hydrologic modeling of small watersheds, Monograph 5. St. Joseph, Michigan (AISO) American Society of Agricultural Engineers, pp 3–16