Assessing uncertainties in WEPP's soil erosion predictions on rangelands

M. Tiscareno-Lopez, M.A. Weltz, and V.L. Lopes

ABSTRACT: An assessment of uncertainties in predictions was performed for the WEPP model by identifying the contribution of parameter variance and model bias on the Mean Square Error of three model response variables. A Monte Carlo simulation scheme with correlated variable generation was assembled into WEPP to produce a large number of model responses to be compared to observed data of a semi-arid rangeland watershed. The behavior of errors and prediction intervals illustrate the model's ability to simulate the system. This analysis answered relevant questions related to model uncertainty and to identify noisy components in the model.

KEY WORDS: soil, erosion, modeling, uncertainty analysis, Monte Carlo, parameter error, model bias, prediction intervals, rangeland, watershed.

Mario Tiscareno-Lopez is a hydrologist with the Centro Nacional de Produccion Sostenible - INIFAP - SAGAR, Michoacan 58260, Mexico. Mark A. Weltz is a hyydrologist with the Southwestern Watershed Research Center, USDA-ARS, Tucson, AZ. and Vicente L. Lopes is a hydrologist and Assistant Professor at the School of Renewable Natural Resources, The University of Arizona, Tucson, 85721.

J. Soil and Water Cons. 50(5) 512-516

Major advances in soil erosion prediction technology have been obtained in the implementation of process-based models that account for the spatial and temporal variability of watersheds when estimating the effects of land-use practices on soil erosion. This is the case of the USDA Water Erosion Prediction Project (WEPP) model, a distributed parameter, process-based, continuous simulation model that simulates erosion and sedimentation processes in cropland and rangeland watersheds. WEPP is based on the modern concepts of weather generation, hydrology, soil physics, plant science, hydraulics, and erosion mechanics (Lane and Nearing 1989).

WEPP simulates many watershed processes to improve the accuracy of predictions. Unfortunately, WEPP is still an abstraction of reality and uncertainty arises in model components. Model complexity and natural variability of the system frequently induce uncertainty in predictions. It is necessary to identify critical periods for which the model produces unrealistic responses as the result of errors in continuous simulations. Typical questions are: Is model error or parameter error decreasing or increasing with time? If error diminishes with time, what is the minimum number of years to obtain reliable predictions? The objective of this study was to assess prediction uncertainty for three response variables of the 93.13 version of the WEPP model when applied to a small semi-arid rangeland watershed.

Relevant literature

Prediction uncertainty is associated with a variety of factors including the model error, caused by the abstraction of a physical system represented by a set of mathematical equations. Model error—also known as structural error—arises because the governing laws controlling the processes are not known precisely. Model error depends on the level of aggregation of components representing the system. Trade-offs exist between model complexity, accuracy of parameters, and input data (Kirchner 1991). As models become more complex, data and parameter requirements usually become greater. Adding complexity to a model may improve its ability to describe the system's behavior, but it may increase prediction uncertainty.

Traditionally, model structure is evaluated using the following four statistical assumptions about model error: 1) the errors are statistically independent of the predictions and are identically distributed, 2) the errors are statistically independent of each other, 3) the error has a zero mean, μ; and a finite variance, σ^2; and 4) the errors are normally distributed. If one of these assumptions is violated then it is said that the model contains structural errors. The experience with process-based models has demonstrated that errors associated with even optimal parameter sets are neither zero nor normally distributed. The difficulty of testing models using traditional statistical schemes results from threshold parameters, parameter correlation, heteroscedasticity in the residuals, and insensitive parameters (Beven and Binley 1992).

Perhaps parameter error has received most attention in WEPP. Sensitivity analyses have shown that small errors in parameters can result in large errors in model responses (Flanagan and Nearing 1991; and Tiscareno-Lopez et al. 1993). Errors due to parameters frequently result from measurement errors and inaccuracy of empirical equations that calculate parameters to represent the natural variability. Model components intend to represent the watershed behavior under natural or stressed conditions. Runoff and erosion calculations depend on parameters representing the spatial and temporal variability. Spatial variability is imposed by changes in soil and vegetation. The temporal variability is the effect of climatic cycles on plant growth.

Procedure

Estimation of errors in model predictions. Total uncertainty for the WEPP model has accounted for two sources of

Table 1. Statistics of watershed characteristics used for this analysis

	Unit	Mean	Std. Dev.
	Input parameters		
Rock fragments	%	23.6	8.9
Sand	%	62.7	8.7
Clay	%	23.0	7.1
Silt	%	14.2	2.5
Bulk density	g cm^{-3}	1.31	0.09
Organic matter	%	1.67	0.5
1/3-Bar soil moisture	%	23.9	5.2
15-Bar soil moisture	%	13.3	2.9
Cation exchange capacity	meq	23.6	8.3
Rock cover	%	28.0	6.0
Litter	kg m^{-2}	0.050	0.028
Standing biomass	kg m^{-2}	0.10	0.036
Saturated hydraulic conductivity	mm hr^{-1}	4.07	2.32
Interrill erosion parameter	kg s m^{-4}	216,286.0	41,814.0
Rill erosion parameter	s m^{-1}	0.00717	0.00135
Critical shear stress parameter	Pa	1.07	0.082
	Observed watershed data		
Rainfall depth	mm	21.05	7.79
Runoff volume	mm	5.87	4.68
Peak runoff	mm hr^{-1}	25.07	20.95
Sediment yield	kg	184.40	342.47

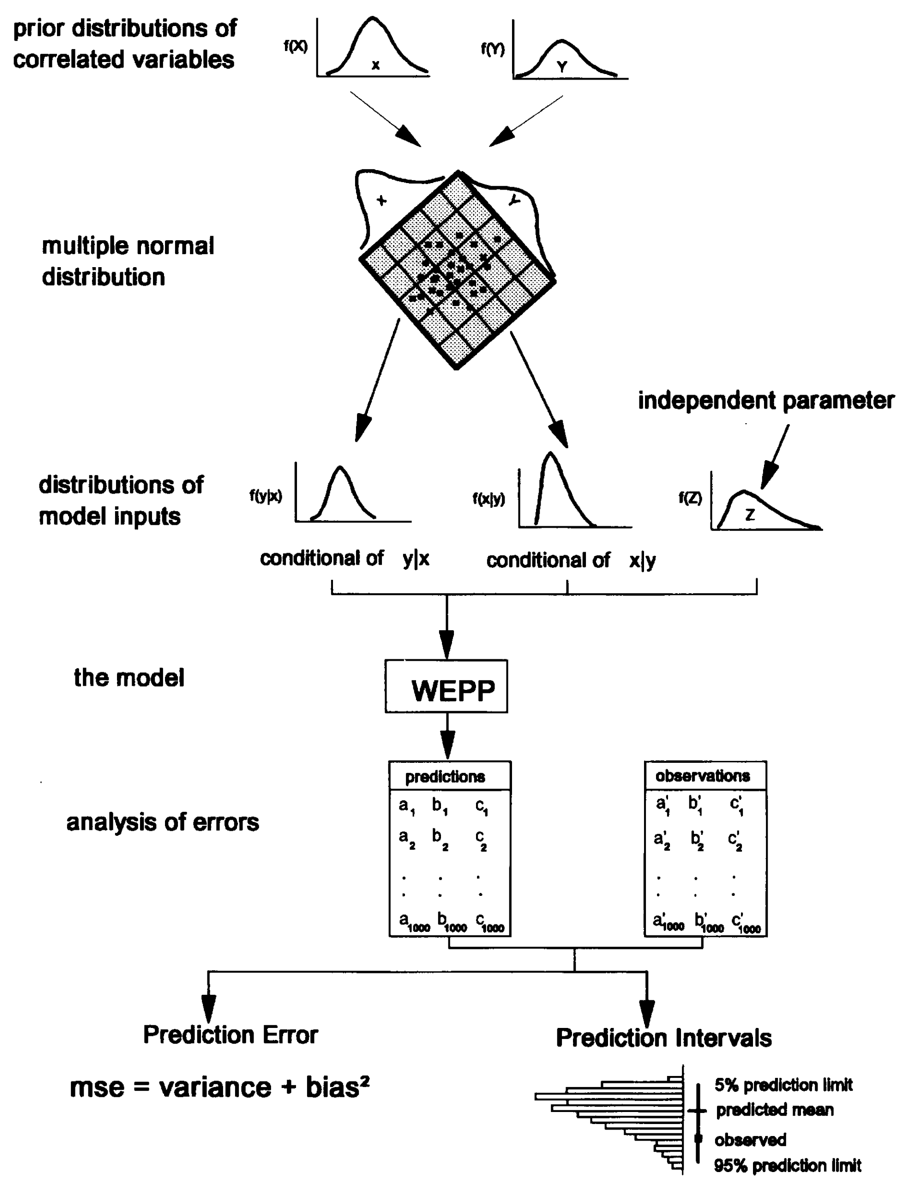

Figure 1. Monte Carlo simulation using correlated deviates generation for error quantification and prediction intervals identification

Note: Where x, y and z represent model imputs, and a′, b′, and c′ are the observed data of the a, b and c model predicted variables.

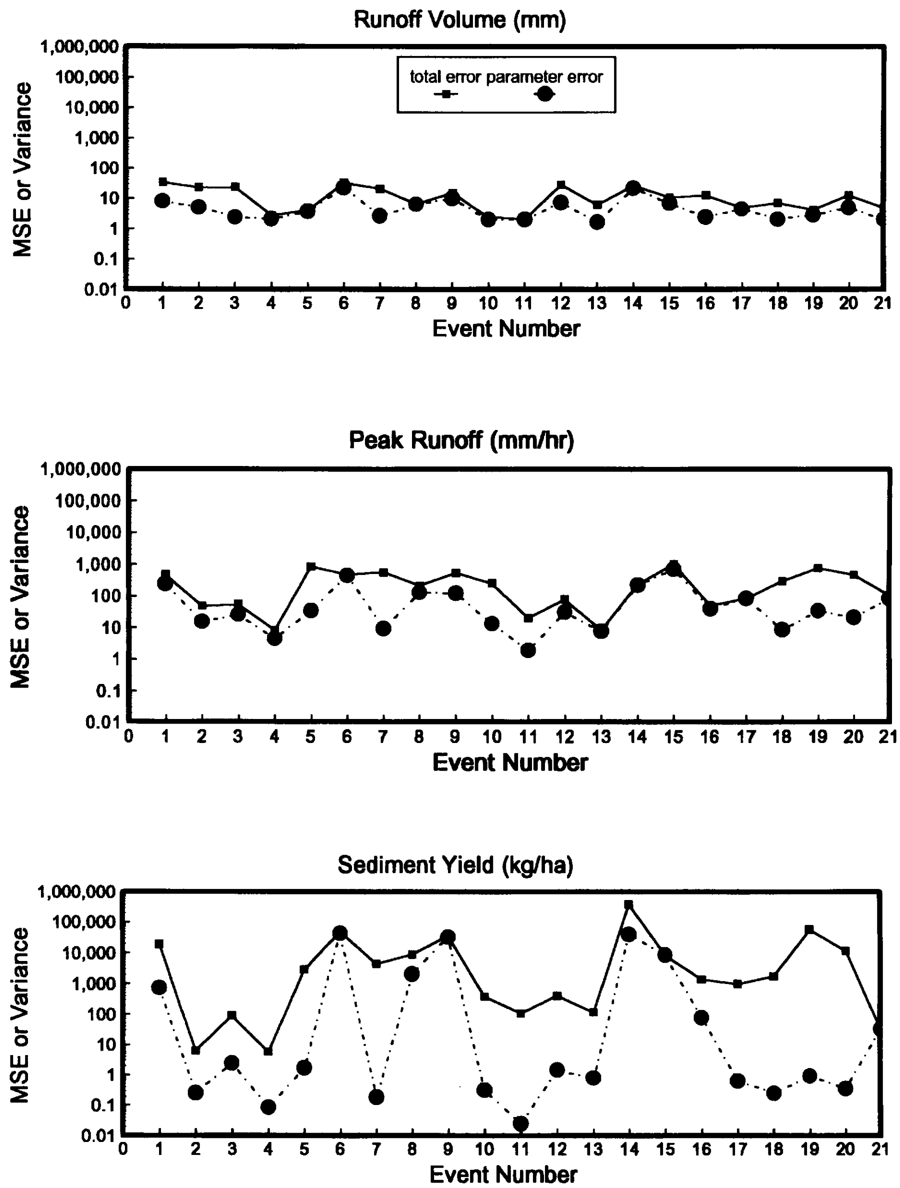

Figure 2. Total error and parameter error by event for runoff volume, peak runoff and sediment yield

uncertainty in predictions: a) model variance, by analyzing the variability of model responses around the mean of predictions caused by parameter variability; and b) model bias, by accounting the difference between the expected behavior of the model when parameters are uncertain and observed data. Total uncertainty was assessed as Mean Square Error (MSE):

$$\mathrm{MSE} = \mathrm{VARIANCE} + \mathrm{BIAS}^2 \qquad (1)$$

where

$$\mathrm{BIAS}^2 = (y - \bar{y})^2 = (\text{observed} - \text{mean of predictions})^2 \qquad (2)$$

$$\mathrm{VARIANCE} = \Sigma(\hat{y} - \bar{y})^2 / \mathrm{n} = (\text{predicted} - \text{mean of predictions})^2 / \mathrm{n} \qquad (3)$$

$$\mathrm{MSE} = \Sigma(y - \hat{y})^2/\mathrm{n} = (\text{observed} - \text{predicted})^2/\mathrm{n} \qquad (4)$$

and where y is the observed value, $\bar{y}$ is the mean of predictions, $\hat{y}$ is the predicted value and n is the number of observations. Since model variance is the result of parameter variability, model variance is considered as a result of parameter error. Model bias, a measure of model accuracy, represents the model's ability to make good estimates. It was assumed that observed data, including precipitation records, were free of errors for an appropriate assessment of uncertainties. The decision criterion utilized to include parameters for this study was based on the results of a sensitivity analysis previously performed (Tiscareno-Lopez et al. 1993).

Confidence in model predictions. A probabilistic assessment of prediction uncertainty was calculated using prediction intervals in this study. Prediction intervals specify the probability that a deterministic prediction lies within a range of model output (Graybill 1976). Based on preliminary model runs under continuous simulation, the probability distributions of model responses are non-normally distributed and vary in shape from day to day. Prediction intervals were obtained by rejecting the upper and lower 5% of the simulations.

Implementation. Monte Carlo simulation with correlated variable generation was implemented to produce a large number of responses for predicted variables as illustrated in Figure 1. Error sources and prediction intervals were identified using observed data of runoff volume, peak runoff, and sediment yield of K2, a small rangeland watershed located in the Walnut Gulch Experimental Watershed near Tombstone, Arizona. A climate file was constructed using the records of a meteorological station located in the watershed. The approach is summarized as follows:

1. Soil and vegetation were sampled throughout K2 to obtain basic statistics (Table 1).
2. To generate random deviates of soil characteristics, a multiple normal distribution was used to preserve correlation between variables. Model parameters were assumed uncorrelated and they were generated from univariate distributions.
3. A set of model inputs was drawn from corresponding probability distributions. A 14-years continuous simulation was performed by entering this set of model inputs and climate data.
4. Model responses for 21 events, for which observed data exist, occurring within the 14-years simulation were saved at the end of the simulation.
5. Steps three and four were repeated for 1,000 simulations.
6. Errors were then quantified for runoff volume, peak runoff, and sediment yield model responses by solving Equation (1) for each event after 1,000 simulations (n=1,000).
7. Using probability distributions of model responses, prediction intervals were identified.

Results

Figure 2 shows variations of total error and parameter error for model responses

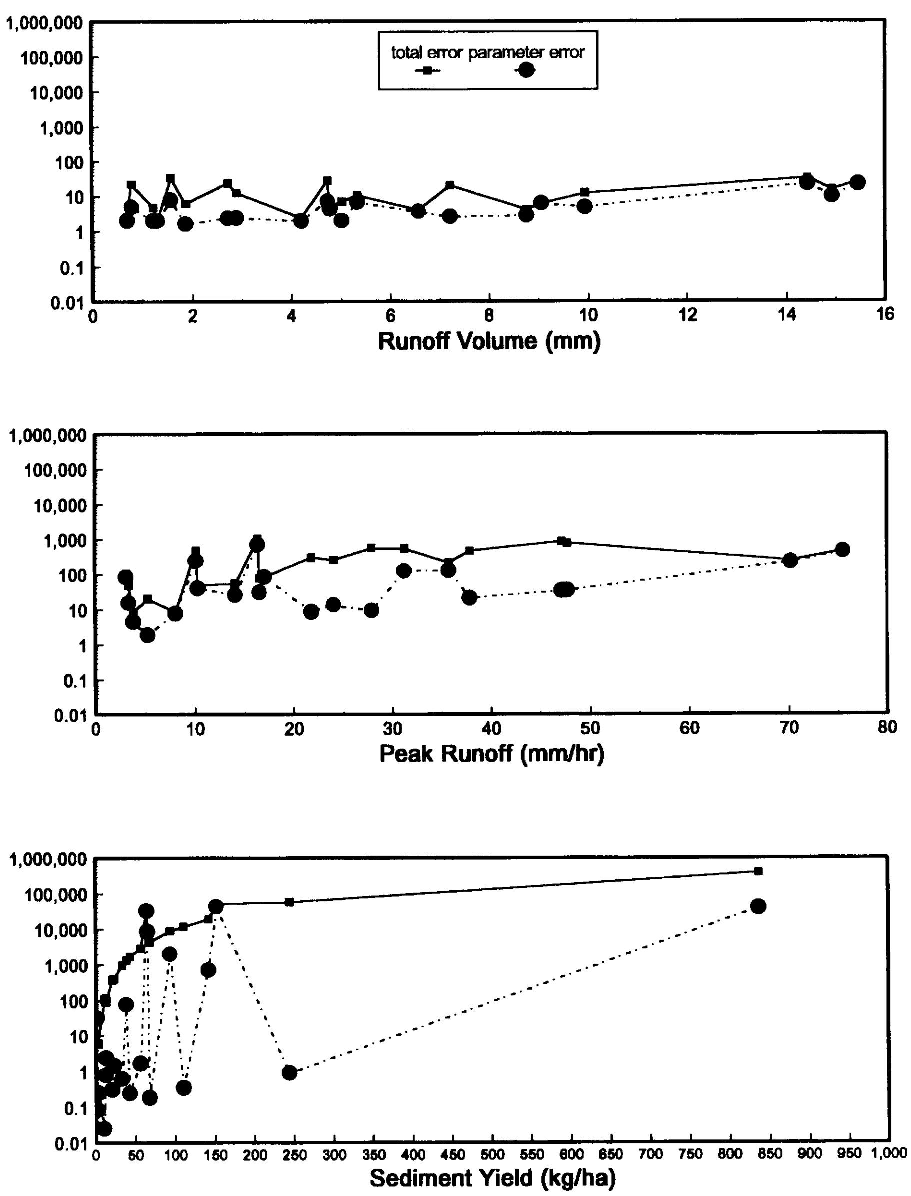

Figure 3. Errors by the magnitude of runoff volume, peak runoff and sediment yield

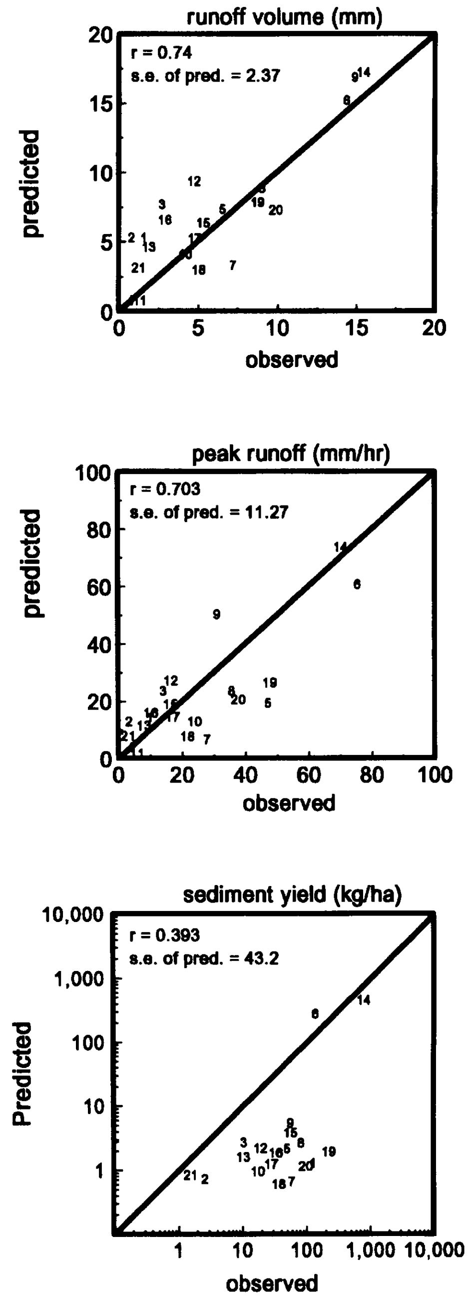

Figure 4. Observed and predicted variables with parameters at calibration

corresponding to 21 rainfall-runoff producing events, chronologically ordered, within the 14-years simulation. Model bias is the difference between total error and parameter error. According to these three plots, all errors maintained a uniform spread over time, indicating that errors do not increase along the time of simulation. However, error propagation was detected because the MSE and variance increased from runoff volume to sediment yield.

To identify heteroscedasticity problems—larger events have larger associated prediction errors—the magnitude of errors was plotted on the magnitude of observed runoff volume, peak runoff, and sediment yield (Figure 3). Errors are uniformly distributed for runoff volume, and there is a moderate tendency for errors to increase for peak runoff. However, errors increase significantly for sediment yield predictions. All three error types tended to increase for large rainfall events and model bias was the most contributing to total error. On average, model bias contributed 45.8, 55.1, and 79.6% of total error for predicting runoff volume, peak runoff, and sediment yield, respectively. A frequent bias in estimating the sediment yield revealed problems in erosion calculations.

Error propagation from lower to higher levels of aggregation was identified. This effect became noticeable when a 14-year simulation was performed with model inputs at their mean value of sample (Table 1) and initiating the model with calibrated parameters (Tiscareno-Lopez 1994).

Figure 4 illustrates one-to-one scattergrams of model responses. Events along the 45° line indicate a perfect fit. The numbers inside the graph correspond to the events. Runoff volume is scattered along the 45° line, showing that WEPP behaves satisfactorily to estimate infiltration variables related with the estimation of runoff volume. Peak runoff was under predicted for some events, however the sediment yield plot shows the highest dispersion about the 45° line. A cluster of events in the lower part of the scattergram revealed sediment yield under prediction. The accumulation of error is reflected in a reduction of the coefficient of correla-

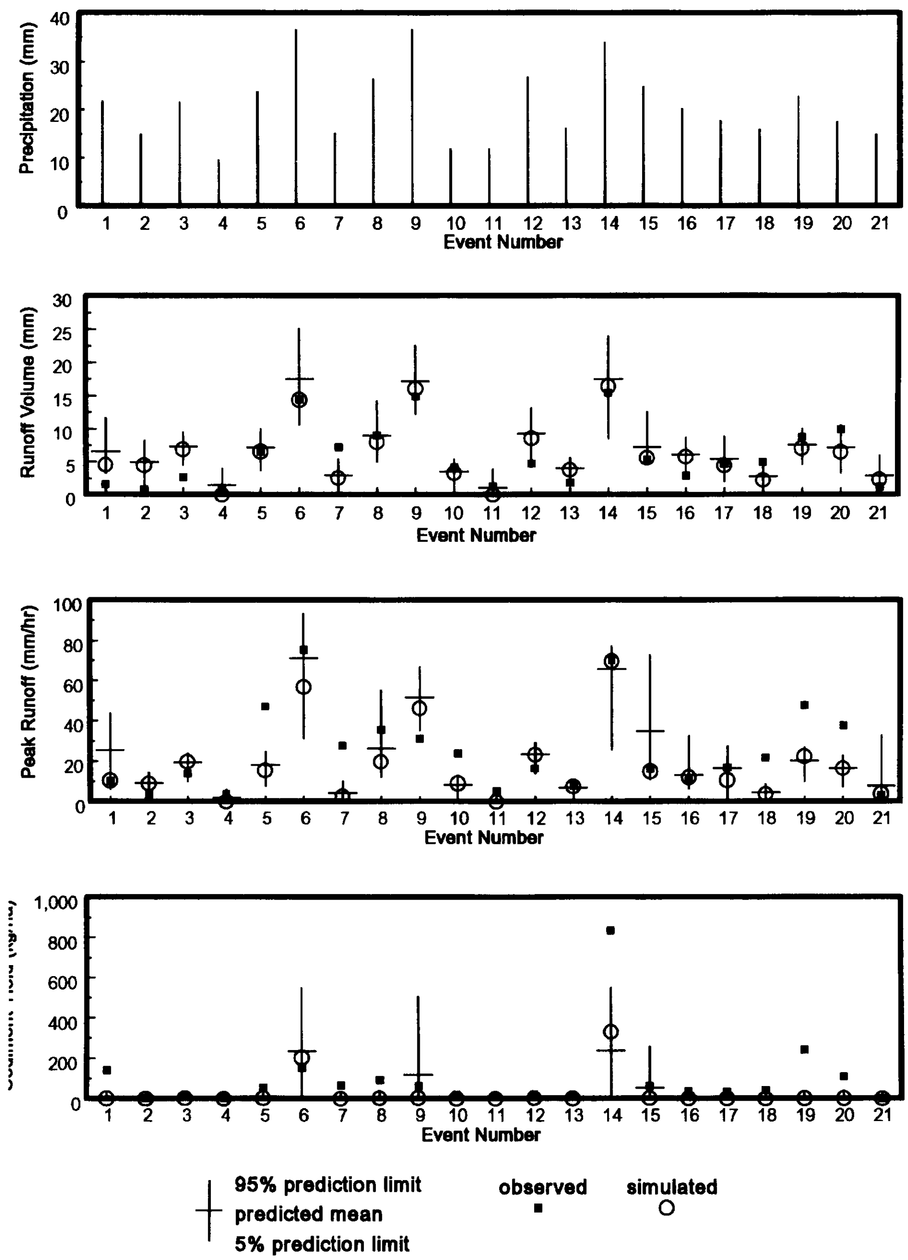

Figure 5. Prediction intervals of runoff volume, peak runoff and sediment yield

tion from 0.74 on runoff volume to 0.39 on sediment yield estimates. Note that runoff events resulting without erosion response were not included in the sediment yield plot since logarithm of zero does not exist.

Figure 5 shows prediction intervals of model responses. The top plot shows the magnitude of the rainfall events. The runoff volume plot indicates that WEPP can make acceptable predictions in continuous simulations since most of the actual observations fell within the 90% prediction intervals. The empty circles represent model predictions when the model was run with calibrated parameters. For peak runoff, 14 of the 21 observed events fell within the range of the prediction. Prediction intervals for sediment yield could not be delineated due to large errors. The reduction of observed events within the 90% prediction intervals is attributed to error propagation between components in the model.

It is important to mention that this and other analyses helped to identify that problems in erosion calculations resulted from interrill erosion under prediction. Then, a sediment transport factor for interrill erosion was included in latter versions of the model to improve predictions.

Conclusions

Based on analyses to estimate uncertainties in the WEPP model, it is possible to conclude the following: (1) errors maintained a uniform spread during continuous simulations, never increasing or decreasing along the time of simulation. This suggests there is not a minimum number of years of simulation required to diminish prediction error; (2) Problems of lack of homoscedasticity were observed, the largest errors associated to the largest rainfall events. This is more evident for components of higher levels of aggregation such as peak runoff and sediment yield; (3) Prediction intervals showed that WEPP simulates acceptable runoff volume responses. Most of the peak runoff observations were comprised within the range of predictions. Because of the large error in estimating sediment yield, most of the sediment yield observations rarely were included inside prediction intervals. This was largely attributed to interrill erosion under prediction.

REFERENCES CITED

Beven, K., and A. Binley. 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrological Processes 6:279–298.

Flanagan, D.C., and M. A. Nearing. 1991. Sensitivity analysis of the WEPP hillslope profile model. ASAE Paper No. 912074. St. Joseph, Michigan.

Graybill, F.A. 1976. Theory and Application of the Linear Model. Duxbury Press, North Scituate, Massachusetts.

Kirchner, T. 1991. Uncertainty and sensitivity analysis. Course notes of the Pathway analysis and risk assessment for enviromental compliance and dose reconstruction. p 30. Radiological Assessments Corporation. South Carolina 29107.

Lane, L.J., and M.A. Nearing (eds.). 1989. USDA-Water Erosion Prediction Project: Hillslope Profile Model Documentation. NSERL Report No. 2. USDA-ARS-National Soil Erosion Research Laboratory, West Lafayette, Indiana.

Tiscareno-Lopez, M., V. L. Lopes, J. J. Stone, and L. J. Lane. 1993. Sensitivity analysis of the WEPP watershed model for rangeland applications I: Hillslope Processes. Transactions of the ASAE 36(6):1659–1672.

Tiscareno-Lopez, M. 1994. A Bayesian-Monte Carlo Approach to Assess Uncertainties in Process-Based, Continuous Simulation Models. Ph. D. Dissertation. The University of Arizona.